FINAL REPORT

AN ASSESSMENT OF THE FEASIBILITY OF DEVELOPING AND IMPLEMENTING AN AUTOMATED PAVEMENT DISTRESS SURVEY SYSTEM INCORPORATING DIGITAL IMAGE PROCESSING

ROGER HOWE
Public Relations Coordinator

GERARDO G. CLEMEÑA, Ph.D.
Principal Research Scientist

V·I·R·G·I·N·I·A
VTRC
TRANSPORTATION RESEARCH COUNCIL

VIRGINIA TRANSPORTATION RESEARCH COUNCIL

Standard Title Page - Report on State Project

Report No.	Report Date	No. Pages	Type Report: Final	Project No.: 9128-010-940
VTRC 98-R1	November, 1997	78	Period Covered: October 1995 - March 1997	Contract No.

Title and Subtitle:
Assessment of the Feasibility of Developing and Implementing an Automated Pavement Distress Survey System Incorporating Digital Image Processing.

Authors:
Roger Howe, Gerardo G. Clemeña

Performing Organization Name and Address:

Virginia Transportation Research Council
530 Edgemont Road
Charlottesville, VA 22903

Key Words:
Digital Image Processing, Pavement, Pavement Distress, Pavement Distress Surveys, Pavement Cracking, Automated Pavement Distress Surveys

Sponsoring Agencies' Name and Address

Virginia Department of Transportation
1401 E. Broad Street
Richmond, VA 23219

Supplementary Notes

Abstract

The rational allocation of pavement maintenance resources requires the periodic assessment of the condition of all pavements. Traditional manual pavement distress surveys, which are based on visual inspection, are labor intensive, slow, and expensive, and they pose a safety hazard when the raters have to get out of their vehicles and inspect the road on foot.

One of the principal goals of this report is to provide the Virginia Department of Transportation with some of the important background required for determining whether, with the current state of development of image processing and camera and computer hardware, it is feasible to develop an automated system to use for pavement distress surveys. This report describes some of the fundamental techniques of image processing that are likely to play a role in a pavement survey system that can automatically recognize and classify cracks. The report concludes that developments in technology during the last 10 to 15 years have made it possible to develop and implement such a system and that the implementation of such a system would mean that surveys would be less expensive, faster, safer, and more objective.

FINAL REPORT

AN ASSESSMENT OF THE FEASIBILITY OF DEVELOPING AND IMPLEMENTING AN AUTOMATED PAVEMENT DISTRESS SURVEY SYSTEM INCORPORATING DIGITAL IMAGE PROCESSING

Roger Howe
Public Relations Coordinator

Gerardo G. Clemeña, Ph.D.
Principal Research Scientist

(The opinions, findings, and conclusions expressed in this report are those of the authors and not necessarily those of the sponsoring agencies.)

Virginia Transportation Research Council
(A cooperative organization sponsored jointly by the
Virginia Department of Transportation
and the University of Virginia)

Charlottesville, Virginia

November 1997
VTRC 98-R1

EXECUTIVE SUMMARY

The rational allocation of pavement maintenance resources requires the periodic assessment of the condition of all pavements. Traditional manual pavement distress surveys, which are based on visual inspection, are labor intensive, slow, and expensive, and they pose a safety hazard when the raters have to get out of their vehicles and inspect the road on foot. Because they are so labor intensive, it is often the case that only a small sample of a state's pavements are assessed in any great detail. There is also a problem with variations in the ratings from manual distress surveys. These variations are the result of three factors: (1) Human subjectivity: Different raters will almost certainly produce evaluations of the same pavement that are different in some respects. In some cases, the differences are significant. Not only are there differences in evaluations between raters, it is to be expected that there will be differences in the evaluations made by the same rater. (2) The rating system: In the SHRP *Distress Identification Manual*, a longitudinal crack is defined as "predominately parallel to the center line," and a transverse crack is defined as "predominately perpendicular to the center line." The borderlines between these categories are not precisely defined. (3) Environmental factors: Lighting is an important environmental factor that affects the way a pavement looks, and lighting changes throughout the day and throughout the year. Often, what is seen as a crack during a windshield survey is the shadow in the interior of a crack. If the interior of a crack is in dark shadow, it will contrast much more strongly with the pavement surface than the interior of a crack that is fully illuminated; consequently, when the shadows change as a result of changes in the light, the crack will not look the same. In some cases, it may be hard to see it at all.

A survey vehicle suitable for comprehensive pavement management should be oufitted with a variety of subsystems, for example, a laser system to calculate rutting and roughness, one or more video cameras to provide a videolog of the right-of-way, and an automated crack identification system. Other systems could, of course, be added; however, the principal goal of this report is to provide VDOT with some of the important background required for determining whether, with the current state of development of image processing and camera and computer hardware, it is feasible to set up an automated system for recognizing and classifying cracks. This report describes some of the fundamental techniques of image processing that are likely to play a role in such a system. Understanding these basic techniques is essential to understanding both the task of coming up with suitable hardware for a system and the task of developing suitable algorithms designed specifically to process pavement distress data. Some of the problems of developing a suitable set of algorithms for the recognition, classification, and measurement of cracks in pavements are discussed, and one complete set of algorithms from an NCHRP project is described in some detail. Finally, a detailed account of the requirements of an image acquisition system is provided.

Digital image processing is not a new technology. NASA pioneered its use back in the early 1960s on the Ranger missions, which were a prelude to the Apollo manned lunar exploration missions. Because the cost of image processing hardware and software has gotten lower and lower during the last 15 years, there has been growing interest in using it for pavement condition surveys. Data collection and data interpretation can both benefit from the application of this technology. The implementation of an automated distress survey system would mean that surveys would be less expensive, faster, safer, and more objective. Depending on the type and extent of the surveying that is now done, all of these benefits may come with the added advantage of a substantial increase in the quantity of pavement covered.

Developments in technology during the last 10 to 15 years have made it possible for VDOT to develop and implement such a system. Several states and several commercial operators have prototype systems in service now. These prototypes differ in many ways: there are differences in the algorithms they use, in the system hardware, in the placement of the cameras, in the types of cameras, in the resolution of the system, in the amount of coverage of the pavement, etc. Some of these systems are fully automated and some are hybrids that incorporate both automated and manual operations.

Probably the best way to describe the state-of-the-art at the present time is to say that the entire field is going through a period of intense development. The application of image processing techniques to the problem of recognizing and classifying cracks is an area that appears to be wide open for development. The principal focus has been on the development of algorithms, although there have been some serious difficulties with hardware that had to be resolved. Actually, many of the problems with hardware that taxed researchers in the 1970s and 1980s have been resolved by the rapid improvements in hardware since that time. Various individuals and organizations have been trying to develop systems of algorithms that will accurately and consistently recognize, classify, and measure cracks in pavements; however, no single system of algorithms has achieved general acceptance as the best. Furthermore, there seems to be more than one way to achieve acceptable results. In many respects, however, it is difficult to assess the "success" of some of these systems because they have been tested only on a small number of images and only on a few types of cracks. Michael Burke, who is the author of the three-volume *Handbook of Machine Vision Engineering*, puts the problem of developing a suitable set of algorithms for analyzing pavement distresses in this way:

> A large number of image processing algorithms have been designed for analyzing video images. This is the heart of the problem: there are literally thousands of algorithms available, but many are not suited for one reason or another. . . . Algorithm selection for image processing system design is still very much an art and not a science. The system tends to be highly nonlinear, and so simply reversing the order in which two processes are applied can drastically change the outcome (the analyzed output). Furthermore, each algorithm tends to have a "magic number" or two, parameter values which must be set (such as gain, window size, etc.), and varying these can also change the outcome. Furthermore, varying the parameters of one processing step can change the effectiveness of the settings of a different processing stage (again, because of the nonlinearities involved). We are looking at on the order of fifty such processing steps, all

interacting with one another in highly nonlinear and unpredictable ways. *Algorithm selection is therefore a very formidable obstacle.* (emphasis added)

A system designed to recognize, classify, and measure cracks in pavements will consist of an image acquisition system and an image processing system. Despite the integral relation between these two components, to some extent, they can be developed separately. If an automated system were developed by VDOT, a long period of research and development would be required before pavement distress surveys could be turned over to it. One of the major failings of some of the previous attempts to develop algorithms is that the algorithms have been tested only on a small sample of pavement images; however, one thing that can be clearly ascertained from a perusal of the literature is that the immense variability in the appearance of actual pavements makes them a far more difficult test for systems of algorithms than any sample. Before a system can be trusted to provide consistently accurate output, the algorithms will have to be tested on high resolution images of many kilometers of actual pavements. This would, of course, require that VDOT either have a suitable image acquisition vehicle in operation or at the very least that the developers of the algorithms have access to the output of some survey vehicle even if it did not belong to the Department. However, it would be possible to *begin* this process of development by using a large number of images of pavements in which the more important types of cracks are represented with as much variety as is practical.

VDOT has in its possession high resolution films of pavement sites that are a part of the long-term pavement performance study of SHRP. These films are made at regular intervals by the PASCO Corporation using a high resolution system that incorporates a 35 mm slit camera. They are of sufficiently high quality that they could be used at least in the initial stages of the development of a system of algorithms in lieu of outfitting an image acquisition vehicle to collect images, which would be expensive.

Using the PASCO films to provide a set of test images would involve the following:

- The PASCO films would be examined for suitable images of pavement cracking. At the very beginning, the number of types of cracks would be limited. Perhaps only longitudinal and transverse cracks would be included. However, as many different variations of each of these types as is practical would be included. In other words, although many of the cracks that would be classified as transverse are similar in shape, those that are chosen to represent the range of cracks included in the category *transverse* would include as wide a range of shapes as possible because there is no point in testing algorithms on a set of images of transverse cracks that look very much alike. It is important that the cracks look as different from each other as possible in order to provide the algorithms with a test that includes some at least of the variety of shapes of cracks certain to be found in a normal pavement. There should be no difficulty getting the required variety of shapes from the PASCO films, which cover many kilometers of pavement.

- Once suitable images of cracks are found on the PASCO films, those portions of the film containing images of the selected types of cracks would be digitized. At the end of this process, we should have several hundred digitized images of cracks. This collection of images would be the first set of test samples.

- This set of test samples would be used in the first stage of the process to develop and refine algorithms that can adequately recognize and classify the longitudinal and transverse cracks that occur in the first sample. Once a suitable set of algorithms has been developed to deal with the first set of samples, a second set of test samples could be culled from the PASCO films. This second set of samples would include images of different types of cracks. In this way, by continually increasing the demands on the algorithms, a set of algorithms could be developed that would be suitable for preliminary testing on the output from a survey vehicle, which would, of course, provide a much more difficult test.

There would be several advantages to proceeding in this way. First, the long process of developing a system of algorithms could commence without the need to purchase and outfit a survey vehicle. Second, the developers of the algorithms will gain experience with the image processing software and hardware on a fixed set of images, and this will be a good prelude to their application to the output of a survey vehicle. Finally, although the system of algorithms developed during this initial stage of the project will almost certainly have to be modified when it is applied to the output of a survey vehicle, there is no reason to believe that all or even a significant amount of the developmental work done during this time would not be of use in the later stages of the project. In fact, it would be better to have one or more alternative systems of algorithms ready to test on the output of a survey vehicle when it is finally outfitted and ready to undertake the distress surveys.

At some point, it will be necessary to test the algorithms on the output from a survey vehicle. There are many options for outfitting a comprehensive pavement survey vehicle, which would also serve as a platform for a bank of lasers for determining roughness and rutting and for a standard video camera providing a windshield view of the road. The choice of a particular hardware configuration for imaging the pavement surface depends on the needs of the user. A general idea of the cost of the imaging hardware on a survey vehicle can be derived from the following. The costs listed below are for a system capable of resolving 2 mm cracks, providing full coverage of the width of the lane and full coverage down lane, and operating at a speed of 72 km/hr (45 mph): (1) four line scan cameras: $20,000, (2) four dedicated frame grabbers: $17,000, (3) 3.66 m of sodium lighting specifically designed for line scan cameras (with a life of approximately 25,000 hr): $30,000, (4) on-board memory using a RAID system with a capacity of 90 gigabytes: $54,000, (5) Pentium PC: $3,000, and (6) two high resolution monitors (1,500 pixel horizontal resolution): $5,000. The total is $129,000. The prices for mounting the cameras, for providing shock-absorbent stands in the vehicle for the computers, as well as other necessities for setting up the survey vehicle are not included, nor is the price for the technical assistance that will have to be purchased to get the system set up and running and to eliminate all of the bugs.

The images that the survey vehicle would make during its regular surveys would make an important contribution to the development of the algorithms, which will be a long incremental process of trial and error. It is important for these images of real pavements to be rated by human raters during this stage of development, otherwise it would not be possible to tell how well the image processing system is performing. It is only by comparing the performance of the image processing algorithms with the ratings of human raters who have been trained to rate pavements using criteria endorsed by VDOT that the algorithms can be tested. In this way, the algorithms can be evaluated by comparing the results of applying them to a set of images that have been examined for cracks by human raters. It would also be important to compare the images from the image acquisition system to the actual condition of the pavement surface. It is essential to verify that the images accurately represent what is perceived to be the actual condition of the pavement surface. The ultimate goal is, of course, to provide a system that assesses pavement surfaces for cracks in the way that VDOT wants them assessed, and this would not be possible without constantly comparing the ratings of the image processing system with the ratings of human raters trained by the Department and by comparing the output of the image acquisition system with evaluations of actual pavements by inspectors.

An important benefit of the long-term process of developing the image processing system is that it allows everyone involved in the process to develop confidence in the performance of the system. Since everyone will be able to see the system's failures and the steps taken to overcome them, it will be possible to know what the system is doing and to have confidence in its performance. Incremental development of this sort is essential to getting over uncertainties about the system's performance. Everyone involved in the development of this system in the Maintenance and Materials Divisions and at the Research Council will be kept informed of each step in the development of the algorithms, which will be essential to ensuring that VDOT will have confidence in the system when it is fully implemented.

Although it would be possible to develop a survey system at the Research Council, it may be more practical to contract for pavement distress surveys with one or more of the commercial concerns that have developed the capability to perform pavement surveys using image analysis. In this case, the Research Council could take on the role of evaluating the accuracy of the surveys and help to customize the algorithms of the contracted survey system so that they are more suitable to the pavements in Virginia. A decision with respect to further research in this area will depend on support from the Maintenance Division and the availability of resources.

TABLE OF CONTENTS

FINAL REPORT

AN ASSESSMENT OF THE FEASIBILITY OF DEVELOPING AND IMPLEMENTING AN AUTOMATED PAVEMENT DISTRESS SURVEY SYSTEM INCORPORATING DIGITAL IMAGE PROCESSING

Roger Howe
Public Relations Coordinator

Gerardo G. Clemeña, Ph.D.
Principal Research Scientist

INTRODUCTION

The rational allocation of pavement maintenance resources requires the periodic assessment of the condition of all pavements. Traditional manual pavement distress surveys, which are based on visual inspection, are labor intensive, slow, and expensive, and they pose a safety hazard when the raters have to get out on the road. Because they are so labor intensive, it is often the case that only a small sample of a state's pavements are assessed in any great detail. In addition to these problems, traditional distress surveys are plagued by problems of subjectivity. Different raters will almost certainly produce evaluations of the same pavement that are different in some respects. In some cases, the differences are significant. Not only are there differences in evaluations between raters, it is to be expected that there will be differences in the evaluations made by the same rater. Some of these differences are the consequence of the inherent subjectivity of the human effort to evaluate or assess anything. For example, fatigue, boredom, and distractions can have a detrimental effect on consistency.

Some variations in ratings are the consequence of the rating system that is being used. In the SHRP *Distress Identification Manual*, a longitudinal crack is defined as "predominately parallel to the center line," and a transverse crack is defined as "predominately perpendicular to the center line." The borderlines between these categories are not precisely defined. It is not likely that the cracks that fall in the center of these ranges will often be inconsistently classified. However, when a crack to be categorized falls on the borderline between "predominantly parallel to the center line" and "predominantly perpendicular to the center line," then it could be classified differently at different times by the same rater. The situation would not be improved by the addition of a category of diagonal cracks between longitudinal and transverse because the problem is the lack of a precise quantified boundary not the absence of a sufficient number of categories. Providing a new category with similarly imprecisely defined boundaries would not solve the problem.

Finally, some variations in ratings are the result of environmental factors. Lighting is a very important environmental factor that affects the way a pavement looks, and lighting changes throughout the day and throughout the year. On a bright sunny summer day, the angle of incidence of the light changes radically. In the early morning and the late afternoon, the angle of the light will produce dark shadows in the cracks, whereas at noon when the direction of the light is from directly above the pavement, the interiors of the cracks will be illuminated. On dark overcast days, the contrast between the interiors of the cracks and the surface of the pavement will be reduced, and on days in which there is a high bright overcast, the interiors of the cracks will not be distinctly shadowed at any time of the day, and the faint shadows that are present will not be nearly as dark as those that are created on bright sunny days. Often, what is seen as a crack during a windshield survey is the shadow in the interior of a crack. A dark shadow contrasts much more strongly with the pavement surface than a fully illuminated interior of a crack; consequently, when the shadows change as a result of changes in the light, the crack will not look the same. In some cases, it may be hard to see it at all. A further complication is moisture on the pavement--a moist pavement does not reflect light the same way that a dry one does.

Developments in digital image processing and in a whole range of digital and analog image acquisition systems have made the automation of distress surveying a real possibility. Digital image processing is not a new technology. NASA pioneered its use back in the early 1960s on the Ranger missions, which were a prelude to the Apollo manned lunar exploration missions. Because the cost of image processing hardware and software has gotten lower and lower during the last 15 years, there has been growing interest in using it for pavement condition surveys. Data collection and data interpretation can both benefit from the application of this technology. The implementation of an automated distress survey system would mean that surveys would be less expensive, faster, safer, and more objective. Depending on the type and extent of the surveying that is now done, all of these benefits may come with the added advantage of a substantial increase in the quantity of pavement covered.

One of the principal goals of this report is to provide the Virginia Department of Transportation with some of the important background required for determining whether, with the current state of development of image processing and camera and computer hardware, it is feasible to set up an automated system for pavement distress surveys. This report describes some of the fundamental techniques of image processing that are likely to play a role in a survey system designed to automatically recognize, classify, and measure cracks in pavement. Understanding these basic techniques is essential to understanding what will be required to come up with suitable hardware and suitable algorithms. Some of the problems of developing algorithms for the detection and classification of cracks in pavements are discussed, and one complete set of algorithms from an NCHRP project is described in some detail. Finally, a detailed account of the requirements of an image acquisition system is provided. Although image acquisition must take place prior to image processing, in this report it is discussed *after* image processing because a knowledge of the nature of the digital image and of image processing requirements is required in order to evaluate the many image acquisition options. Gregory Baxes states in his book *Digital Image Processing* that ". . . the spatial and brightness resolutions of the camera directly dictate the quality of the digital image that can be acquired," so it is necessary to understand the quality of the image that is needed in order to be able to choose an image acquisition system.

IMAGE PROCESSING

The term *image processing* in its broader sense refers to image processing operations, image analysis, image compression, image synthesis, and indeed to the whole field. However, in its narrower sense, it is used to refer to a specific set of operations that are a preliminary to image analysis. In this narrower sense, it is easy to see the distinction between image processing and image analysis, because the results of image-processing techniques are pictorial in nature, whereas the results of image analysis operations are almost always the measurement and classification of the information in the image. Another way of putting this would be to say that the point of image analysis operations is "to understand an image by quantifying its elements" (Baxes 1994), whereas the goal of image processing operations is to prepare the image for the quantification and analysis that is to take place. However, it is important to remember that image processing operations are not always a preliminary to analysis; image processing can be used simply to improve the appearance of an image.

The discussion of image processing that follows is limited in scope. It is not meant to provide an account of all of the major areas of image processing or of all of the operations in any particular area. Some of the major areas of image processing--such as frequency domain processing, geometric transformation processing, image compression, image synthesis, and the processing of color images among others--are not described. The goal is to provide a brief account of the way that a few of the fundamental operations in two of the fundamental areas of image processing function. The operations described here are merely a sample. They were chosen because they are fundamental and because they are likely to play a role in any system designed to recognize and classify cracks in pavements. (The account of image processing presented in what follows is indebted to Gregory Baxes' *Digital Image Processing*. The following books have also been found to be helpful: John Russ, *The Image Processing Handbook*; Mike James, *Pattern Recognition*; and Louis Galbiati, *Machine Vision and Digital Image Processing Fundamentals*.)

In Part 1, a basic description of the nature of the digital image is provided. The discussion of image processing in Part 2 focuses on image enhancement; it provides an introduction to pixel point and pixel group processing. In Part 3, the three stages of image analysis are described: segmentation, feature extraction, and object classification. The emphasis in this section is on binary morphological processing. Part 4 is a brief survey of some the common hardware configurations used for image processing. Finally, in Part 5, the current state of algorithm development is briefly described, and one complete system of algorithms used for the detection and classification of cracks is described in some detail.

Part 1: The Digital Image

A continuous-tone image is one in which the various shades and colors blend with no disruptions. Continuous-tone images can either be made up of various shades of gray or of various shades of color. By contrast, a digital image is composed of discrete points of tone (or brightness), which can be shades of gray or colored. To make a digital image from a continuous-tone image, the continuous-tone image must be sampled and quantized. The sampling process samples the intensity of the continuous-tone image at specific locations. The quantization process determines the digital brightness value of each sample. In a gray scale image, the brightness values range in shades of gray from black to white. In a digitized image, a quantized sample is referred to as a *pixel*, which is short for *picture element*. Each pixel has an (x, y) coordinate that corresponds to its location in the image. The x location is usually referred to as the *pixel number*, and the y location as the *line number*. The pixel numbering convention is displayed in Figure 1.

Resolution

How well the digital image represents the original continuous-tone image is dependent on the number of pixels and lines and the range of brightness values in the image. The capability of the digital image to resolve the elements of the original scene is its *resolution*. There are two types of resolution: spatial resolution and brightness resolution.

Spatial Resolution

Spatial resolution refers to the number of pixels in the digital image. The greater the number of pixels, the greater the resolution (see Figure 2). A digital image's spatial resolution is directly dependent on its spatial density and its optical resolution. *Spatial density* refers to the number of pixels in a digital image. *Optical resolution* refers to the capability of the *entire* physical imaging system to resolve the spatial details of an original scene. Optical resolution is thus contingent on the quality of the imaging system's optics, photosensor, and electronics and on the spatial density. The spatial resolution will be determined by the spatial density or the optical resolution, whichever is the lowest.

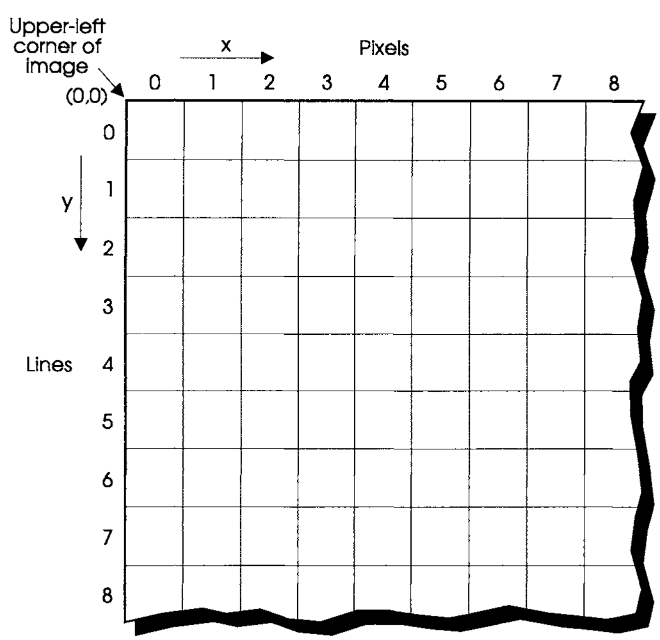
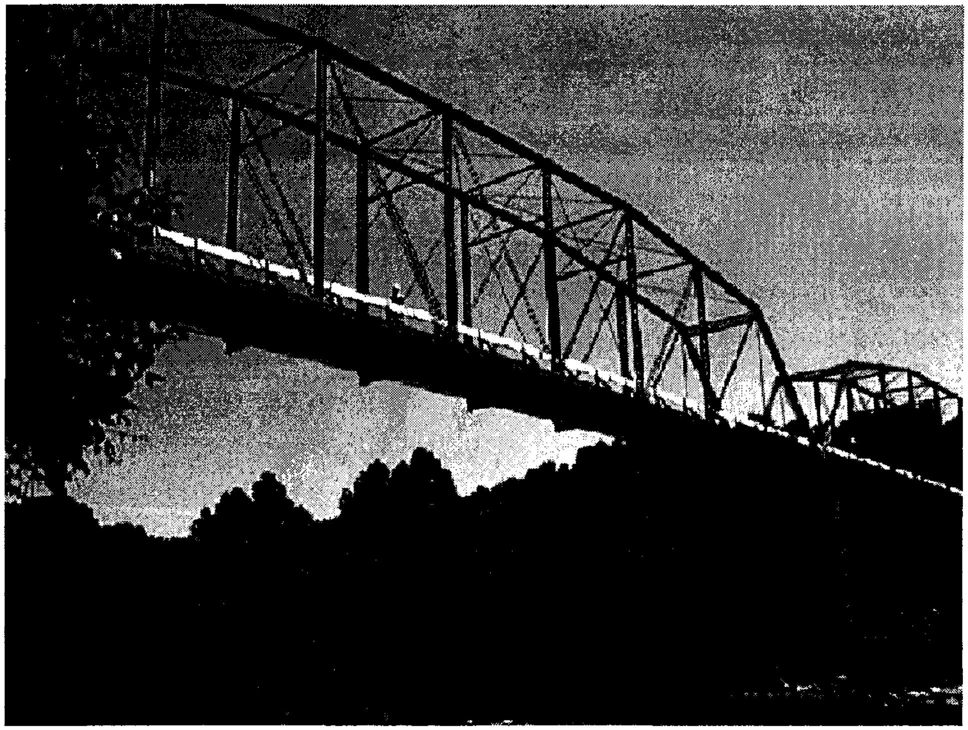
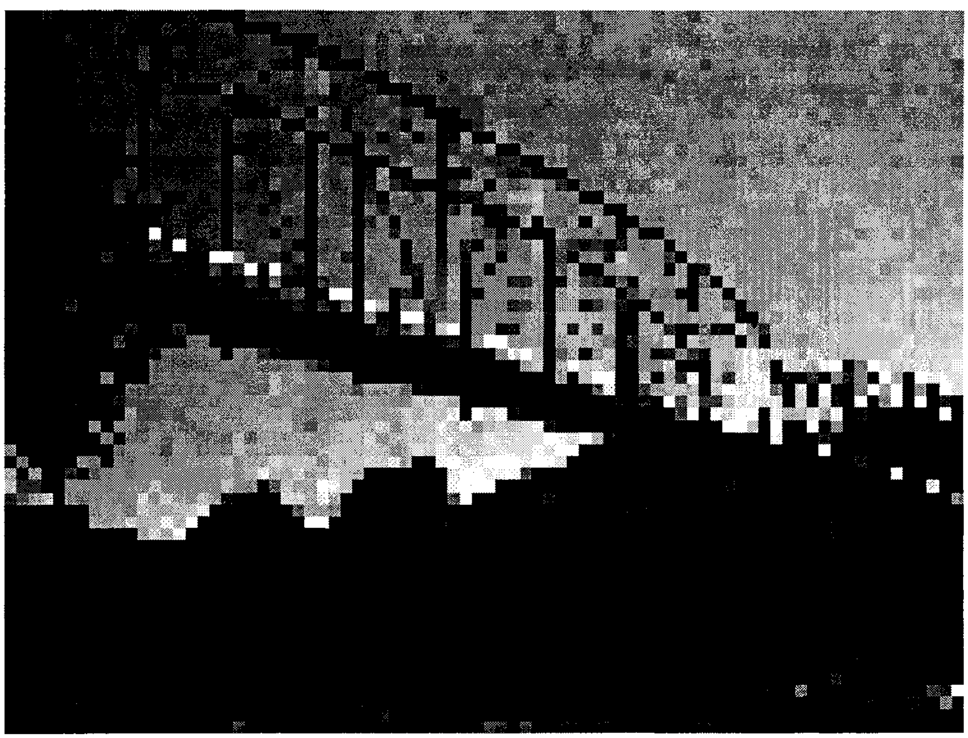


Figure 1. The pixel numbering convention.



a



b

Figure 2. Image b has much lower resolution because of the smaller number of pixels used.

The goal of the digitization process is to break the continuous-tone image into a large enough number of discrete pixels that the digitized image contains the same information that the original contains. The digitized image should look the same to a human observer as the original continuous-tone image. However, if the digitized image is to be analyzed by a computer, it may be necessary to have a larger or smaller number of pixels than that required for use by human vision. The question how finely an image should be sampled is thus important, and the answer given should be dependent on the purpose of the digitized image. There is no one correct sampling rate. Understanding the concept of *spatial frequency* can help determine how finely an image should be sampled.

All images are made up of transitions in brightness. To avoid confusion, it is important to remember that in the world of image processing, the word *brightness* refers to a tone, so it is possible to have dark brightnesses as well as light brightnesses. The rate of transition in brightnesses--from light to dark and back to light--is the spatial frequency (see Figure 3). The more frequently there is a change from light to dark and back to light, the higher the spatial frequency. Figure 4 shows an image that contains more than one spatial frequency. The area of the open sky has a very low spatial frequency, whereas the area of the steel superstructure of the bridge is an example of an area with a much faster rate of change in brightnesses. The leaves on the trees have an even higher spatial frequency. In order to adequately capture all of an image's spatial detail, it is necessary to sample it at a rate twice that of its highest spatial frequency. Consequently, to capture an image's finest detail, sampling must occur at a rate that allows at least two samples to fall on each detail. This ensures that both the light and dark portions of the detail are sampled. If the sampling rate is lower than twice the rate of the highest spatial frequency of an image, then some of the frequency details will be missed, and the digital image will appear to have lower spatial resolution than the original image because there will not be enough pixels to represent all of the original image's spatial detail. Figure 2 shows the results of inadequate sampling rates. Any inadequacies in resolution caused by undersampling become part of the digital image; no image-processing technique can remove them. Oversampling is not a problem; in practice, it simply leads to the existence of superfluous pixels.

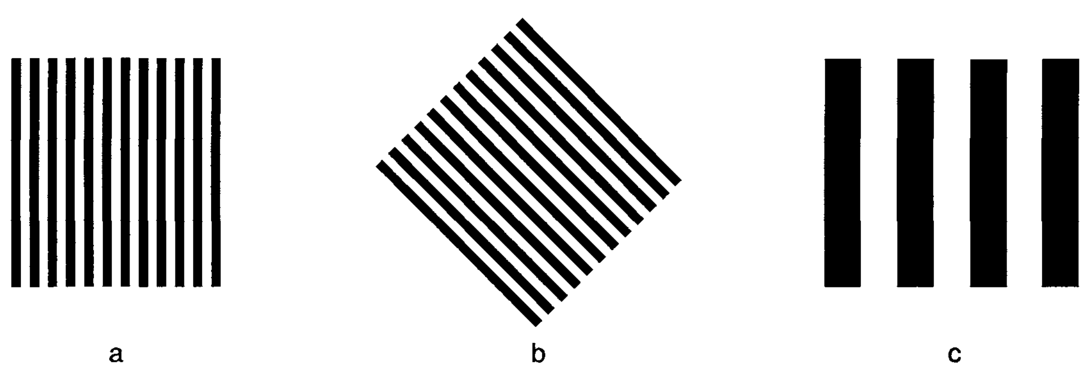


Figure 3. Though a and b have a different angular orientation, they have the same spatial frequency, which is different from the spatial frequency of c.

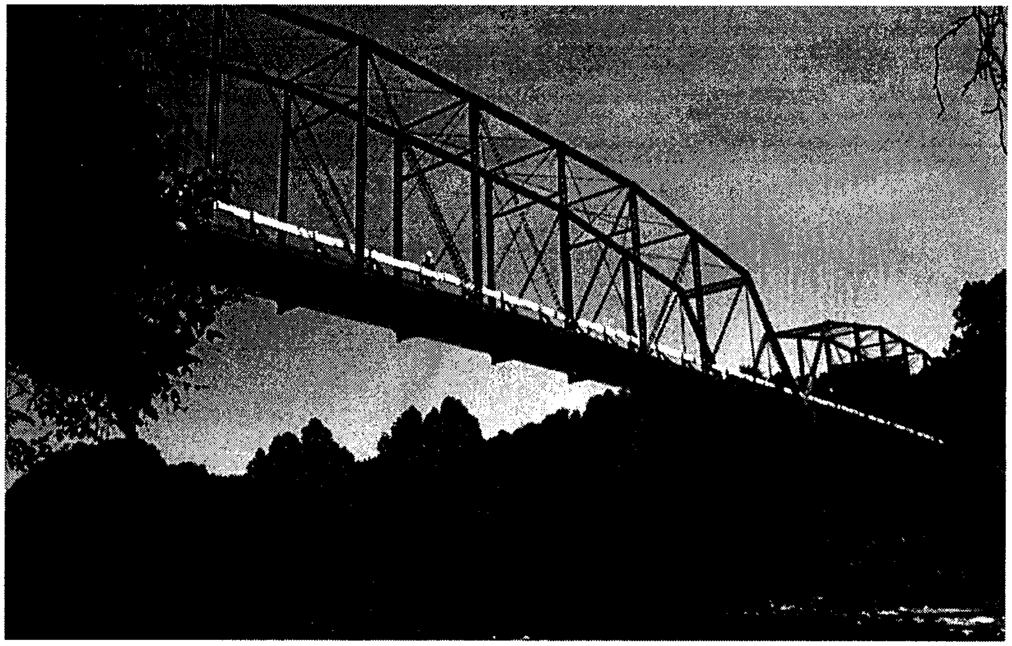
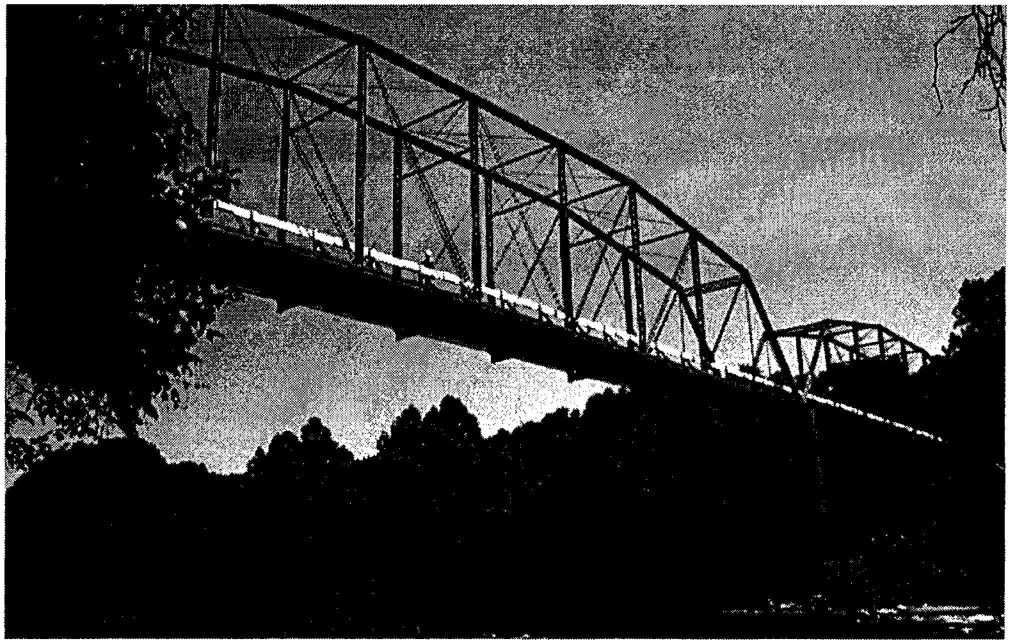


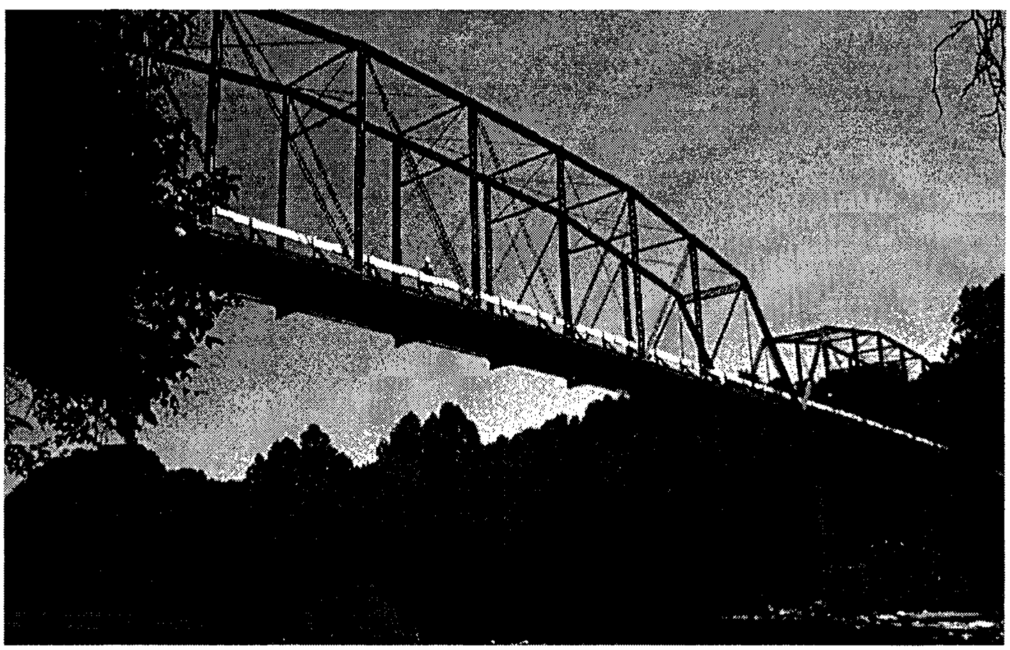
Figure 4. An image containing different spatial frequencies.

Brightness Resolution

The term *brightness resolution* refers to how accurately a pixel's brightness represents the intensity of an original continuous-tone image. A digital image is supposed to represent the intensity of a continuous-tone image, and every pixel is supposed to represent the intensity of the spatial location where it was sampled. Before going further, it would perhaps be useful to clarify the distinction between intensity and brightness. The term *intensity* refers to the amount of light energy actually transmitted from a physical scene. The quantization process converts the intensity of the continuous-tone image at the sampling point to a digital brightness value. The term *brightness* refers to the measured intensity at the sampling point *after it has been sampled, quantized, and displayed*, so it is pixels that have a brightness property. The accuracy of the brightness value is dependent on the number of bits used in the quantizer: 3 bits provide 8 gray levels, 4 bits provide 16, 5 bits provide 32, and 8 bits provide 256, etc. The smoothness of the transition from one gray level to another increases when more bits are used to represent brightnesses. The larger the range of the gray scale, the higher the brightness resolution. When there are not enough gray levels to accurately represent the brightness of the original image adequately, brightness contouring occurs. Brightness contouring is also known as *posterization*. Figure 5 shows an image at different brightness resolutions. Posterization is most obvious in areas in which there are fine gradations in levels of gray, and in this image, it is most apparent in the sky in 5b.



a



b

Figure 5. Image b has lower brightness resolution than image a, which results in posterization, which is especially noticeable in the sky.

Part 2: Image Enhancement

There are two basic types of digital processing: (1) pixel point processing and (2) pixel group processing. But before going into detailed accounts of them, it is necessary to discuss brightness histograms, because they will play a significant role in each type of processing. A brightness histogram displays the overall brightness characteristics of an image. It can therefore serve to indicate an image's strengths and weaknesses. A brightness histogram is a graph that shows the gray levels of all of the pixels in an image, and it shows how many pixels are at a particular gray level (see Figure 6). The scale of brightnesses runs along the horizontal axis, and the number of pixels is displayed along the vertical axis. The horizontal scale in the histogram in Figure 6 has 256 levels of gray (0 plus 255), so it is an 8-bit gray scale. The left end of the scale (0) is black, and the right end (255) is white. Since every pixel has a determinate gray level, a brightness histogram shows how many pixels are at each gray level, and this can be a very useful preliminary to beginning any processing. For example, from a brightness histogram, it is possible to tell whether an image is dark or light or whether it is low or high contrast. The histogram in Figure 6 reflects the fact that the image in Figure 6 has low contrast. This can be seen from the fact that all of the pixels are grouped in a narrow range on the brightness scale. Figure 7 shows the histogram for a high-contrast image. Rather than being clumped close together, the pixels are peaked at each end of the brightness scale. When applied to digital images, the term *dynamic range* refers to the range of gray levels used. An image with a small dynamic range uses only a small portion of the available gray levels. An image of this type has low brightness resolution and low contrast. The image in Figure 6 has a small dynamic range, and the image in Figure 7 has a large one. A large dynamic range indicates an image with good contrast except in cases in which the pixels are peaked at each end of the scale as in Figure 7.

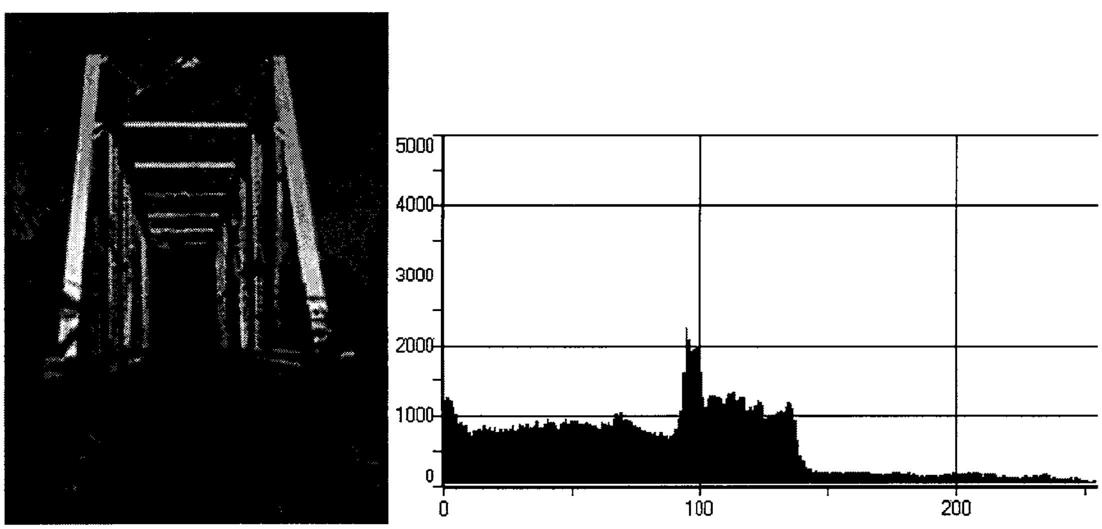


Figure 6. A sample histogram.

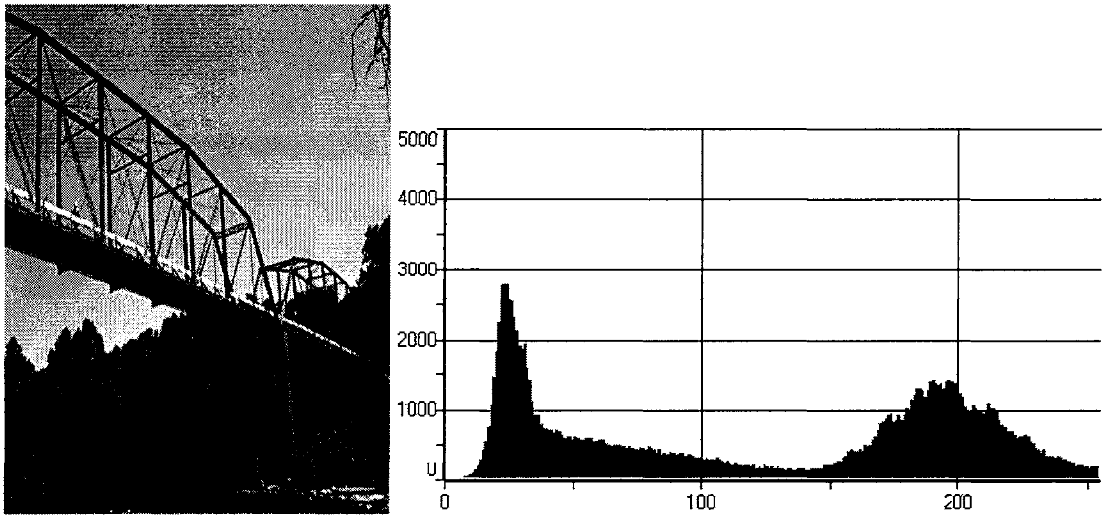


Figure 7. A high-contrast histogram.

Pixel Point Processing

Point processes are the most fundamental of all image processing techniques. With these processes, the gray level of each pixel is modified in some way. If we think of the original image as the *input image*, then the gray level of each pixel in the input image is changed using mathematical or logical techniques to a new value and placed in the *output image* at the same spatial location. In other words, the brightness value of each pixel in the input image is converted to another brightness value. Once the conversion is done, we will have an image (the output image) that is altered in some way that will be useful to the image processing or image analysis that we wish to undertake.

The most straightforward pixel point process is the complement operation, and the most common example of two images that are complements of each other is that of a photographic negative and the print (or positive) made from it. If the complement operation is applied to a digital image, each pixel's brightness value is changed to its complement. If the input image is a positive image, then the output image will be a negative. For example, if a pixel's input gray level is 0 (black), then that pixel will have a gray level of 255 (white) in the output image. A pixel with a gray level of 128 (which is right in the middle of the range) will remain the same. Figure 8 shows an image and its complement.

Point processes are most commonly used to enhance the contrast of an image. This usually involves the addition of a constant value to every pixel, the subtraction of a constant value from every pixel, or the multiplication or division of a pixel's value by a constant value. Two of the most common of these processes are called *histogram sliding* and *histogram stretching*. Both of these processes enhance the contrast of an image by redistributing brightnesses. Figure 9 presents a low-contrast image and its histogram in which the gray scale distribution has a narrow dynamic range. The sliding operation is nothing more than adding or subtracting a constant brightness to all pixels in the image. The amount added or subtracted is limited only by the scale, but in either case, the effect is simply to slide the range toward the black (0) end of the scale or the white (255) end. Figure 10 shows the results of subtracting 70 gray levels from all of the pixels in the image. At this point in the operation, the dynamic range has not increased. The occupied range of gray levels can be stretched by multiplying all of the brightness values by 3, which will put the lighter end of the range close to the limit of 255. The results of the stretch can be seen in Figure 11. The contrast of the image has been greatly improved, and this is shown by the histogram and by the improved appearance of the image. However, it must be remembered that whether an image's contrast is good or bad is dependent on its intended use. In some cases, it is the intent of the image processing to improve the appearance of the image to the human eye; however, in other cases, if the image analysis is to be made as easy as possible, it may be necessary to modify the image in such a way that it appears less attractive to the eye.

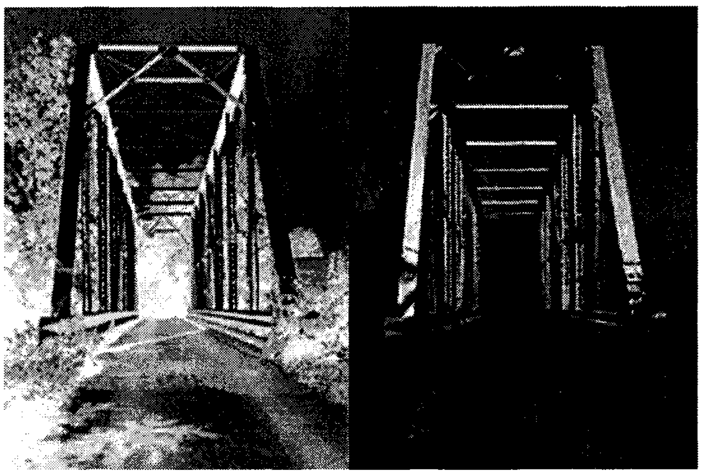


Figure 8. An image and its complement.

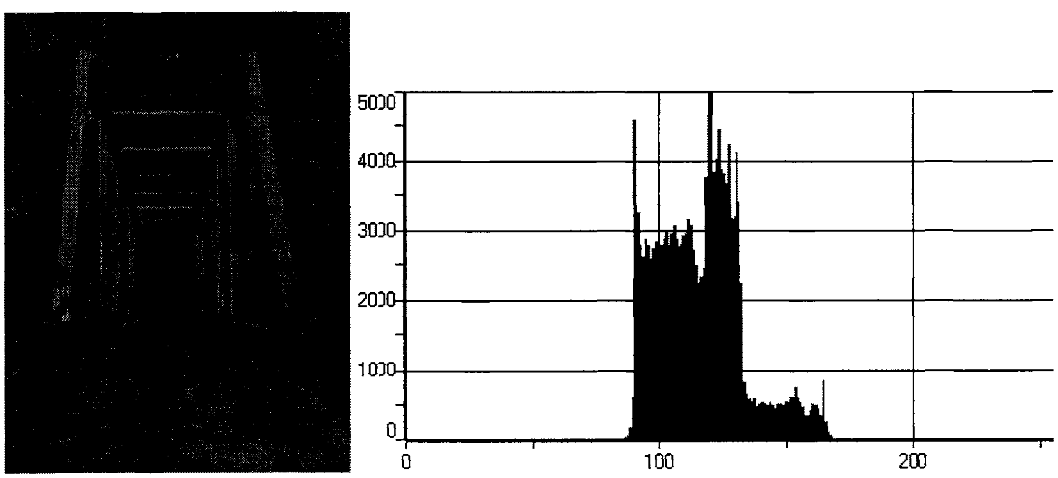


Figure 9. A low contrast image and histogram.

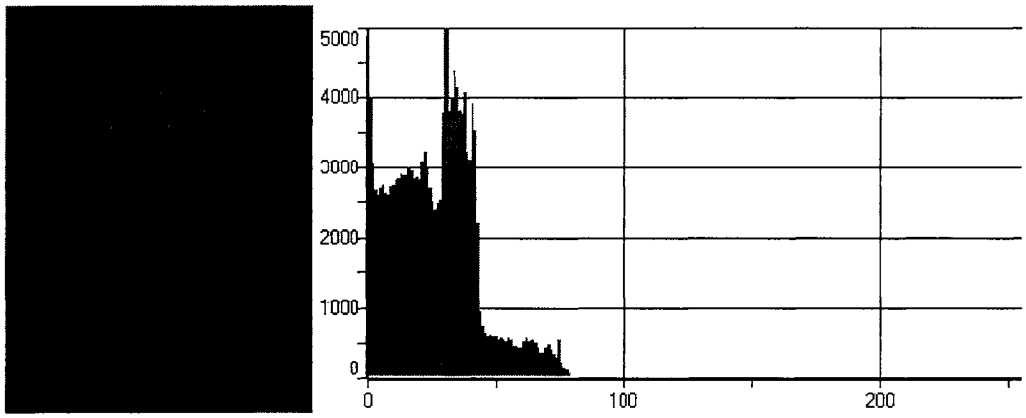


Figure 10. An image and histogram after the slide.

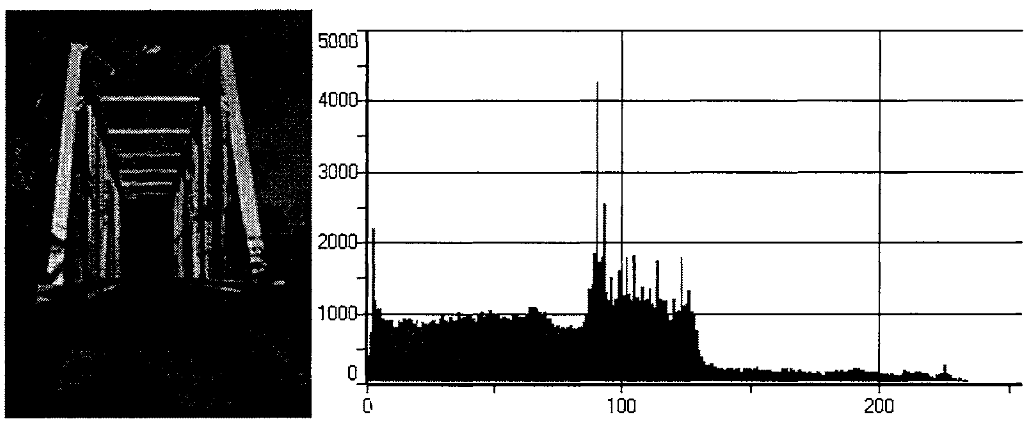


Figure 11. An image and histogram after the stretch.

Binary contrast enhancement, which is also known as *thresholding*, is another operation that can be applied to a low-contrast image. In an image in which there is little difference in gray level between the object of interest and the background, it is sometimes possible to set a threshold level that will separate the object of interest from the background so that it will be possible (or simply easier) to perform subsequent image processing or image analysis. It is not the goal of this operation to improve the appearance of the image. After viewing the histogram for the input image, a threshold level is chosen. Everything below and equal to the threshold level will be given the brightness value of 0 (black), and everything above the threshold level will be given the value 255 (white). Thus the output image will be a binary image: everything will either be black or white. The difficulty with this kind of operation is picking the best threshold level. It is obviously the case that the object of interest cannot contain any gray levels below the threshold level, because these gray levels will become part of the background. Figure 12 shows a low-contrast image and its histogram. Figure 13 shows the same figure after a thresholding operation has taken place.

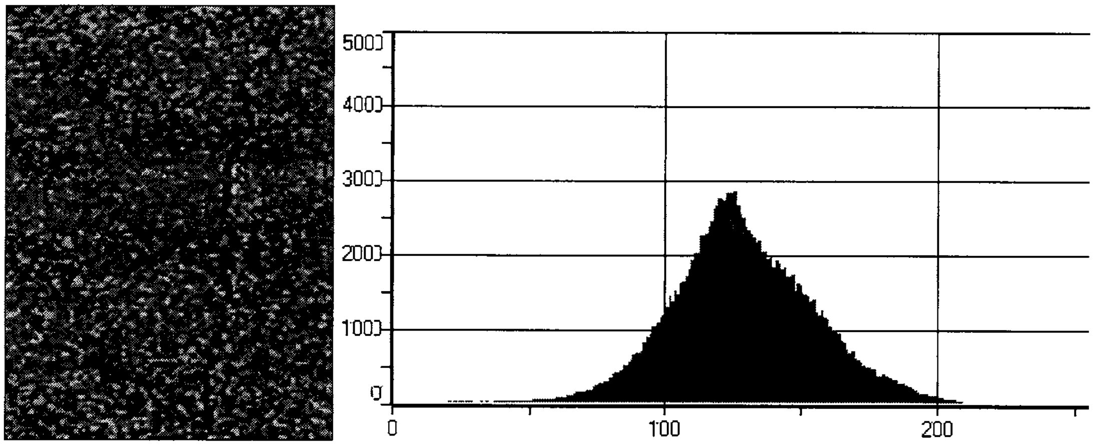


Figure 12. A low-contrast image of a crack in concrete pavement and its histogram.

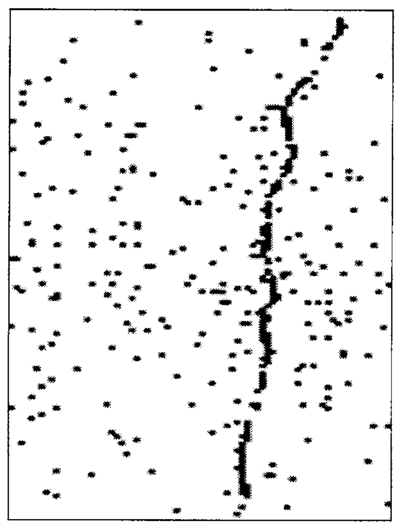


Figure 13. The image in Figure 12 after application of a thresholding operation.

Pixel Group Processing

As previously discussed, the areas of an image in which there are rapid transitions in brightness have high spatial frequencies, whereas the areas that contain slowly changing transitions in brightness have low spatial frequencies. The areas with the highest frequencies are those in which there are points or edges in which the transition from light to dark takes place within a distance of one or two pixels. Spatial frequencies are thus an indication of the frequency of *transitions* in brightness. An image can be filtered to alter its spatial frequencies. Certain frequencies can be removed altogether. On the other hand, there are other spatial filtering operations that serve to accentuate edges or points. These operations are called *edge-enhancement operations.* In order to alter the spatial details of an image it is necessary to take into account more than one pixel at a time. In point processing, the gray level or the brightness value of each pixel in the input image is processed without taking into account in any way the brightness values of the pixels adjacent to the input pixel. Unlike point processing, group processing takes the neighboring pixels into account. Group processing incorporates a process called *spatial convolution* to provide an account of the spatial frequencies in the neighborhood of the input pixel, and this makes it possible to perform spatial filtering. The brightness of each pixel in the output image is the product of the calculation of the weighted average of the input pixel and its neighbors.

The input pixel and its neighbors are called a *kernel.* Kernels are usually square. Although the dimensions of the squares vary widely for different types of operations, they commonly have an odd number of pixels in each direction: 3 x 3, 5 x 5, etc. A 3 x 3 kernel looks like this:

a	b	c
d	e	f
g	h	i

The letter e represents the input pixel, and the rest of the letters represent neighboring pixels. The brightness values of the adjacent pixels will have an effect on the brightness value of the output pixel. The output brightness value is a product of a weighted average calculation, which involves the summation of pixel brightnesses in the kernel multiplied by the weights provided by a particular *convolution mask.* The convolution mask is an array of convolution coefficients, which differ depending on the spatial filtering operation desired. The following is a common convolution mask for a high-pass filter:

-1	-1	-1
-1	9	-1
-1	-1	-1

Imagine an arbitrary array of pixels from an input image:

a	b	c	d	e
f	g	h	i	j
k	l	m	n	o
p	q	r	s	t
u	v	w	x	y

Imagine the pixel designated as g to be the first input pixel. The pixel g and all of its immediate neighbors in the input image (a, b, c, f, h, k, l, and m) have a determinate brightness value (between 0 and 255). When, for example, a high-pass mask (as above) is applied to this input pixel, each of the neighboring pixel brightness values will be multiplied by -1, and the input pixel will be multiplied by 9. The sum of these values will be the brightness value for the output pixel g. The mask will then move on to h, for example, and the brightness value for the output pixel h will be a sum of the multiplicands of the input pixel h and its neighboring pixels. This process continues until every pixel in the image or in an area of interest is processed. The pixels on the edges of the image or the area of interest are treated in a variety of ways to compensate for the absence of neighboring pixels beyond the edge of the image or beyond the edge of the area of interest.

High-Pass Filters

High-pass filters accentuate high-frequency spatial components. The high-pass mask mentioned above is a common one:

$$\begin{matrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{matrix}$$

The coefficients add up to 1. This shows that in areas of constant brightness, this filter will have no effect. Figure 14 shows an imaginary array of pixels in which the input pixel (I) has a brightness value of 200 as do all of the neighboring pixels in the kernel; consequently, when the high-pass filter is applied, the output pixel (O) will have the same brightness value (1800 - 1600 = 200).

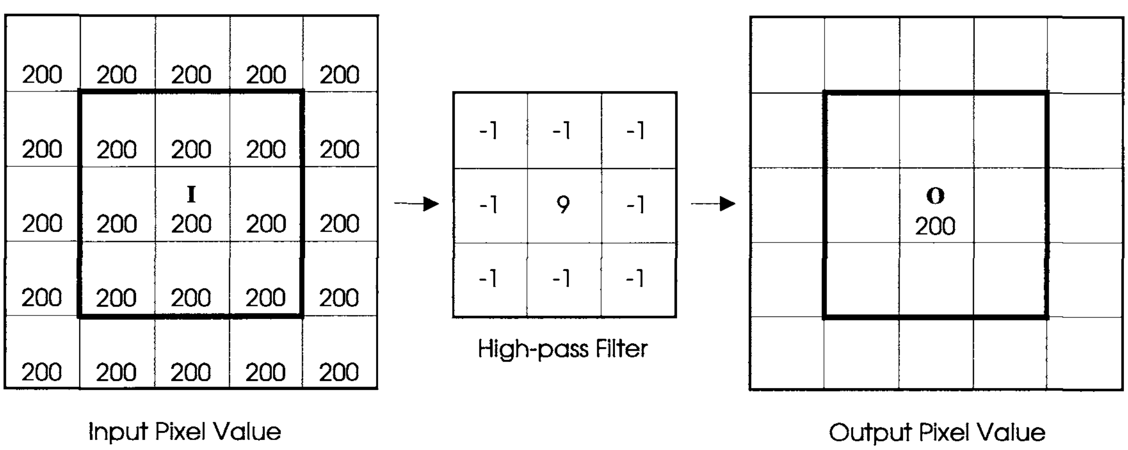


Figure 14. The application of a high-pass filter in an area of constant brightness,

If there is a large difference in the brightness values of the input pixel and the neighboring pixels, the high-pass filter will, in effect, accentuate the input pixel. Figure 15 shows the result of applying the high-pass filter to an area containing high-frequency components. It is possible for the value of the output pixel to be lower than 0 or higher than 255 when this operation is applied. If an output pixel has a value below 0 or above 255, its value is set at 0 or 255, whichever is the closest to its calculated value. In Figure 15, the output pixel has a value of -700 (450 - 1150), so it is set at 0 (black), thereby accentuating the difference in brightness between it and the surrounding pixels. Figure 16 shows the same array, but with the filter applied to a different input pixel. The value of the output pixel is 500 (1800 - 1300), so it is set at 255. Again, the effect is to accentuate the difference between the input pixel and the neighboring pixels that have significantly different brightness values. Figure 17 shows the effect of applying a high-pass filter to an image.

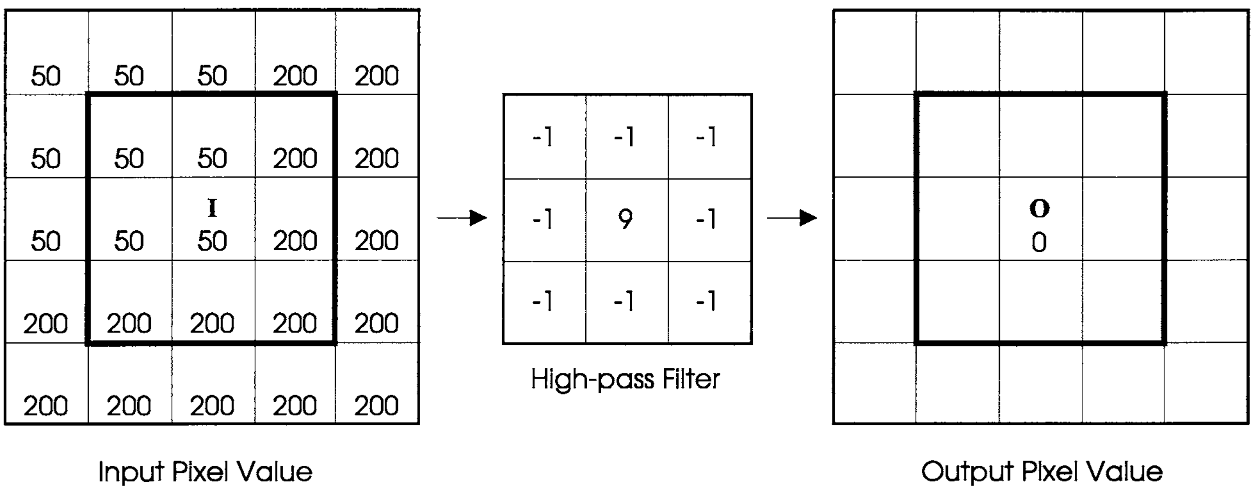


Figure 15. The application of a high-pass filter in an area containing high-frequency components.

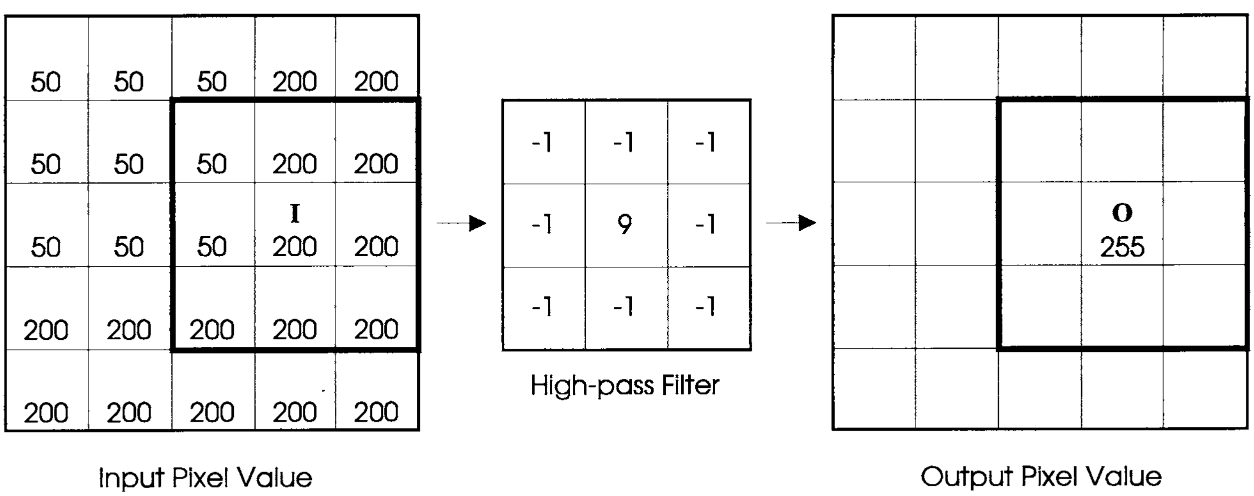


Figure 16. The application of a high-pass filter to a different kernel.

Input Array

50	50	50	50	200	200	200	200
50	50	50	50	200	200	200	200
50	50	50	50	200	200	200	200
50	50	50	50	200	200	200	200
50	50	50	50	200	200	200	200
50	50	50	50	200	200	200	200
50	50	50	50	200	200	200	200

-1	-1	-1
-1	9	-1
-1	-1	-1

Output Array

	50	50	0	255	200	200	
	50	50	0	255	200	200	
	50	50	0	255	200	200	
	50	50	0	255	200	200	
	50	50	0	255	200	200	

Figure 17. The application of a high-pass filter.

Edge-Enhancement Filters

Edge-enhancement filters are spatial filters. Like high-pass filters, they are implemented using group processes. The effect of their application is to show only the edge details in an image. These edge outlines are used to facilitate the recognition of features or objects. There are many edge-enhancement filters; however, this discussion will be limited to the *Laplacian edge filter*. In order to understand edge-enhancement filters, it is first necessary to understand the concept of *slope*. As it is used in image processing, it refers to differences in brightness values. A large difference in brightness between two adjacent pixels corresponds to a steep slope, and a small difference corresponds to a small slope. In a digital image, an edge is an abrupt change in brightness; consequently, a steep slope is an indication of an edge.

Some edge-enhancement filters show only a portion of the edges in an image. For example, the *shift and difference* edge-enhancement operation can be used to enhance horizontal edges or vertical edges. A Laplacian edge enhancement enhances all edges in an image. Here are three common Laplacian masks:

$$\begin{matrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{matrix} \qquad \begin{matrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{matrix} \qquad \begin{matrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{matrix}$$

The first of these is very similar to one of the masks used for a high-pass filter; however, in this case, the coefficients add up to 0. This means that in an area of constant brightness (that is, an area with 0 slope) the output will be 0 (black). Another way of putting this would be to say that in all areas of an image in which there are no steep slopes, all pixels will be converted to black. Figure 18 shows an array of pixels with their brightness values indicated. On the far left and the far right are areas of constant brightness. Moving from left to right, there is a gradual increase in brightness and then a steep increase to a brightness level of 200. After the Laplacian filter is applied, all of the pixels in the areas of constant brightness and in the areas of slowly changing brightnesses are changed to black. At the point of the steep slope, the pixels have a value of 225, which is close to white. The Laplacian filter was applied only to the pixels inside the bold boundary. The outer perimeter of pixels is there only to provide a full kernel to work with for the pixels that are on the outside border of the area marked out by the boundary. The values in the output array are the product of the application of the filter and the convention that places all values calculated to be below 0 and above 255 at 0 and 255, respectively. For example, for the input pixels that have a brightness value of 50, the calculation would be 400 - 475 = -75 = 0. Figure 19 shows the results of the application of a Laplacian filter to an image of cracked concrete.

Input Array

50	50	75	100	125	200	200	200
50	50	75	100	125	200	200	200
50	50	75	100	125	200	200	200
50	50	75	100	125	200	200	200
50	50	75	100	125	200	200	200
50	50	75	100	125	200	200	200
50	50	75	100	125	200	200	200

-1	-1	-1
-1	8	-1
-1	-1	-1

Output Array

	0	0	0	0	225	0	
	0	0	0	0	225	0	
	0	0	0	0	225	0	
	0	0	0	0	225	0	
	0	0	0	0	225	0	

Figure 18. The results of applying a Laplacian filter.

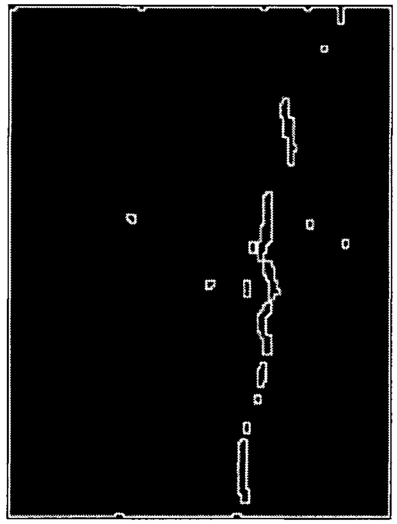


Figure 19. A Laplacian filter applied to the image of cracked concrete that appeared in Figure 12.

Part 3: Image Analysis

Unlike image processing, the results of image analysis are not usually pictorial. The goal of image analysis is to understand an image, and this means measuring and classifying its elements. Image analysis involves three types of operations: (1) segmentation, (2) feature extraction, and (3) classification.

Segmentation

"The first step in any image analysis endeavor is to simplify the image, reducing it to its basic components" (Baxes 1994). Any operation that highlights or isolates the individual objects within an image is a form of segmentation. There are three stages of segmentation: (1) preprocessing, (2) initial object discrimination, and (3) object boundary cleanup.

The preprocessing of the image and the initial discrimination of objects are both preparatory steps to analysis, and, to that extent, the type of operation used must be closely coordinated with the nature of the particular analysis operations for which they are preparation and the goals of these operations. The goal of preprocessing is to simplify an image by removing useless information from it; however, it is important to remember that what counts as useless is dependent on the goal of the analysis operations to be undertaken. Almost any image enhancement operation can be used in preprocessing. Image-enhancement operations are also useful in the initial discrimination of objects. For example, binary contrast enhancement isolates objects of interest from the background, and edge-enhancement operations can be used to highlight the edges of objects. Further clarification of the structure of objects is almost always needed after preprocessing and the initial discrimination of objects. This clarification is done by simplifying the boundaries of objects until they are only a single pixel wide. These simplifications are necessary to ensure that the measurements to be undertaken as part of the analysis of the objects are as accurate as possible. This will further contribute to the accuracy of the classification of these objects. The operations that are used to simplify the boundaries of objects are called *object boundary cleanup operations. Image morphological operations* are a form of boundary cleanup operation. They are often used to clarify an object's structure. The following account of morphological processing focuses exclusively on binary processing; however, morphological processing can also be used on gray scale images.

Every pixel in a binary image is either black or white. Images of this sort are often the product of binary contrast enhancement (see the section of this report on pixel point processing). Binary images can be represented by an array of 0s and 1s. In the material that follows, pixels that have a value of 0 will be black and part of the background, and pixels that have a value of 1 will be white and will represent objects of interest. Because the pixels in these images must have one of two values, it is possible to use logical operators in their processing.

The logical operators are *and*, *or*, and *not*. Figure 20 shows their truth tables. The letters p and q are variables that represent assertions each of which can be true or false independently. With two assertions, there are four possible combinations of truth and falsity, which are represented by the rows of Ts and Fs below the letters p and q. The Ts and Fs under the logical operators represent their truth values. So, for example, the operator *or* is true when one or both of the variable assertions is true. This is, by the way, the nonexclusive *or*; there is an exclusive *or* that is true only when one or the other assertion is true but not when they are both true. The operator *and*, which is a symbol for conjunction, is true only when both assertions are true. If the conjunction had contained more than two conjuncts (for example, p, q, and r), it would still be true only when they were all true. Finally, the operator *not* is true when the assertion that it modifies is false and false when the assertion that it modifies is true.

not	p
F	T
T	F

p	or	q
T	T	T
T	T	F
F	T	T
F	F	F

p	and	q
T	T	T
T	F	F
F	F	T
F	F	F

Figure 20. Truth table for the logical operators.

There is a superficial resemblance between binary morphological processes using logical operators and the spatial convolution processes already described. Both processes are group processes, and both use a mask that moves across the image pixel by pixel. However, spatial convolution processes compute the brightness value of an output pixel by taking into account the spatial frequencies of the input pixel's neighbors, whereas morphological processes combine pixels logically into a resulting output pixel value, which is either 0 (black) or 1 (white).

The two most fundamental morphological operations are *erosion* and *dilation.* Erosion reduces the size of objects in relation to their background, and dilation expands them. Both of these operations can be used to remove noise in an image or to remove the ragged edges of an object. The morphological mask for erosion is configured like this:

$$\begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}$$

Remember, this is a two-valued system, so these are not numerical values, they represent one of two states: white or black. This mask resembles spatial convolution masks; however, in this case, the mask is placed over each input pixel, and the values of the input pixel and its eight neighbors are logically compared to the values of the mask. When the mask values match their respective input pixel values, the output pixel value is set at either 1 or 0, and if there is not a match, then the output value is set at the opposite value. This is sometimes called the *hit-or-miss transform.*

If we have an input kernel of the following form:

$$\begin{matrix} y_3 & y_2 & y_1 \\ y_4 & I & y_0 \\ y_5 & y_6 & y_7 \end{matrix}$$

and a mask of the following form:

$$\begin{matrix} x_3 & x_2 & x_1 \\ x_4 & X & x_0 \\ x_5 & x_6 & x_7 \end{matrix}$$

then the general form of a binary morphological process is: the output pixel $O = 1$, if $X = I$ and $x_0 = y_0$ and $x_1 = y_1$ and $x_2 = y_2$ and $x_3 = y_3$ and $x_4 = y_4$ and $x_5 = y_5$ and $x_6 = y_6$ and $x_7 = y_7$; otherwise, $O = 0$. All of these equalities are mathematical assertions each of which is capable of being true or false independently of the others, and the output pixel has a value of 1 only if this conjunction of them is true, and this conjunction is true only if each of the equalities is true. Whether the output pixel's value is 1 or 0 is a matter of convention as is whether 0 represents black and 1 represents white; however, once it is established that the output pixel will have a value of 1 if the conjunction is true, then if the conjunction is false, the value of the output pixel must be set at 0.

Looking specifically at the erosion operation, we can say that if the input pixel and all of its neighbors are 1s, then there is a match or *hit*, and the output pixel's value will be 1. If the input pixel or any of its neighbors are 0s, then there is a *miss*, and the output pixel's value will be 0. Figure 21 shows the four possible cases. In a, b, and d, the value of the output pixel is the same as the value of the input pixel; however, in c, where the input pixel has a value of 1 and its neighbors are a mix of 1s and 0s, the output pixel's value is 0. The first two cases (a and b) show that when an input pixel is surrounded by pixels of the same value, the value of the output pixel will be the same as the input pixel. So, an input pixel that is a part of the background (0) and is surrounded by background pixels will not change in value. The same is true of a pixel that is part of an object (1). When an input pixel is surrounded by neighboring pixels with a mix of values as in the cases c and d, then this pixel is on an edge. If the input pixel has a value of 0 (black), then the output pixel will retain the same value. If, on the other hand, the input pixel has a value of 1 (white), the output pixel's value will change to 0 (black); consequently, the edge of the object erodes away. Figure 22 shows the effect of the erosion operation. One row of white pixels (1s) on the edge of the object erodes away. Figure 23 shows how the rough edges of an object are removed when an erosion operation is applied. Figure 24 shows the effect of erosion on a binary image.

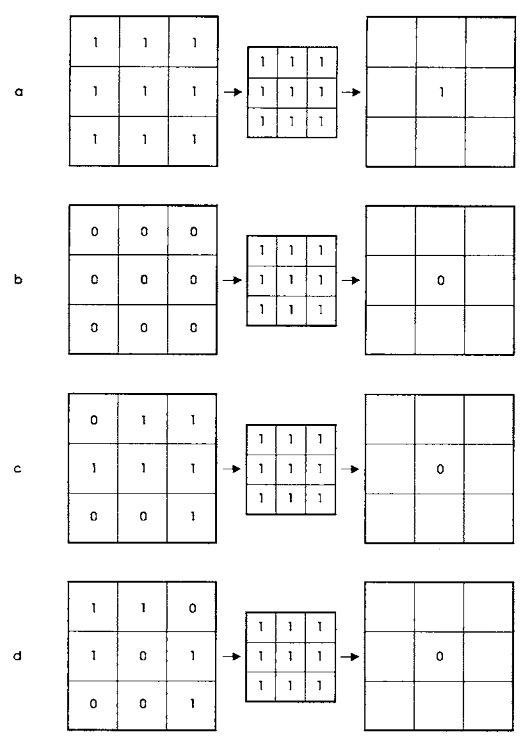


Figure 21. The four input pixel possibilities for binary erosion.

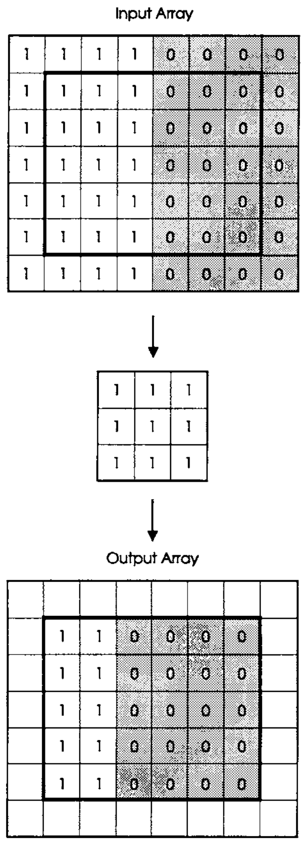


Figure 22. The effect of the erosion operation.

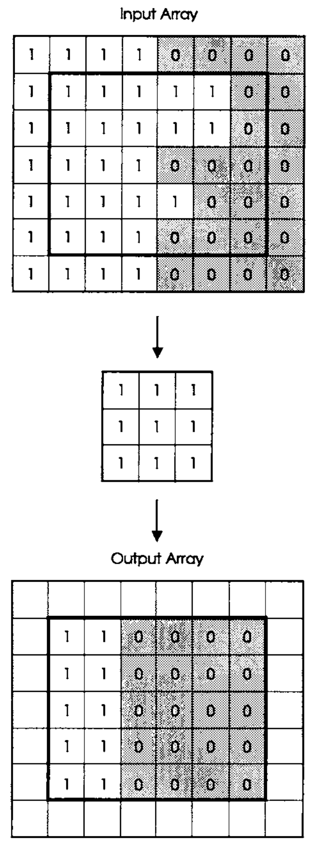


Figure 23. The effect of the erosion operation on the rough edges of an object.

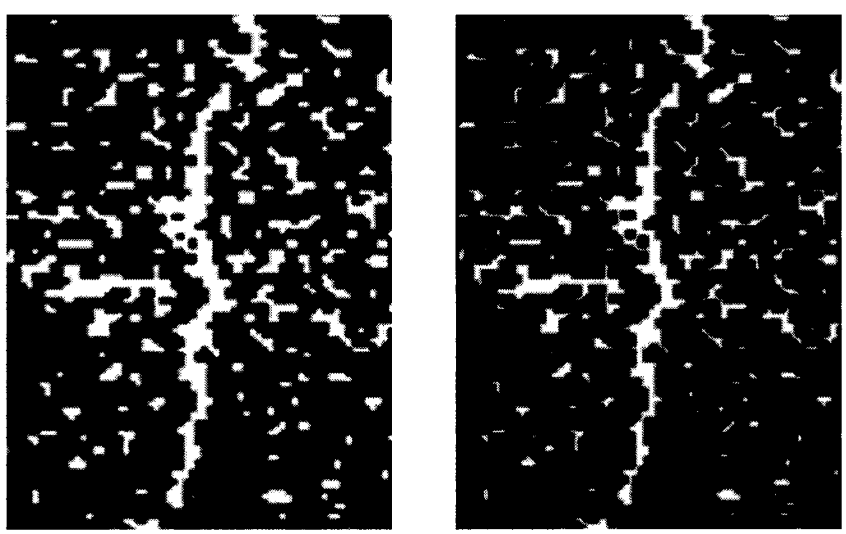


Figure 24. A binary image before and after the application of an erosion operation.

The mask for dilation looks like this:

$$
\begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix}
$$

The effect of this operation is to increase the size of an object in relation to its background. Here, the output pixel has a value of 0 for a hit and 1 for a miss. The four input pixel possibilities are presented in Figure 25. As in the case of the application of the erosion operation, whenever the input pixel and its neighbors have the same value, the output pixel will have the same value as the input pixel (cases a and b). If the input pixel has a value of 1, and it is surrounded by a mixture of 1s and 0s, then the value of the output pixel will remain the same. However, if the input pixel has a value of 0, and it is surrounded by a mixture of 1s and 0s, then the value of the output pixel will be 1, i. e., the object's boundary will increase in size. Figure 26 shows how the size of an object increases by one pixel when the dilation operation is applied. Figure 27 shows how the dilation operation fills in gaps in the edge of objects. Figure 28 shows a binary image before and after the application of the dilation operation.

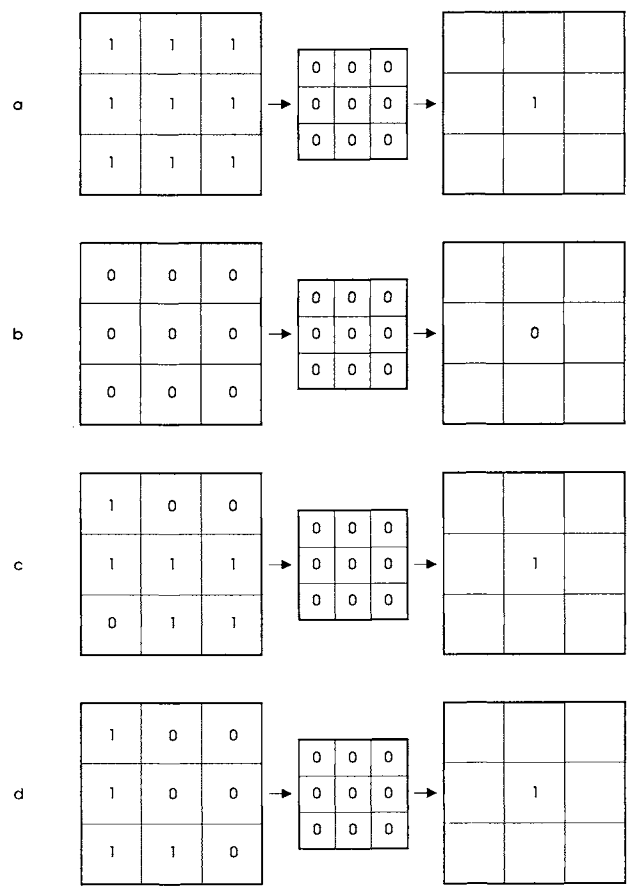


Figure 25. The four pixel possibilities for binary dilation.

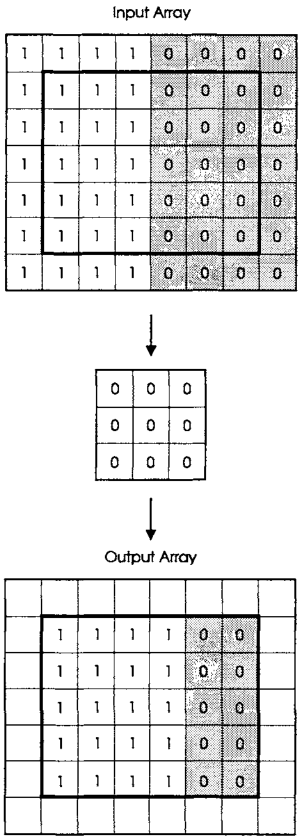


Figure 26. The application of the dilation operation.

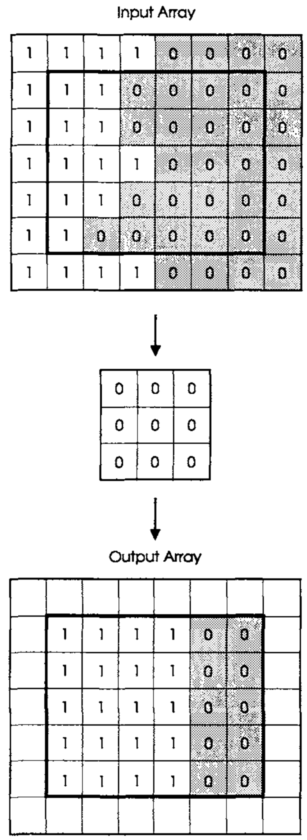


Figure 27. The use of the dilation operation to fill in gaps in the edge of an object.

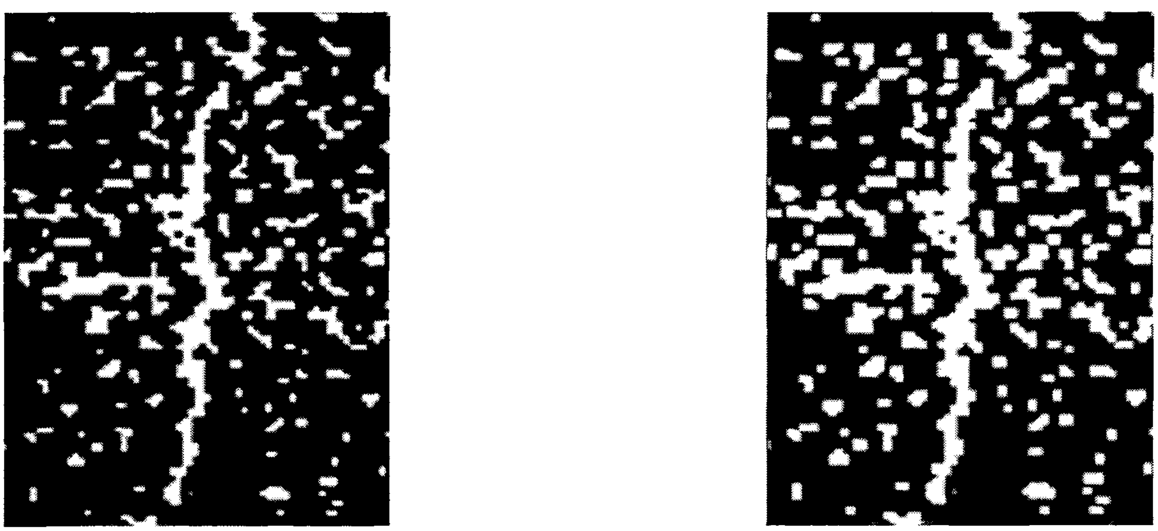


Figure 28. A binary image before and after the application of the dilation operation.

The erosion and dilation operations are used together in an operation called *opening*, which involves the application of erosion first and then dilation. This operation removes small spurs from the border of an object. It also can remove noise from the background of an image. The advantage of this operation over erosion is that it maintains the original size and shape of the object of interest. Figures 29 and 30 show the results of applying the opening operation. In Figure 29, the erosion operation is applied. The spur is removed from the object, and the boundary is reduced in size by one pixel. Figure 30 shows how the dilation operation applied to the results of the erosion operation brings the boundary back to its original size--but without the spur. There is also an operation called *closing*; it is the reverse of opening: dilation first and then erosion. Closing smoothes the boundaries of objects by filling in gaps.

Input Array

1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0
1	1	1	1	1	0	0	0
1	1	1	1	1	0	0	0
1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0

Erosion

Output Array

1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0

Figure 29. The first step in an opening operation.

Input Array

1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0
1	1	1	0	0	0	0	0

Dilation

Output Array

1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0

Figure 30. The second step in an opening operation.

Erosion is also used in a process called *outlining*. The erosion operation is applied to an image as a result of which the object or objects will be reduced in size by one pixel. A dual-image point process is then used to subtract the result of the erosion operation from the original boundary of the object, which leaves a one-pixel-wide boundary, or, in other words, the outline of the original object.

Feature Extraction

Feature extraction follows segmentation. Segmentation operations isolate individual objects within an image. Once the objects are isolated, their relevant features are measured. As always, what features are relevant is dependent on the goals of the analysis. Various kinds of features can be measured including, among others, brightness, texture, and color; however, this account will focus on boundary descriptions which are the most precise way to measure shape.

A boundary (or outline) is a single-pixel-wide sequence of pixels that makes up the perimeter of an object. Figure 31 shows the description of an object boundary. Each pixel in the object's boundary adjoins another pixel. The list of pixel locations provides a specification of the object's boundary. This list could be used to recreate the object, or it could be used to identify the object in an object classification operation. The accuracy with which the sequence of pixels represents the object is contingent on the resolution of the imaging system.

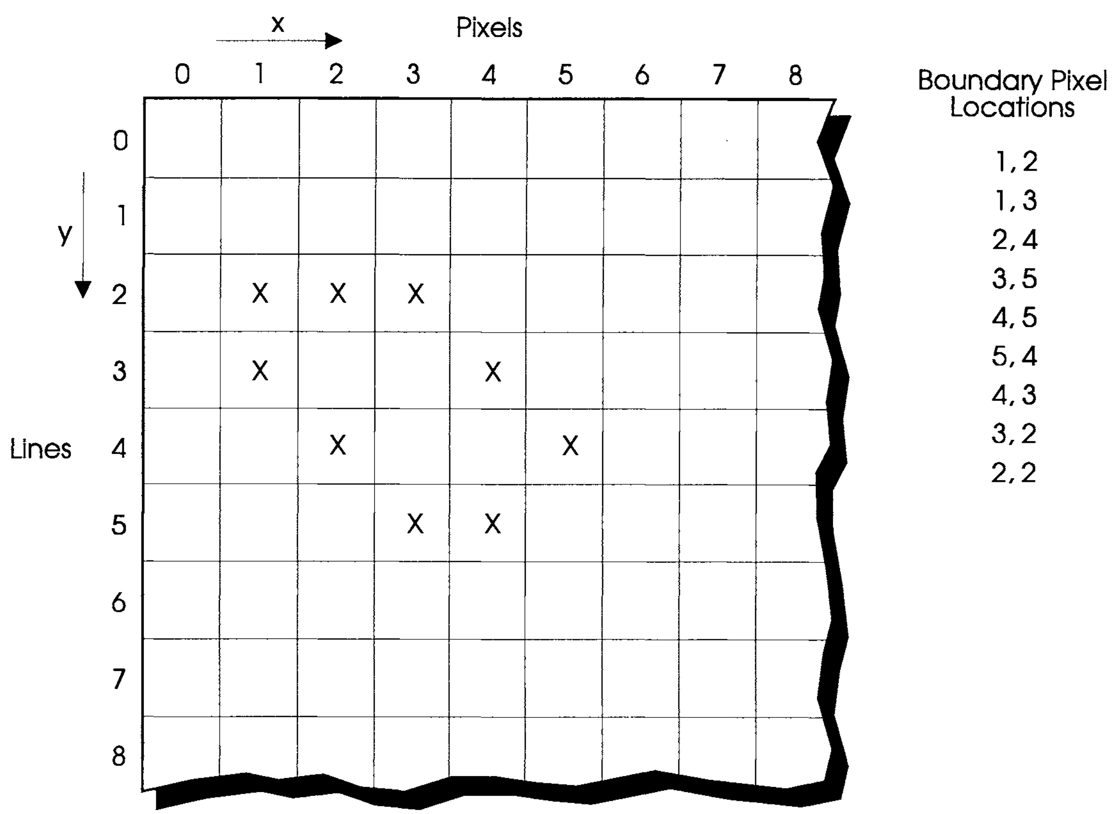


Figure 31. Using sequential pixel locations to describe an object's boundary.

The use of sequential pixel locations as a means of describing an object's boundary is not very useful for classification operations, because no identification would take place unless the object were in the exact same location. For example, if the object in Figure 31 were shifted to the right by one pixel, it would not match the identifying pixel locations even though it is exactly the same shape and size. In order to get around this problem, it is necessary to use *chain codes*, which make it possible to describe an object's boundary without tying it to a particular set of pixels. Any pixel in the object's boundary can be used as a starting point. The pixel adjoining the starting pixel will be in one of eight directions, and each of these directions has a different code. Every pixel in the object's boundary will be given a chain code. Figure 32 shows a convention for assigning chain codes. Figure 33 shows the object in Figure 31 defined by an absolute chain code using the convention shown in Figure 32. The absolute chain code must keep track of the absolute location (1, 2 in Figure 33) of the starting pixel, otherwise, it would not be able to stop at the right time. The absolute chain code is called absolute because the chain-code directions are absolute directions relative to the "0" direction. Like the use of sequential pixel locations for describing the shape of an object, absolute chain codes are sensitive to an object's location and orientation. However, there is a variation of the absolute chain code that is called a *relative chain code*. Relative chain codes are useful for object classification operations, because they define an object's shape in a way that is not dependent on the object's location in the image or on its orientation. The relative chain code differs from the absolute chain code in that the chain-code directions are not absolute. Figure 34 shows how the directional framework reorients itself depending on the direction between two pixels. Figure 35 shows the object in Figure 31 defined by the use of a relative chain code. From the starting pixel, the directional conventions shown in Figure 32 are used. Since the direction is straight down, the code is a 6. However, now that a direction is established, the orientation of the conventional set of directions is altered with each change in direction.

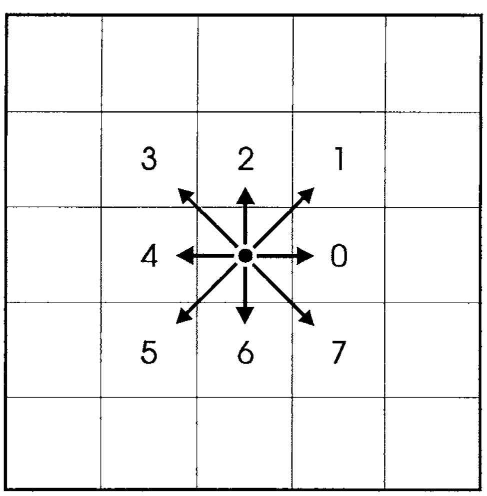


Figure 32. A convention for assigning chain codes.

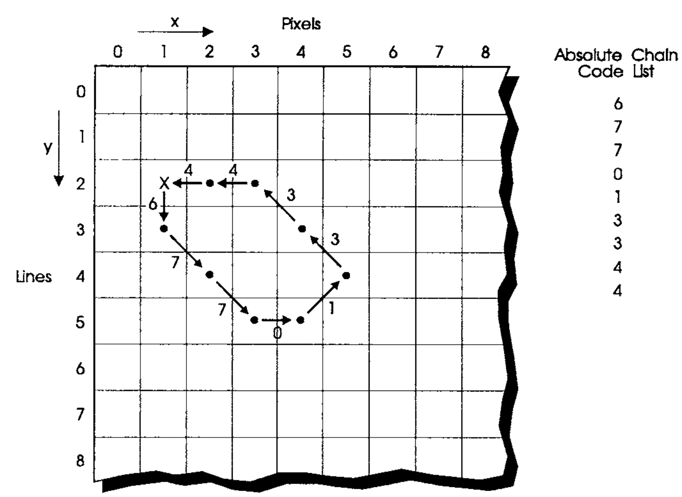


Figure 33. Use of a chain code to designate an object's boundary.

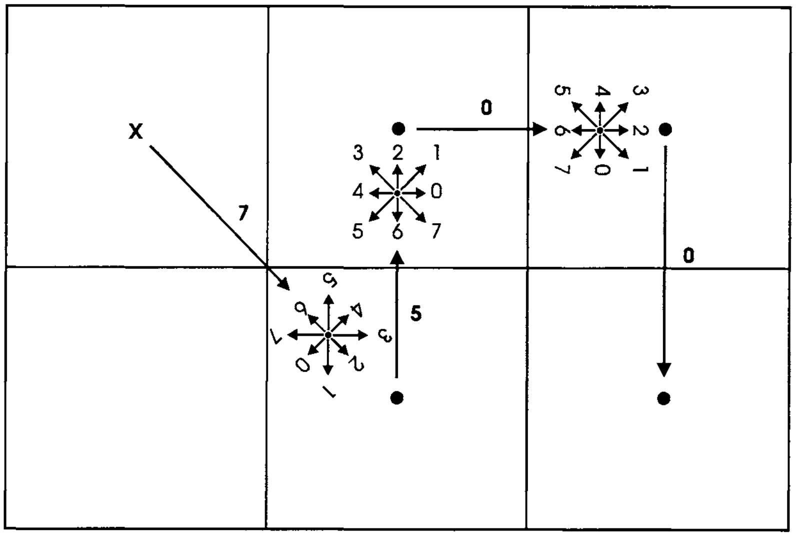


Figure 34. Use of a relative chain code.

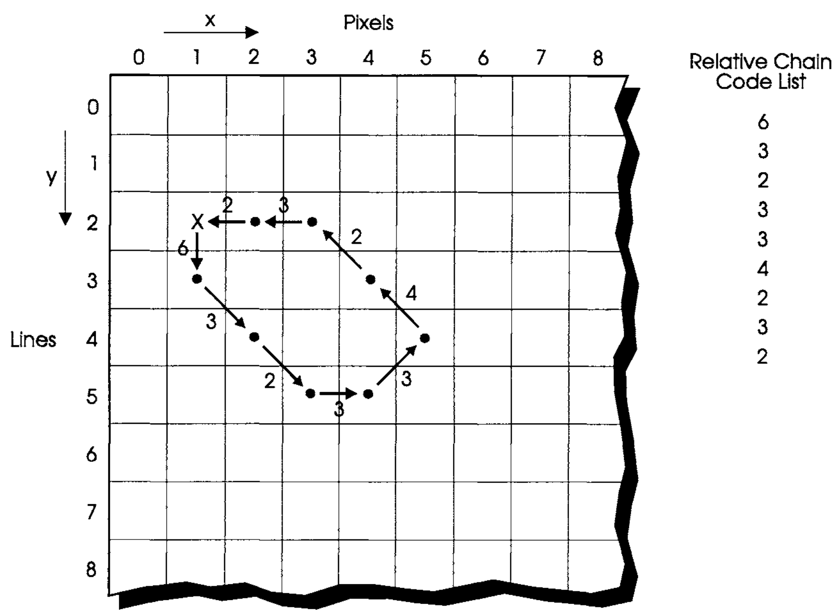


Figure 35. A boundary defined using a relative chain code.

Often, classification operations do not require the designation of every pixel in the boundary of an object. In many cases, it is possible to use a technique called *line segment boundary representation*. This technique simply replaces some of the designations of individual pixel locations with lines. The choice of the length of the lines is crucial in determining how accurately the technique of line segment representation will represent the object: the shorter the line, the more accurate the representation will be. And the required level of accuracy is dependent on the recognition technique to be used. If the goal is to be able to distinguish a 5 cm bolt from an 8 cm bolt on an assembly line, then a precise representation of the entire outline of these items is not necessary. What is needed is simply a representation sufficient for an accurate representation of their length. A variation of the line segment technique is called *variable line segment boundary representation.* The goal of this technique is to represent an object's boundary with as few line segments as is consistent with an accurate representation of the boundary. This technique uses a variety of line lengths to more accurately represent the shape of an object. Figure 36 shows how a judicious use of long line segments can reduce the number of pixels that need to be designated to define a boundary.

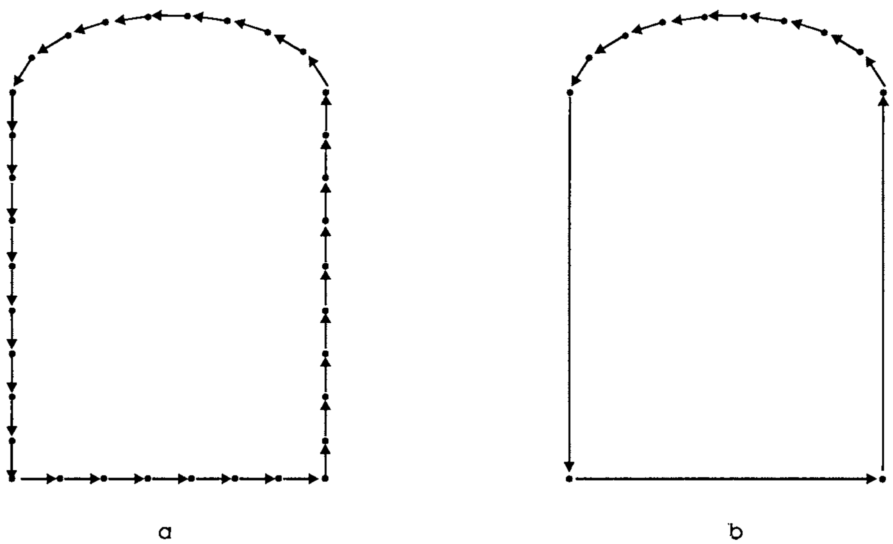


Figure 36. The use of variable line segments to simplify the designation of a boundary.

Object Classification

The first two steps in the image analysis process are (1) the isolation or highlighting of objects in the image (segmentation) and (2) the measurement of features of these highlighted objects (feature extraction). The classification of objects involves the comparison of the measurements of these highlighted objects with the measurements of a known object or with a set of criteria of identification. Generally, this process of comparison involves (1) determining which features of the object are to be used to classify it, (2) determining how close to the established criteria these measurements must be (tolerances), and (3) creating categories (or classes) to which the object will be assigned depending on how its measurements compare with the established criteria.

Classification operations can take a wide variety of forms. The following is an example that comes from Gregory Baxes' book *Digital Image Processing*:

Task: Determine whether the specific shape and size of a molded part match an original sample part to within 10% of the original part, measured using a 10-segment line segment boundary description.

Feature Measures: Line segment boundary description.

Tolerances: Line segments must all be within +/- 10% in the length and the direction angle of the original sample part.

Classes: Good, if the shape is within tolerance to the original part; otherwise, bad.

There are essentially an unlimited number of ways to measure the features of objects, and any of them might prove successful depending on the goal of the classification. If all that is necessary to see whether an object should be classified in a certain way is to determine whether it is round, then it is not, for example, necessary to provide a detailed description of the brightness values of the pixels in the area inside the object's perimeter, nor is it necessary to measure its size. In some cases, it may only be necessary to determine an object's area, in which case, counting the number of pixels will achieve the desired end. In an industrial application, it may be necessary to distinguish two types of objects moving down a conveyer belt. No matter how complex their respective shapes, if one of them is supposed to have a hole in it and the other is not, then it will be possible to distinguish them by using the common shape measure that determines whether an object has a hole in it. In this case, the description of each object's boundary could be greatly simplified. In some cases, there may be a hole in each of the objects, so the size of the hole--or its location--could possibly be used to distinguish them. On the other hand, if the goal of the classification operation is to determine whether a part meets a set of specifications, it may be necessary to determine whether the part is the right shape, whether it is the right size, whether there is a hole in it, whether the hole has the correct shape, whether the hole is in the right place, and so forth.

Part 4: Image Processing Hardware

There are three data-handling functions that any image processing system must be able to perform: (1) image digitization, which is the conversion of the analog signal to a digital form, (2) image storage, which is the short- and long-term storage of the image, and (3) image display, which is the conversion of the digital image back into analog form for display on a monitor. The flow of image data through the system has the following distinct stages (Baxes 1994, 305):

1. The analog video signal enters the system.

2. The analog signal is digitized.

3. The digital image is stored.

4. The digital image is processed by the image data processor.

5. The processed image is converted back into an analog video signal.

6. The analog video signal exits the system for display.

All of the data handling is performed by a frame grabber. This includes every stage except the fourth. The image processing is performed by an image processor.

Image processing is usually undertaken using one of three hardware configurations:

1. *The host computer carries out all of the image processing.* In a system of this sort, there is no way to digitize images, so digitized images must be imported from some other source. Figure 37 shows a simple schema of such a system. The addition of a frame grabber (see Figure 38) makes it possible for the system to acquire images from a video source. Further, more storage is added, and the images can now be displayed on the frame grabber's monitor, thereby leaving the host computer's monitor for use in interacting with the frame grabber.

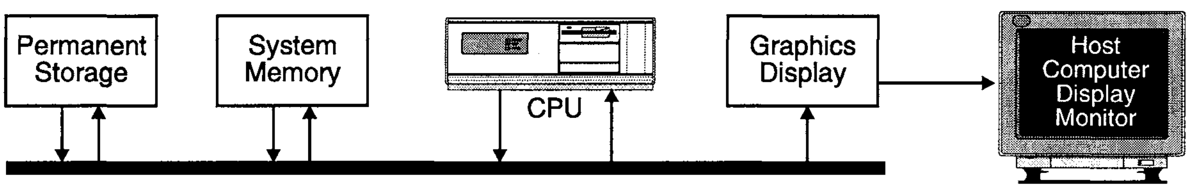


Figure 37. An image processing system using a host computer for all digital processing functions.

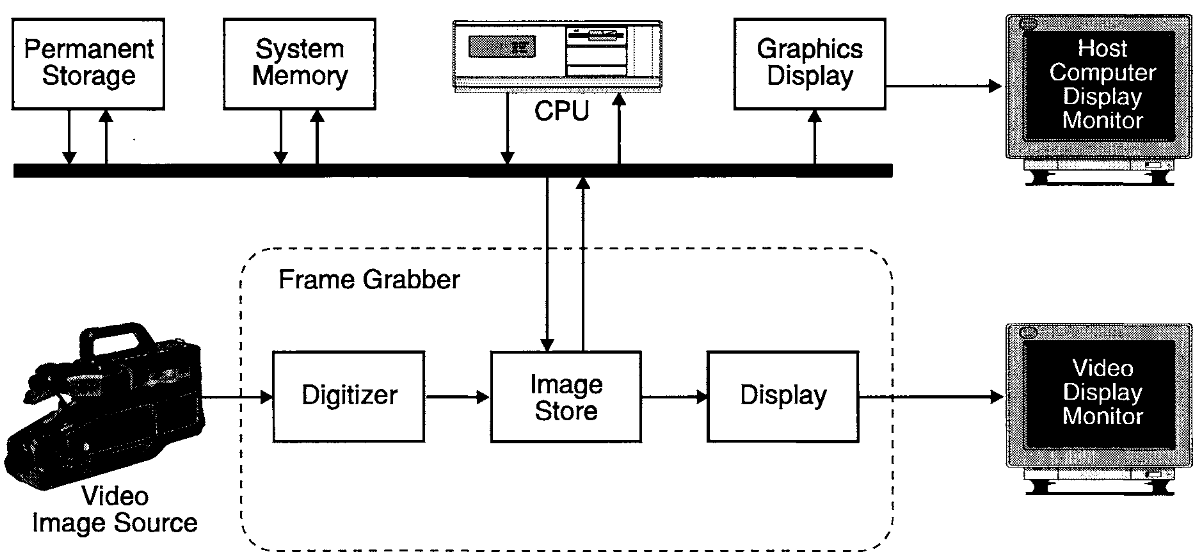


Figure 38. An image processing system that includes a frame grabber.

2. *An internal accelerator processor that has direct access to the image store is used.* An accelerator processor, or a DSP (digital signal processor) as it is often called, is a high-speed microprocessor designed to process digital signals. A DSP works together with a frame grabber (see Figure 39). A DSP accesses image data directly from the image store, and this means that the data does not have to be moved over the host computer's data bus. Further, the host computer no longer has to do any image processing. A DSP downloads software from the host computer to undertake a particular image processing operation. A DSP accesses the image store directly and writes the processed image back into it. The image is displayed on the frame grabber's monitor.

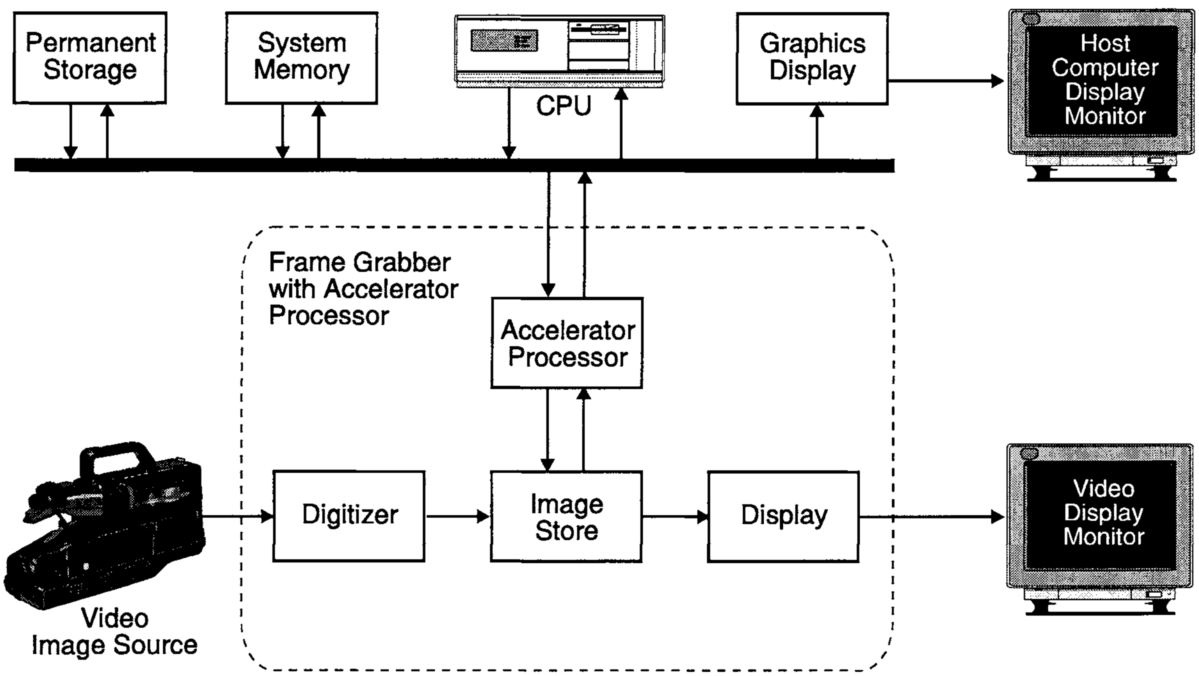


Figure 39. An image processing system that includes an accelerator processor.

3. *Pipelined, special-purpose image processing hardware is used.* Pipeline processing is used when speed is critical or when an image processing operation is used frequently. These processors are usually board-level modules. These modules are configured to perform particular operations that require high speed or that are repeated frequently, which makes real time processing possible. The remaining image processing operations are implemented using software on a host computer or on an accelerator processor. Figure 40 shows a schema of a system that includes a pipelined processor.

One image processing operation that can easily be incorporated in hardware is thresholding. Thresholding converts a gray scale image into a black and white image by setting a threshold brightness value above which all pixel brightness values are converted to white (255) and below which all pixel brightness values are converted to black (0). Thresholding is a pixel point process: the operation is applied one pixel at a time to each pixel in the image. The total range of input values for any pixel in the image is 0 through 255, and the total range of output values is the same. The thresholding operation can be used to calculate the appropriate output value for any possible input value. The result of this series of operations will be a table of output brightness values corresponding to all of the possible input brightness values. These results are then loaded into a look-up table (LUT). Once the LUT is loaded, the thresholding operation takes place by means of a simple procedure of looking up the appropriate output value for each input value rather than calculating it. A LUT can be implemented in the software of a host computer or an accelerator processor where it is stored in the processor's system memory. It can also be implemented in hardware in a pipelined processor circuit, in which case, it is implemented as a 256-location RAM memory device in which each location stores an 8-bit data value. When the LUT is incorporated into hardware in this way, the data can be processed in real time; however, this is the most expensive way to process image data.

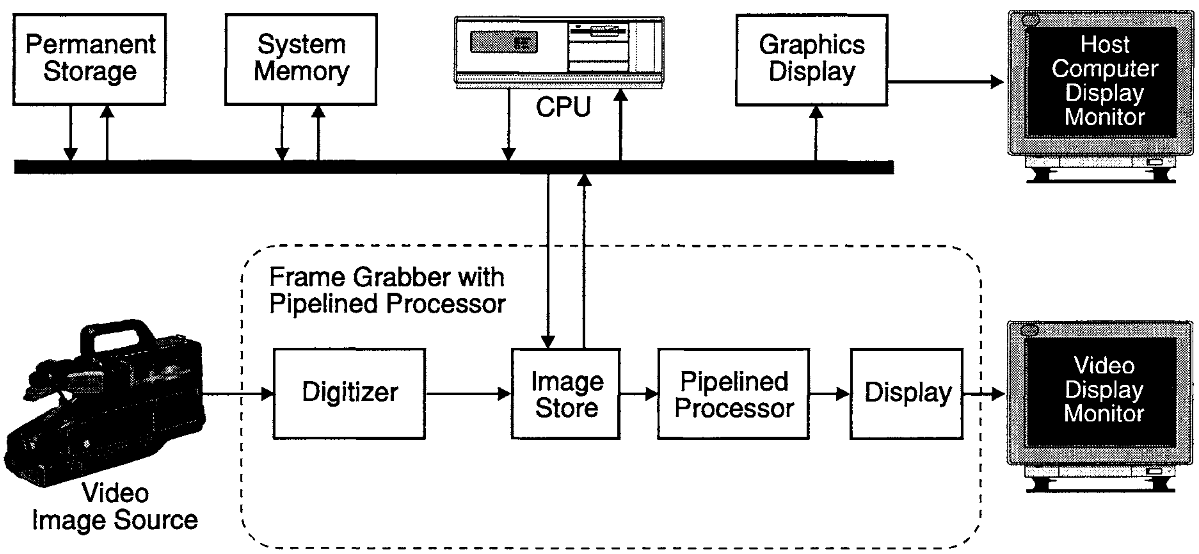


Figure 40. An image processing system that includes a pipelined processor.

Many image processing operations can be implemented in hardware. The thresholding operation and other pixel point processes can be incorporated in circuitry, as a result of which, many of these operations are often included in frame grabbers; however, no matter how computationally complex an image processing operation is, only 256 input values and 256 output values are possible in an 8-bit system. For example, in a dual-image point process, two input pixel brightnesses--one from each image--are operated on to create one output pixel brightness. If a dual-image point process were implemented using a LUT, the LUT would have to have 65,536 locations (256 x 256 possible combinations); however, there would still only be 256 possible output values. Dual-image (or multiple-image) operations as well as all forms of group processing can be implemented in hardware, in which case, they will be implemented as a RAM memory device with the required number of locations.

There are other specialized processors that can be implemented in hardware. Two that are likely to be of use in acquiring and analyzing pavement images are the image compressors (and decompressors) and the histogram generators. Some form of real-time image processing will almost certainly be required on the image acquisition vehicle, because the data rate is so high. One way to reduce the amount of storage needed on the survey vehicle would be to compress the raw data. A real-time histogram generator could prove to be useful in screening raw data for the purpose of eliminating images of pavement containing no distresses, thereby eliminating the need to store data that is of no interest.

Part 5: Algorithms

The development of image processing algorithms for use in pavement distress surveys has not been linear. Despite growing interest in this area during the last 15 years, there is currently no one system of algorithms singled out in the literature as the best way to proceed. The entire enterprise is fraught with complexities because (1) there is a lot of variability in the materials used to construct pavements, (2) cracks have many shapes and alignments, and (3) the presence of tire marks, paint stripes, shadows, oil stains, and debris such as pieces of wire, fan belts, hoses, mufflers, etc., all interfere with image processing operations (Li 1991, 133). However, in part, the difficulty comes from the nature of the enterprise. Michael Burke, who is the author of the three-volume *Handbook of Machine Vision Engineering*, puts the problem of developing a suitable set of algorithms for analyzing pavement distresses in this way (1992, 9):

> A large number of image processing algorithms have been designed for analyzing video images. This is the heart of the problem: there are literally thousands of algorithms available, but many are not suited for one reason or another. . . . Algorithm selection for image processing system design is still very much an art and not a science. The system tends to be highly nonlinear, and so simply reversing the order in which two processes are applied can drastically change the outcome (the analyzed output). Furthermore, each algorithm tends to have a "magic number" or two, parameter values which must be set (such as gain, window size, etc.), and varying these can also change the outcome. Furthermore, varying the parameters of one processing step can change the effectiveness of the settings of a different processing stage (again, because of the nonlinearities involved). We are looking at on the order of fifty such processing steps, all interacting with one another in highly nonlinear and unpredictable ways. *Algorithm selection is therefore a very formidable obstacle.* (emphasis added)

A review of the literature confirms this picture. One factor that contributes to the sense of a lack of linearity in the development of algorithms is the fact that different researchers are able to achieve "good" results even though they use different sets of algorithms. Perhaps this is to be expected. There is certainly nothing in the literature to suggest that there is only one set of algorithms suitable for the solution of a particular problem. As a matter of fact, a perusal of the literature would strongly suggest just the opposite. For all anyone knows, there may be 20--or

200--sets of algorithms that will produce equally accurate results in the detection and classification of cracks in pavements. Some of the systems of algorithms that have been developed have been more successful than others. No system is generally acknowledged to be the best. The development of these systems is an empirical matter: the evaluation of a particular algorithm or set of algorithms involves looking at the results of applying it. And because there are so many variations in the appearance of pavement surfaces, it is essential that the algorithms be tested on extensive sections of actual pavement. Fundakowski et al. tested the algorithms described in their NCHRP report on 3,000 images covering over a lane mile, which is a very large sample in comparison with the sample used in many other studies; however, they conclude that "in order to develop confidence in the methods, the development and testing must move from the processing of individual images to the processing of many lane miles of pavement" (1991, 67). This is an important point. Many of the studies carried out so far have tested their algorithms on an insufficient number of samples in which only three or four types of cracks are represented. Many of these studies also ignore the enormous amount of variation in the appearance of actual pavements and in the appearance of actual cracks. Consequently, it is really not possible to tell how these "successful" algorithms will perform on real pavements. This is not to say that nothing is gained by undertaking these studies. Indeed, some of these systems look promising, and it certainly is the case that the algorithms that are successful in the "lab" will be the ones that get further testing in the field. At some point, however, they must prove themselves on actual pavements, and it is hard to predict how well they will do. Several of the commercial operators offering automated distress survey systems appear to be achieving good results. They are working with proprietary software, which, for obvious reasons, they are unwilling to divulge. However, it appears that these systems are constantly being tested on many kilometers of actual pavement, and--presumably--their algorithms are constantly being modified to deal with pavements with which they have had difficulties.

The image processing that takes place in the detection and classification of cracks in pavement can be divided into four sequential stages: preprocessing, segmentation, feature extraction, and classification. These stages are described briefly below. The emphasis is on the problems associated with the detection and classification of cracks. Since the order and relationship of the algorithms are important, it was felt that the best way to provide a sense of image processing in this context was to present the complete set of algorithms of a proposed system. The system chosen is the final result of an NCHRP project from the early 1990s (Fundakowski, et al. 1991). By comparison with visual ratings carried out by human raters on the same sample, this automated system exhibited an error of 5--17% in its estimate of the extent of the cracking; however, it was over 95% accurate in detecting cracks that were three or more pixels in width. The system did tend to consistently give the detected cracks a higher severity rating than the human raters did.

The areas that have garnered the greatest amount of attention are segmentation and feature extraction/classification; however, preprocessing is an important preliminary to segmentation. In some circumstances, it might be possible to perform segmentation without having undertaken any preprocessing--it would depend on the nature of the raw image. The relationship between feature extraction and classification is much closer: classification is inconceivable without feature extraction.

Segmentation and Preprocessing

Segmentation plays a pivotal role in the processing of images of pavements: satisfactory segmentation is essential if cracks are to be measured and classified successfully. The purpose of segmentation is to separate the cracks in an image from the remainder of the pavement surface and from the various forms of noise that pavement images often contain. If it were not possible to successfully isolate the cracks, it would be impossible to measure them accurately. However, segmentation is not a straightforward matter. Haralick and Shapiro (1985, 100) claim that there is no theory of segmentation: "image segmentation techniques are basically ad hoc and differ precisely in the way they emphasize one or more of the desired properties and in the way they balance and compromise one desired property against another." Writing specifically of the segmentation of images of distressed pavement, Michael Burke (1992, 9) says:

> The biggest problem area is in the segmentation (binarization) of the gray-scale image. . . . Fixed thresholding is extremely idiosyncratic and unstable--a small change in the threshold setting causes large changes in the resulting image, and it is not at all obvious how to automatically determine the proper threshold value. Furthermore, the optimal threshold value is not fixed but varies from image to image (and even from one side of an image to the other side), because the image contrast is itself not constant. It changes with lighting levels, and with pavement reflectivity (which is known to vary from 10-40% in reflectance).

Although Burke's comment might suggest that all segmentation is equivalent to binarization, this is not so. It is possible to segment a gray-scale image without converting it to a binary image; however, almost all of the work done to develop algorithms for detecting and classifying pavement distresses converts the gray-scale image into a binary image. Although working with binary images is not necessary, it does have certain advantages:

> If you can obtain a suitable binary image showing the required object, then there is a wide range of techniques which you can use to enhance the image and extract features that make classification possible. Binary image processing is easier from both the theoretical and practical point of view. Binary images are more amenable to analysis because they have clear cut properties such as boundaries, areas and shape. . . . We can ask questions about the shape of a *binary object* because it has well-defined boundaries. Questions about the shape of an object in a grey level image depend on where we decide its edges are. (James 1988, 102-103)

Often, the source of the greatest difficulty in binary processing is simply getting a satisfactory binary image. James claims that "this can be 99% of the real image processing/recognition problem" (1988, 102).

Getting a suitable binary image is dependent on coming up with a suitable threshold. This has proved to be a difficult task for images of pavement distresses. It is important to remember that an image processing system suitable for analyzing many kilometers of pavement at normal highway speeds will have to be an automated system. If, for example, rather than setting a fixed threshold for the entire pavement, a variable threshold is used to take into account variations in the pavement images, then the image processing system will have to set the threshold levels for each image. So, there will have to be a threshold-setting algorithm incorporated into the system. The goal of a threshold-setting algorithm is to set the "best" threshold for each image, which is the threshold that completely separates all of the cracks--and only the cracks--from the remainder of the image. This is an ideal that is seldom achieved using thresholding alone; however, it is certainly the case that some threshold-setting algorithms are more successful than others.

The simplest way of automatically setting a threshold is by using the histogram method. In an image in which the brightness levels of the cracks are all darker than the brightness levels of the pavement surface, the histogram will be bimodal (see Figure 41a). In this kind of case, it is easy to set the threshold. If it is set in the valley between the two peaks, then the binary image will show the cracks as black and the pavement surface as white (or vice versa). Most often though, the histograms of images of cracked pavements are not bimodal. The histogram in Figure 41b shows a more common distribution of brightness levels. With this type of distribution, it is much more difficult to determine where to set the threshold. Obviously, a person could set it using trial and error, but this is not an option for an automated system.

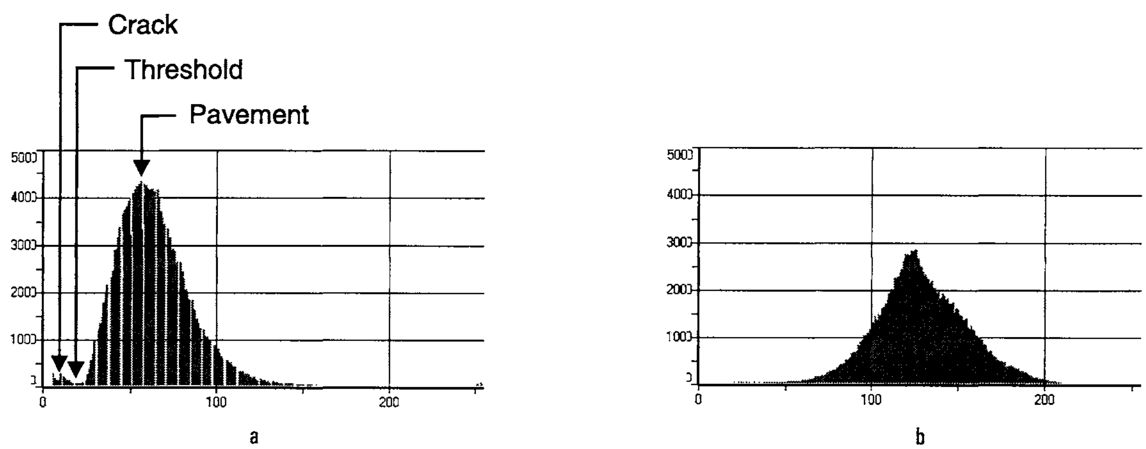


Figure 41. Histograms showing two types of intensity distributions.

Almost all images of pavement require preprocessing to ensure that segmentation will be more successful. One of the mechanisms by which the image processing system will distinguish cracks from the pavement is contrast: cracks are usually darker than the pavement surface. Real pavement surfaces, however, have a wide range of gray levels; consequently, the level of contrast between cracks and the pavement surface varies. If the contrast is too low, it is not possible to distinguish the crack from the surface. The background in an image of a pavement can be made more uniform in intensity through a variety of averaging techniques. Noise is another common problem with images of pavement. Noise is often caused by the texture of the pavement surface. Roughness in the surface of the pavement can create shadows, and small pits in the surface can collect oil or other debris. Very small spots of noise in the gray scale image can often be removed by using low-pass filters or averaging filters. Dark-colored debris in the image, such as a black fan belt or a piece of black rubber hose, cannot be removed by averaging operations. Because of their size and color, they will appear in the binary image along with other large dark spots. It is virtually impossible to screen them out using a thresholding operation because the brightness levels of the pixels that represent them are the same or similar to those that represent cracks. Figure 42 shows two binary images containing noise that remained after the thresholding operation. Figure 42b shows a piece of black rubber hose lying adjacent to the crack. A human rater would have no trouble disregarding the hose, but an image processing system is going to treat the hose in the same way that it treats all of the other black pixels in the binary image. This means, of course, that at this point in the process, it is classified as an object of potential interest, that is, as a possible crack. In order to prevent a misclassification of the hose, it will have to be removed from the image, or some mechanism will have to be devised that will exclude it from consideration even if it can't be excluded from the ranks of black pixels. One possiblity is that the criteria for identifying cracks that will be used in the classification stage will exclude it from being classified as a crack. The removal of large-sized noise can be a tricky matter. The operations that can remove large clumps of black pixels can also remove narrower cracks.

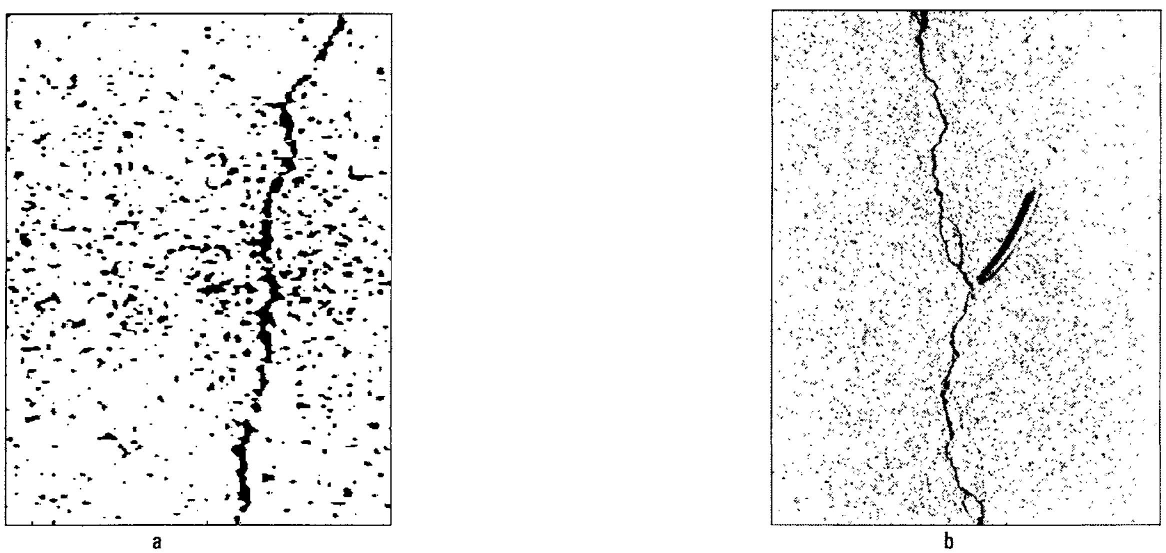


Figure 42. Two noisy binary images.

Feature Extraction and Classification

Writing of pattern recognition operations, Fundakowski et al. point out how important understanding the nature of the problem is to the successful application of a particular type of technique (1991, 11-12):

> There are many different approaches to performing the pattern recognition function, including statistical methods, syntactic or structural graph oriented methods, decision trees incorporating fuzzy logic, pattern associators using neural networks, binary decision trees, and many variants and hybrids of these afore-

> mentioned methods. It is safe to say that no one method is best suited for all pattern recognition applications. . . . All these methods have their strengths and preferred applications. We view the structure of the particular application approach as a secondary issue which must be preceded by an understanding of the problem's unique characteristics.
>
> The main technical challenge in designing a classification approach for a machine vision application is the determination of where sufficient information can be determined for making the necessary decisions and how those decisions can best be made.

This emphasis on the importance of understanding the unique characteristics of a particular problem points up the close relationship between feature extraction and classification. Feature extraction provides the measures by which it is determined whether a particular object meets the criteria embodied in the classification rules. The criteria in question are criteria for identification and not just for classification. The goal of segmentation is to separate all of the objects of interest from the remainder of the image; however, whether the segmentation operation has succeeded cannot be determined until classification takes place. The so-called "objects" in a binary image on which feature extraction and classification are performed are clusters of black (or white) pixels that have been separated from the background, and until feature extraction and classification take place, it is not known whether they are the objects of interest.

The identification and classification of cracks is a complex matter. In some kinds of pattern matching applications--for example, in some industrial applications of machine vision--there is a paradigm that has known features and measurements. In this sort of case, the classification system is measuring the items coming down the assembly line and determining whether they match the paradigm. The objects to be examined are supposed to be just like the paradigm (within certain tolerances). This is not the case when dealing with cracks. There is no paradigm in this sense. What identifies a transverse crack as a transverse crack and a longitudinal crack as a longitudinal crack are criteria--not paradigms. This is not to say that there cannot be a sample of a longitudinal crack. However, the sample does not play the role of a paradigm: "matching" the sample is not a defining criterion of a longitudinal crack.

The criteria used for the different types of cracks in the SHRP *Distress Identification Manual* (1993) point up the type of problem that has to be dealt with. The SHRP description of longitudinal cracking in jointed portland cement concrete surfaces is: "Cracks that are predominantly parallel to the pavement centerline" (1993, 42). The term *predominantly* is forced on the authors by the nature of cracking: cracks do not develop in straight lines, they meander. When the authors say that longitudinal cracks are "predominantly parallel to the center line," they are forced to include the many cases of longitudinal cracks that are not actually parallel to the center line for *any* discrete segment of their length. The consequence of this is that many millions of differently meandering cracks will be treated as longitudinal because they all run predominantly parallel to the center line.

The problem for an image processing system is that the term *predominantly parallel* is not precisely defined. Human raters are able to perform useful distress ratings using SHRP criteria as guidance. In some sense, they discount or ignore the meandering of the crack in many directions: they have a "sense"--no doubt developed by training--that a particular crack is or is not roughly parallel to the center line. However, image processing systems are very "literal minded," so this defining criterion must be translated into a more precise set of rules that can be incorporated into the algorithms of the feature extraction and classification stages of the image processing system. It would obviously be completely impractical to set up algorithms that treated the meandering shape of each of the millions of possible longitudinal cracks as unique. It may be possible, for example, to use the method of least squares to substitute a straight line for the meandering shape of any particular crack. Then, the classification of the alignment of a particular crack as longitudinal, transverse, or diagonal would depend on rules that more precisely specified the relationship of this straight line to the center line.

The classification of cracks by type, e. g., longitudinal, transverse, etc., requires a specification of defining features or criteria that will distinguish the types. Each specific type must have its own set of unique defining features. The problem of classification in its simplest terms is whether a particular cluster of pixels belongs to one of the types of cracks previously defined. Features of the cluster are measured. The features of the cluster that are measured are the features that will serve to determine whether the cluster is a crack and, if it is, to what type it belongs. The classification of a cluster is dependent on how closely the measured features of the cluster compare with the defining criteria. The defining criteria are not necessarily the criteria that would normally be used to define the crack type. They serve a particular role in the differentiation and classification of clusters of pixels. A feature that may normally serve as one distinguishing criterion of cracks, would serve no purpose in a classification system that cannot measure it. The depth of a crack may normally be one of several important criteria used for determining the severity of a crack. An image processing system cannot measure the depth of cracks; however, if the severity of cracks can be determined using measures other than depth, then it will still be possible to classify them successfully.

NCHRP Project 1-27

Preprocessing

Fundakowski et al. (1991) see the adjustment of contrast in images as the principal reason for preprocessing operations. An adjustment in contrast is necessary because of local variations in the intensities (brightness values) of the pixels that make up the background of an image. These differences in background intensities are the result of differences in the color of the pavement, uneven lighting, shadows, oil stains, differences in wear, surface texture, etc. Variations in the darkness of the backgrounds of pavement images complicate the image processing and analysis. These variations are common, and they occur from image to image and even within single images. It is not at all uncommon for the local background intensities to be different from one side of an image to the other.

The goal of the preprocessing in this case is to make the pavement surface appear more uniform, which makes it easier to distinguish surface distresses from normal pavement surfaces.

The authors examined two different techniques for achieving what they call the "normalization of the background." The first they call "local background suppression with global contrast enhancement," and the second is called "local background suppression and local contrast enhancement."

Local background suppression with global contrast enhancement. There are four steps to this algorithm. First, local average pixel intensities are calculated. Each pixel in the image serves as a center pixel in a square kernel. The average intensity for the kernel is calculated. This serves as an estimate of the mean intensity. Second, the mean intensity value determined in step one is subtracted from the intensity value of the center pixel in each kernel. Third, 127, which is the midrange value for an 8-bit gray scale, is added to the result of the second step. Fourth, a histogram modification of the results of step three is used to "stretch" the dynamic range of the image so that it has better contrast. The results of applying this operation can be seen in Figure 43, which shows an image of a pavement surface with variable contrast between the background and the pavement distresses, and Figure 44, which shows the same image after the application of the operation. (The images used here and throughout the remainder of this account of the work of Fundakowski et al. are of poor quality. They are essentially copies of copies; however, they are not available in any other form. Despite their poor quality, they are a reasonable representation of the results of applying the algorithms that are discussed here.)

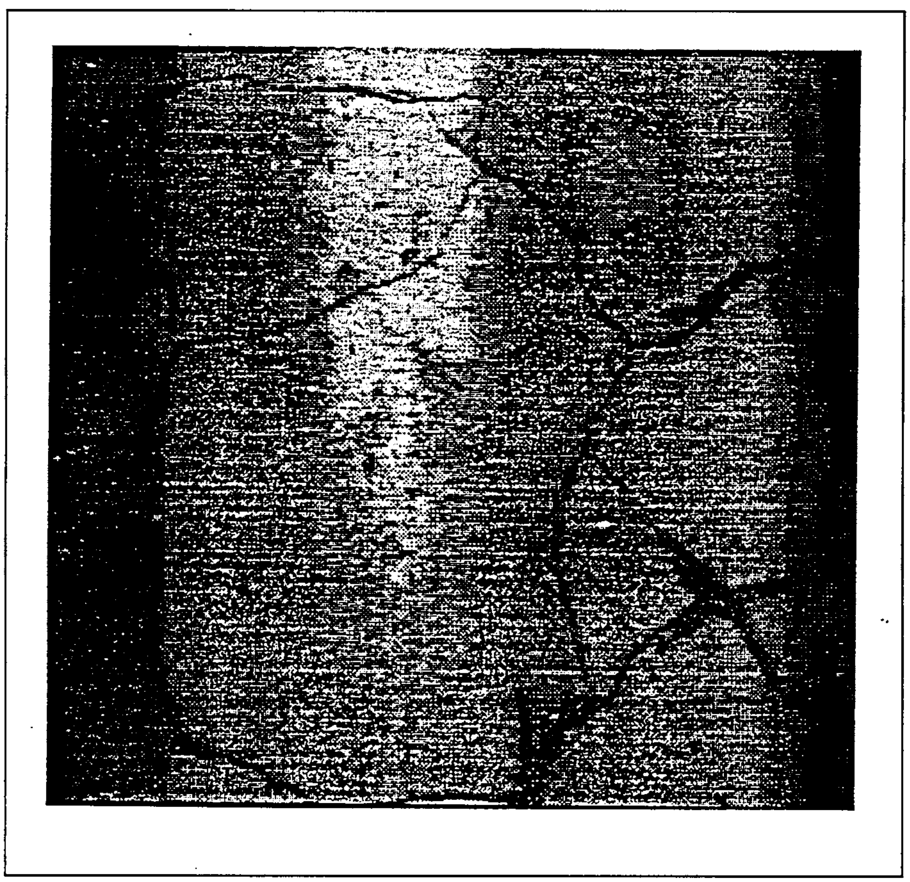


Figure 43. An image of cracked pavement with varying contrast between the pavement surfaces and the cracks. Source: *Video Image Processing for Evaluating Pavement Surface Distress*

Figure 44. An image of cracked pavement after the application of a contrast enhancing operation. Source: *Video Image Processing for Evaluating Pavement Surface Distress*

Local background suppression and local contrast enhancement. This operation modifies the intensity level of each pixel in the image by the following transformation:

$$I\_out_{i,j} = Gain \bullet [(I\_in_{i,j} - NM_{i,j})/NSD_{i,j}] + Bias$$

where $I\_out_{i,j}$ and $I\_in_{i,j}$ represent the output and input values at position i, j; $NM_{i,j}$ and $NSD_{i,j}$ represent the neighborhood mean and the standard deviation of the intensity values in the neighborhood of i, j; Gain and Bias are specified constants. This technique modifies the intensity of each pixel in the input image in proportion to its deviation from the mean of the pixel intensities of its neighborhood and inversely proportional to the standard deviation of the pixel intensities of its neighborhood. The modification of intensity values in inverse proportion to the *local* variability is what distinguishes this approach from the first one. The increased local sensitivity of this approach to variability in the contrast of different parts of an image makes it superior to the first approach with images in which the relative contrast is low between the distress and the neighboring background.

Segmentation

The purpose of segmentation is to separate the pixels in the image that represent cracks from the pixels that represent the pavement surface. For the most part, the pixels that represent cracks are darker than the pixels that represent the pavement surface--but not always (El Korchi and Wittels 1990). Fundakowski et al. developed a two-step segmentation operation. The first step is the selection of a threshold, and the second is the application of a connectivity operator. The application of a thresholding operation creates a binary image. Figure 45 shows that the image created in this way has a lot of noise in it. The application of a connectivity operator serves to identify clusters of dark pixels that have passed the thresholding stage, and it eliminates small objects that are determined to be insignificant (that is, not likely to be a part of any crack). Figure 46 shows the result of applying a connectivity operator to the binary image in Figure 45. Many of the smaller clusters of black pixels have been included in the background.

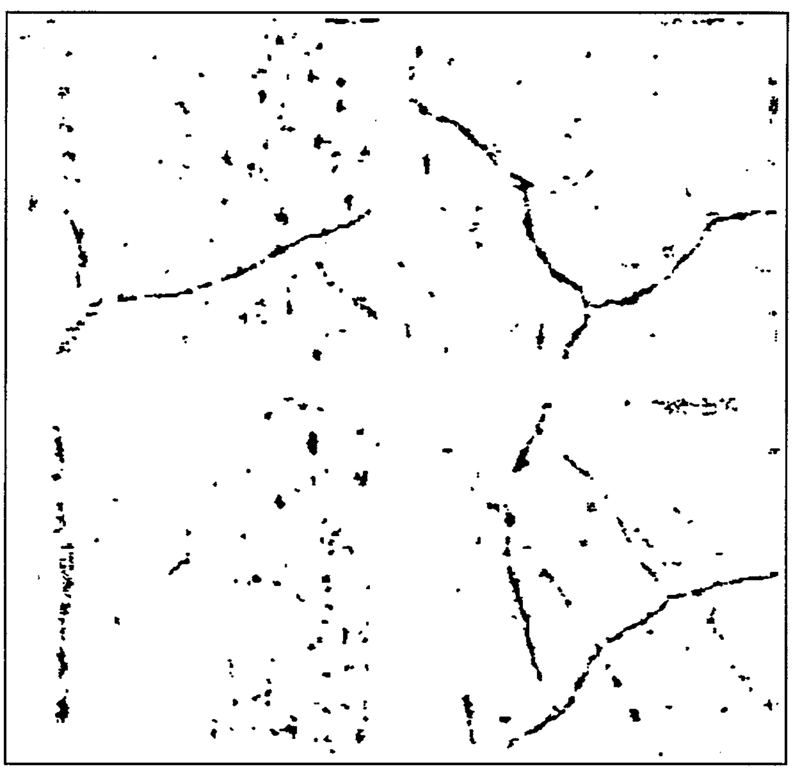


Figure 45. A binary image created by the application of a thresholding operation. Source: *Video Image Processing for Evaluating Pavement Surface Distress*

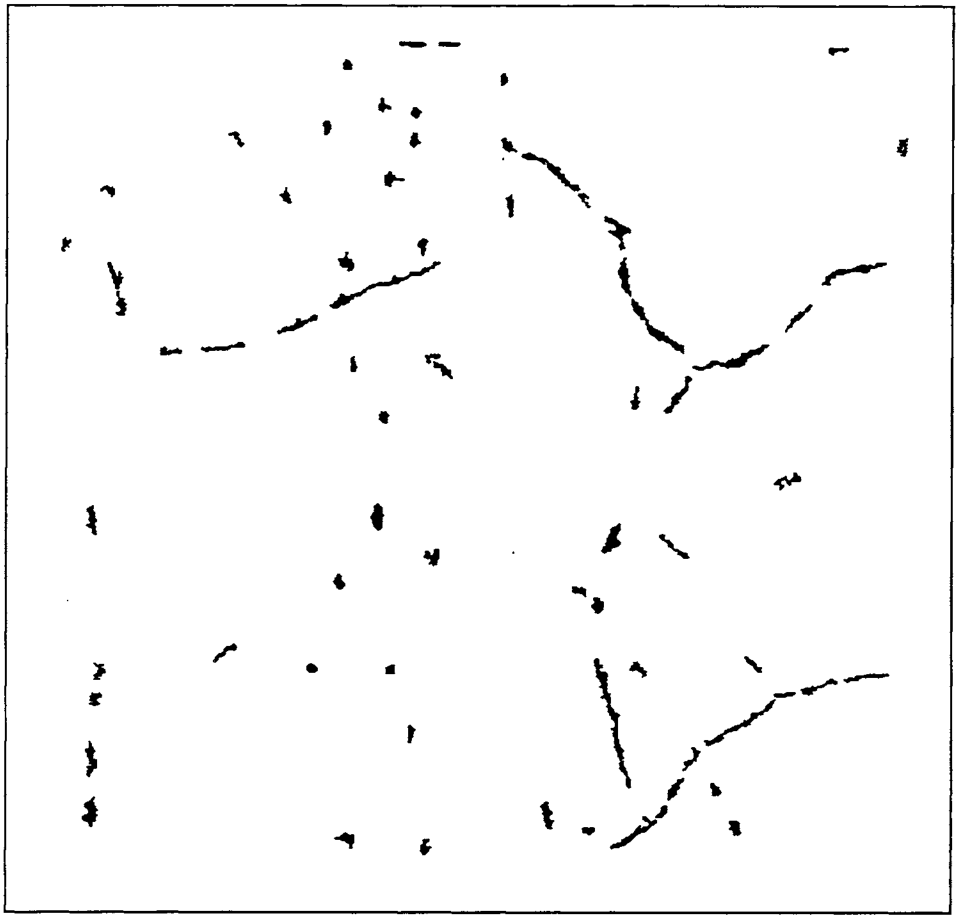


Figure 46. A binary image after the application of a connectivity operator.
Source: *Video Image Processing for Evaluating Pavement Surface Distress*

The thresholding operation used here applies a threshold to each pixel independently. Each pixel's threshold is determined by computing the standard deviation of the intensities of neighboring pixels. A pixel should be considered dark in relation to its neighboring pixels if it is K units of standard deviation below the mean intensity value of these neighboring pixels; consequently, the threshold value for each pixel is set at K units of standard deviation below the local mean.

The result of applying the thresholding operation and the connectivity operation is a binary image containing groups of contiguous dark pixels on a white background. At this stage of the process, no decision has been made about the status of these groups of black pixels; whether they represent cracks has not yet been determined. At this time, they are referred to as regions of interest (ROI). This rather noncommittal classification is really necessary because, although some cracks or portions of cracks will be accurately represented, segmentation operations applied to images of cracked pavement will not normally produce binary images containing cleanly outlined cracks.

Feature Extraction

The effect of converting the original gray scale image into a binary image is that, at least in most cases, the majority of the data in the original image is removed from consideration. The image processing has separated ROIs from the remainder of the image, and from this point on, the focus will be on these areas only. The final goal of the entire process is the classification of these ROIs as cracks with an indication of their type, severity, and extent or as something other than a crack, which should be dropped from consideration. Whether the ROIs are classified as cracks or not and how they are characterized will depend on whether they have a specified set of measurements. In image analysis, the process of making these measurements is called *feature extraction.* The crucial problem in classification operations is the selection of appropriate features. An inappropriate set of features would make classification impossible.

Fundakowski et al. extract about 60 measures from each ROI, although all of them are not always used in every classification. They have chosen measures that describe the following general characteristics of the ROIs:

- Position: These features are used to describe the position of ROIs on the pavement and the proximity of pairs or groups of ROIs.

- Shape: Features that characterize the shape of ROIs are essential for distinguishing between types. A straightforward example of this would be a feature describing the degree of elongation (the width of a crack is small compared to its length). Measures that characterize the orientation of the ROI are included for distinguishing between longitudinal, transverse, and diagonal cracks. The need to distinguish between cracks and joints means that some measure of the straightness of ROIs must be included.

- Size: Features describing the size of the ROIs contribute to classifying them by their severity and their extent. Some of the measures used were length, width, area, and transverse and horizontal extent.

Classification

There are many different approaches to the classification of ROIs. Fundakowski et al. view the selection of the particular structure of the classification design as a secondary issue that must be preceded by an understanding of each problem's unique characteristics. They see the main technical challenge in designing a classification system as the determination of where sufficient information can be found for making the decisions and how those decisions can best be made (1991, C-12).

They investigated three different methods for classifying ROIs in images of pavements. The first method used the following features to discriminate between cracks and joints: degree of elongation, orientation, RMS error about best fitted line, compactness, average width, and length. In tests on 150 images of cracks and 20 of joints, a classifier based on these features was successful 75% of the time. The errors in classification were caused (1) by ROIs that were a small portion of a crack or joint, which were the result of a crack being fragmented during segmentation and (2) by complex ROIs consisting of several cracks of different orientations merged together or longitudinal cracks merged with a joint. In order to deal with this problem, two other classification methods were investigated.

The merging of one crack with another is the simplest form of a type of cracking that is not linear but covers an area in a complex pattern. Alligator cracking could be viewed as an extreme example of this type of cracking. So, these problems with the merging of cracks and with the fragmentation of cracks required that classification techniques sensitive to image regions and to the grouping of ROIs be developed. The two techniques investigated were projection signatures and the Hough Transform.

A projection signature is the summation of the pixel intensities of an image or a region of an image in a particular direction. A projection signature looks very much like a histogram; however, it is not a summation of all of the pixels at each of the possible brightness levels of the gray scale. In a binary image, it is the summation of all of the pixels in one line along the x-axis or the y-axis. The pixels summed are object pixels not background pixels. In the case at hand, they are black pixels. Figure 47 shows a vertical projection signature placed over the image it represents. The image, which is a product of the segmentation process, contains a transverse joint with a long longitudinal crack crossing it. It is easy to see the projection signature's directional character. The large spike just beyond 250 on the scale reflects the large number of black pixels at that location in the image. A horizontal projection signature would reflect the presence of a large number of black pixels in the area of the image where the transverse joint is located. A longitudinal or transverse ROI will produce a spike in a projection signature. Since a projection signature is a summation over a region of an image, it can accurately pinpoint groups of ROIs that are the fragments of a crack, and because of its strong directional character, it can separate cracks that have merged (in the way that it separated the longitudinal crack from the transverse joint in Figure 47). The weakness of the projection signature is that it is not able to recognize ROIs that are not longitudinal or transverse. Diagonal ROIs do not produce strong spikes in horizontal or vertical projection signatures. To make up for this shortcoming, the Hough Transform was employed. It is equally sensitive to any orientation, not just longitudinal and transverse. Not only did this technique prove to be useful in detecting diagonal ROIs, it also proved to be useful in interpreting complex objects such as those resulting from patterns of cracking.

The design of the classifier used by Fundakowski and his associates uses all three of these techniques:

- features of individual ROIs

- projection signatures of the pixels comprising all ROIs

- orientation and position descriptions of the ROIs using the Hough Transform.

These techniques are complimentary. By including all of them, the system is not dependent on only one technique to provide detection of all possible ROIs.

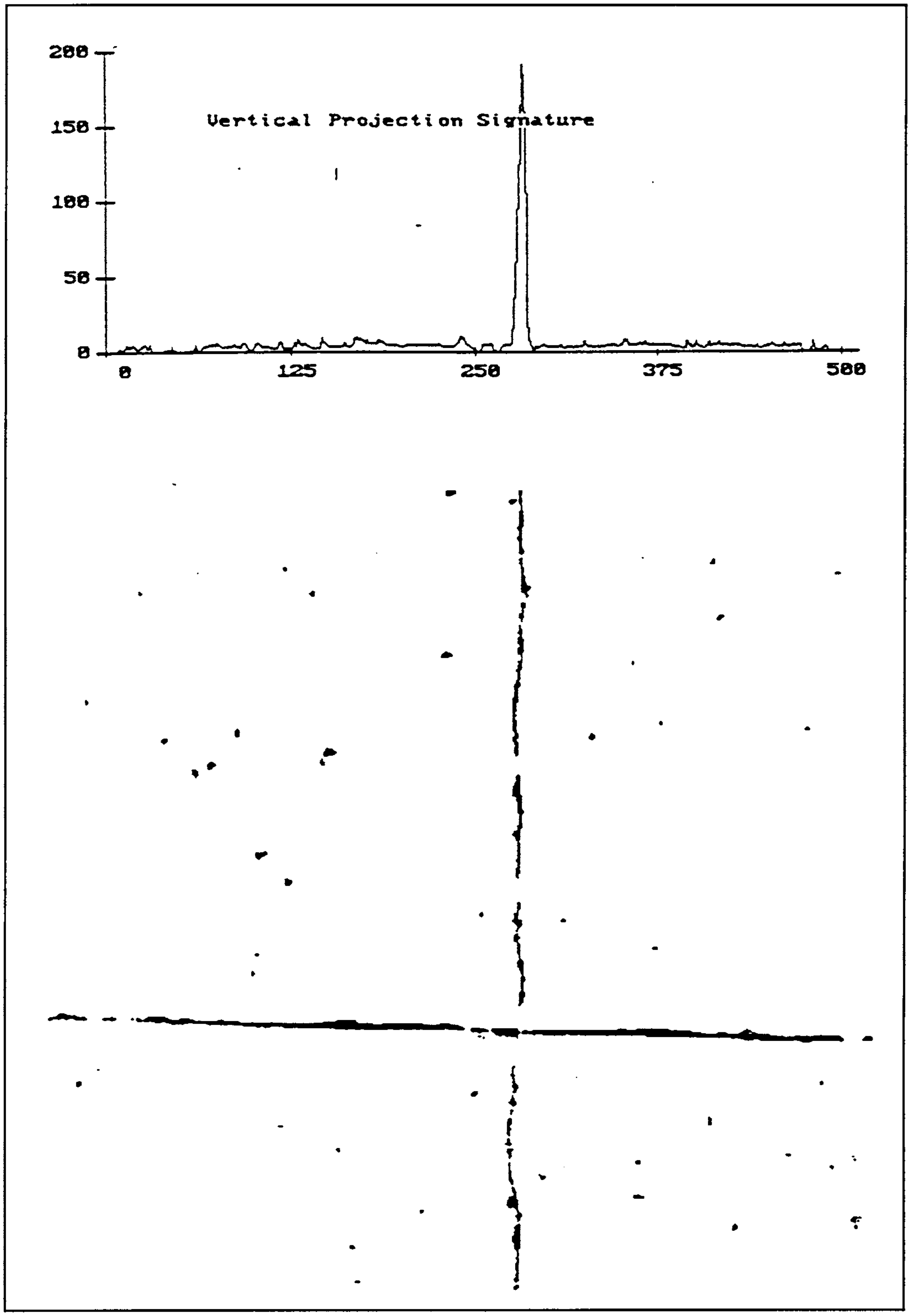


Figure 47. The representation of a group of ROIs using a vertical projection signature.
Source: *Video Image Processing for Evaluating Pavement Surface Distress*

The first stage of the classifier is the *classifier interest operator*, which determines whether the image has any "interesting" areas in it in virtue of the features of the ROIs or the clustering of their orientation and position. The characteristics of an image that have a sufficiently high interest are used to "seed" the next stage of the process. If no "seed" is generated, the image is assumed to have no cracks, and it is discarded. The *pattern/group associator* assembles the group of ROIs that appear to be related to the "seed" and interprets them as a whole. This process entails assessing the ROIs contained in the image and making note of those that are related to the "seed" condition in virtue of their position in the image and/or their orientation, shape, etc. As a result of this association process, a graph is generated that describes all ROIs in the image that correspond directly or indirectly to the characteristics in the image that resulted in the "seed" generation. The result of this is that the fragments of a crack or patterns of cracking are grouped for interpretation. The final stage of the classifier is called the *interpreter*. The interpreter decides whether each group of ROIs grouped by the pattern/group associator should be classified as a particular type of crack. If it determines that a particular group should be so classified, then it determines its severity level and extent. The interpreter is a rule-based system. These sets of rules are criteria against which the interpreter compares the groups of ROIs. For example, the rules for jointed PCC pavements characterize a joint as being long, narrow, and approximately straight, whereas a transverse crack is characterized as long and not very straight; consequently, any group of ROIs that is long, narrow, in a transverse direction, and relatively straight will generate a high confidence value as a transverse joint. Once the type of the crack is determined, its severity and extent are determined using the same procedure. Figure 48 shows an example of the output of this system.

IMAGE ACQUISITION

The requirements for an image acquisition system are largely dependent on the nature of the processing that the images will undergo and on the goals of the pavement distress survey. For example, one of the most important considerations would be: What is the minimum size of the cracks that the system should be able to recognize (i.e., what resolution is needed)? Questions of resolution cannot be dealt with in isolation; these questions are bound up with questions of the speed of the surveying vehicle, the capacities of the imaging hardware, what portion of the width of the pavement needs to be surveyed, etc. The emphasis in Part 1 is on describing the problems that are encountered in providing images suitable for use by an automated image processing system. In Part II, image acquisition hardware is described. Particular emphasis is placed on how the hardware options that are available can contribute to the solution of the problems mentioned in Part I.

Part 1: Nature of the Problem

To begin with, it is necessary to provide a set of provisional requirements for the image acquisition system. The set described below is provided as a means of clarifying their interrelationships and showing how the problems associated with any attempt to implement them could

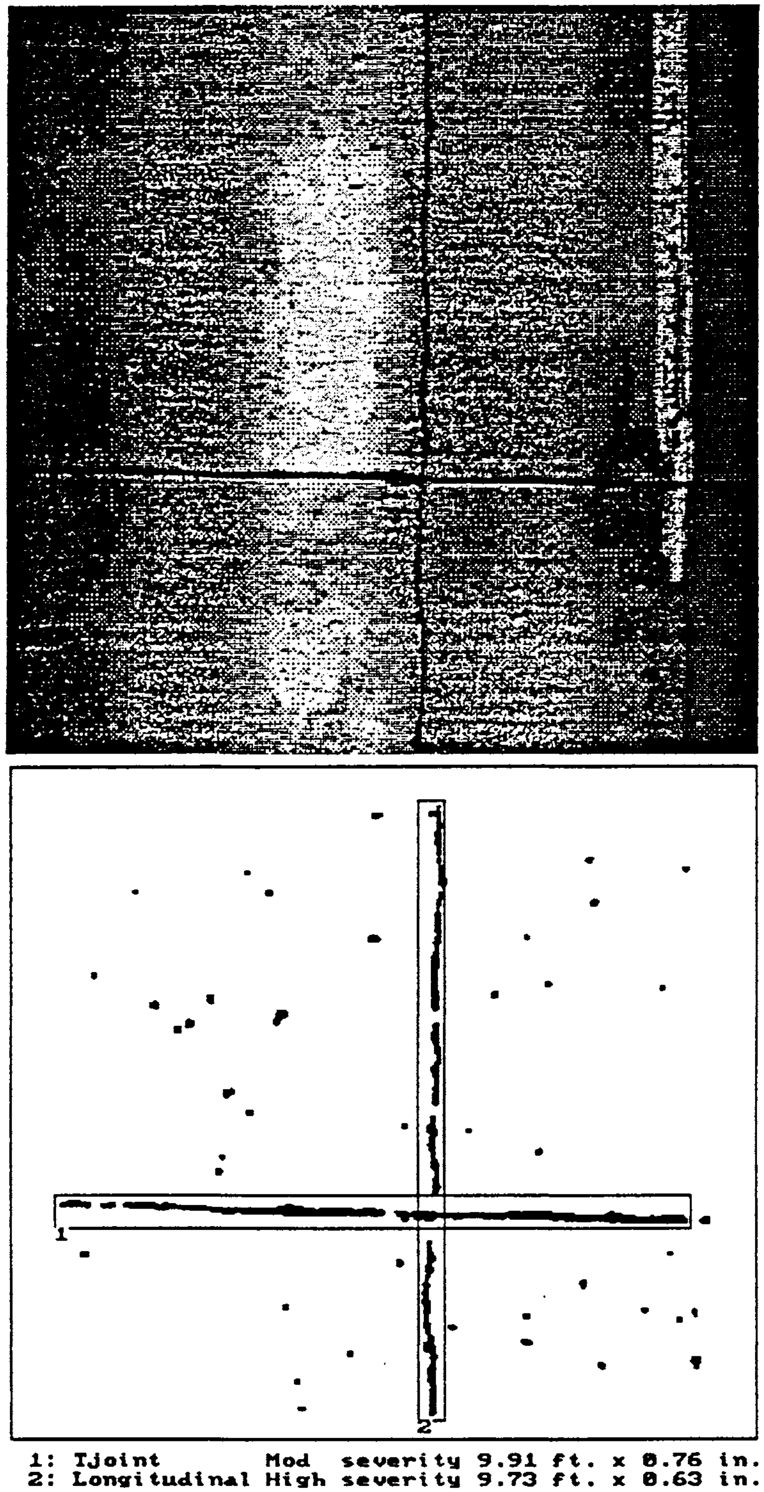


Figure 48. Typical output of the system. Source: *Video Image Processing for Evaluating Pavement Surface Distress*

be dealt with. It is not being suggested that VDOT should adopt them. A system that met these requirements would be a very high resolution system, so these requirements should be seen as an upper limit. The requirements are as follows:

- The images should show cracks as small as 1 mm (.039 in).
- The full 3.66-m (12-ft) width of the lane should be covered.
- Down-lane coverage should be 100% (the entire pavement should be imaged, not just a sample).
- The survey vehicle should be able to maintain a speed of 72.41 km/hr or 20.12 m/sec (45 mph).

In order for an automated image processing system to be able to detect cracks as small as 1 mm, the sampling rate must be .5 mm; consequently, 7,320 pixels are required to cover the full width (3,660 mm) of the lane. If the survey vehicle is traveling 20.12 m/sec, then the number of pixels traversed down lane with 100% coverage in 1 second is 40,240. The nature of the task that must be accomplished by the image acquisition system can now be made clearer by multiplying the cross-lane resolution in pixels by the down-lane resolution in pixels (7,320 x 40,240 = 294,556,800). Thus, at this resolution, speed, and coverage, the image acquisition system must be able to process more than 294 million pixels every second. This means that the storage hardware on the survey vehicle must be capable of storing the 1,060,404,480,000 pixels (which is more than 987 gigabytes) that are processed every hour that the surveying takes place.

Table 1 shows how variations in any one of these requirements affect the rate at which data is acquired. The rate of data acquisition is important because it gives some indication of the required capacity of the image acquisition system. At its most basic, the image acquisition system in the survey vehicle will consist of a camera and some means of recording the output of the camera. If, for example, the survey vehicle is to travel at 20.12 m/sec and the image acquisition system is to fulfill the other requirements mentioned above, the camera must be able to process and output to a recording device (or devices) more than 294 million pixels per second. If it is determined that the pixel resolution need only be 1 mm, then the data rate drops to 73,639,200, which is one fourth the capacity required with the first alternative. If it is determined that 20.12 m/sec is too slow to be safe (on an interstate, for example), and the vehicle speed is changed to 24.57 m/sec (55 mph), then the acquisition system will have to be able to process 359,704,800 pixels each second. Table 1 does not reflect the results of changes in more than one of the requirements at a time; it would, of course, be possible to set the speed of the survey vehicle at 24.57 m/sec and to reduce the required pixel resolution to 1 mm. This would put the data rate at 179,852,400 pixels per second, which is approximately a 39% reduction in the required capacity of the image acquisition system and a 22% increase in the speed of the survey vehicle. Similar compromises can be made by reducing the width of the lane to be covered or by reducing the amount of downlane coverage. What will turn out to be reasonable compromises will be based on the goals of the pavement distress survey. It may be reasonable to equip one survey vehicle with a lower resolution, less expensive image acquisition system suitable for surveying the entire pavement system at higher speeds, and another vehicle with a high resolution acquisition system for those limited circumstances in which high resolution is essential.

TABLE 1

Variations in the Requirements for Image Acquisition

Variations in	Pavement Width	Pixel Resolution	Image Resolution	Speed	Coverage	Data Rate (pixels/sec)
Pavement Width	3.66 m (12 ft)	0.5 mm (.020 in)	1 mm (.039 in)	20.12 m/sec (45 mph)	100%	294,556,800
	1.83 m (6 ft)	0.5 mm	1 mm	20.12 m/sec	100%	147,278,400
	.91 m (3 ft)	0.5 mm	1 mm	20.12 m/sec	100%	73,236,800
Resolution	3.66 m (12 ft)	0.5 mm (.020 in)	1 mm (.039 in)	20.12 m/sec (45 mph)	100%	294,556,800
	3.66 m	1.0 mm (.039 in)	2 mm (.078 in)	20.12 m/sec	100%	73,639,200
	3.66 m	1.5 mm (.059 in)	3 mm (.118 in)	20.12 m/sec	100%	32,727,720
	3.66 m	2.0 mm (.078 in)	4 mm (.156 in)	20.12 m/sec	100%	18,409,800
Speed	3.66 m (12 ft)	0.5 mm (.020 in)	1 mm (.039 in)	24.57 m/sec (55 mph)	100%	359,704,800
	3.66 m	0.5 mm	1 mm	20.12 m/sec (45 mph)	100%	294,556,800
	3.66 m	0.5 mm	1 mm	15.66 m/sec (35 mph)	100%	229,262,400
Downlane Coverage	3.66 m (12 ft)	0.5 mm (.020 in)	1 mm (.039 in)	20.12 m/sec (45 mph)	100%	294,556,800
	3.66 m	0.5 mm	1 mm	20.12 m/sec	75%	220,917,600
	3.66 m	0.5 mm	1 mm	20.12 m/sec	50%	147,278,400
	3.66 m	0.5 mm	1 mm	20.12 m/sec	25%	73,639,200

Part 2: Imaging Hardware

The hardware for the image acquisition system will consist of a camera or cameras and recording devices. It will also be necessary to have some form of playback hardware, but that needn't be in the survey vehicle. The configuration of the hardware in the survey vehicle will be dependent on the type of camera chosen and the number of cameras required. Before the data from the camera can be recorded, it may be necessary to have it "buffered," i.e., temporarily stored or converted into a format compatible with the recording device. For example, if line scan cameras are used, then the buffering may take the form of converting from line scan data format into a video image frame for recording in a video format on video cassette recorders. Recorders are a form of storage device for electronic cameras; for cameras that use film, film is the storage medium.

Cameras

There are two types of cameras: electronic and film. Both types of cameras convert a scene's intensity into a form suitable for storage. The output of electronic cameras is an electronic signal, and the output of a film camera is a chemical exposure on film. There are a number of factors that are important for determining how many and what types of cameras will serve the purposes of an image acquisition system for pavement distress surveys.

The resolution of the camera. The horizontal resolution of the camera is one important factor in determining how many cameras are required to cover the full width of the pavement. The horizontal resolution of all of the cameras together must equal or exceed the desired horizontal resolution of the pavement. For example, if the required horizontal resolution of the pavement is 7,320 pixels (3,660 mm at .5 mm resolution), then 12 cameras with a horizontal resolution of 640 pixels will be needed.

The framing rate. The framing rate is the number of frames shot per second. The framing rate must exceed the rate at which the camera moves through the number of lines of pixels that make up the vertical resolution of the camera. A standard RS-170 video camera has a framing rate of 30 frames per second. It is common for such cameras to have a resolution of 640 (horizontal) by 480 (vertical) pixels. At a speed of 20.12 m/sec and a pixel resolution of .5 mm, such a camera would have to have a framing rate of 84 frames per second. At 30 frames per second, the speed of the vehicle could be no more than 7.2 m/sec (16 mph). If the vehicle goes faster than 7.2 m/sec, there will be gaps between the frames, that is, some of the pavement will not be photographed. This would mean that the system would be sampling the pavement rather than providing 100% coverage. If the vehicle goes slower than 7.2 m/sec, the frames will overlap. The slower the speed, the greater the overlap. The consequence of this would be that the image processing system would be analyzing and counting the same crack more than once.

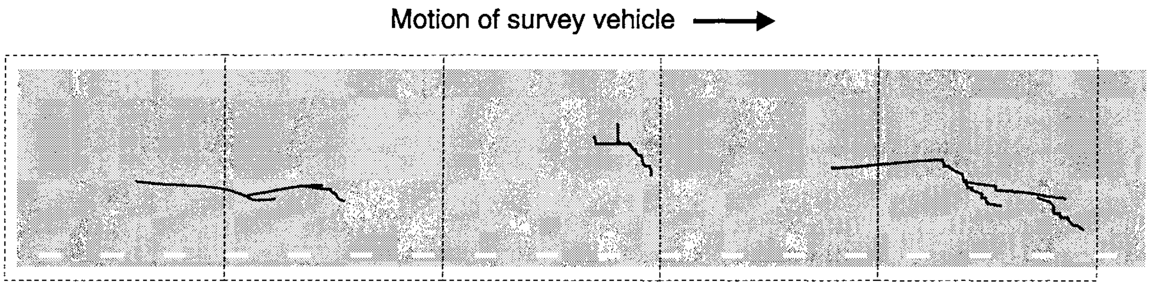
The data rate per frame. The data rate per frame is simply the number of pixels in a frame for a given camera. For example, a camera with a pixel resolution of 640 x 480 will have a data rate per frame of 307,200 pixels. At 30 frames per second, the camera's data rate will be 9,216,000 pixels per second.

Number of pixels in the sample size. If the image acquisition system is only to provide a sample of the pavement, that is, the downlane coverage will be something less than 100%, then it is important to know the number of pixels in the sample. This is determined by multiplying the width of the sample (divided by the pixel resolution) by the length of the sample (divided by the pixel resolution). If the width to be covered is 3,660 mm, the speed is 20,120 mm/sec, the resolution is .5 mm, and the coverage is to be 50%, then each sample will contain 147,278,400 pixels.

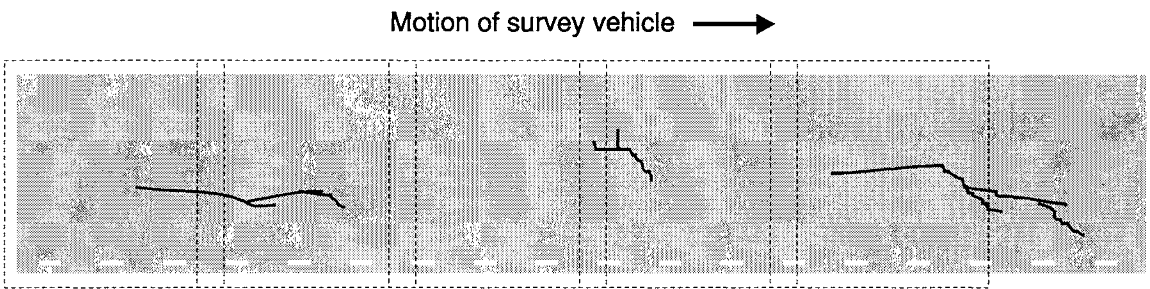
Shutter speed. Image smearing is a problem when imaging a moving object or when imaging a stationary object from a moving camera, which is, of course, what occurs when using video cameras to image the pavement during a distress survey. In either case, the subject of the imaging moves or appears to move relative to the film while the shutter is open. Smearing is usually not a problem if the camera or the subject of the imaging does not move more than 10 to 20% of the desired resolution during the time the shutter is open. Illumination is directly related to the problem of smearing: if there is insufficient illumination of the subject of the imaging, it will not be possible to reduce the shutter speed to the point that smearing does not occur. Some smearing can be corrected by image processing.

Electronic Cameras

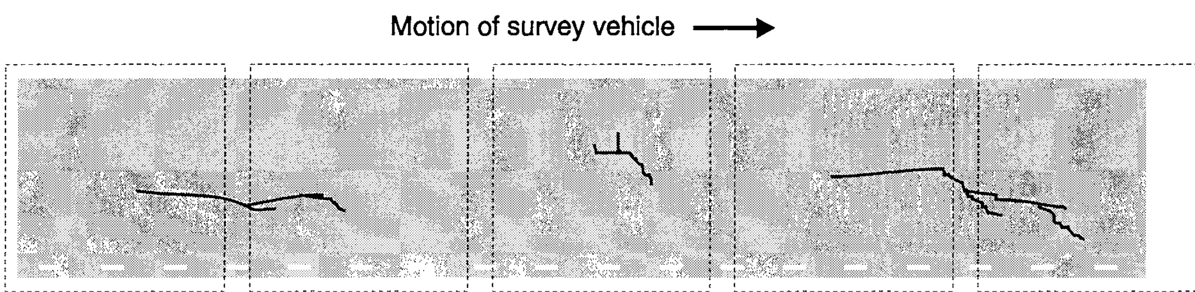
There are two types of electronic cameras: framing and line scan. Video cameras and electronic still cameras are both framing cameras. As a matter of fact, the vast majority of cameras are framing cameras. A framing camera shoots an image of a certain size and resolution. While the shutter is open, an area of a certain size is imaged. The size of the area imaged is dependent on the focal length of the lens and the distance of the subject from the lens. For example, a lens with a 100 mm focal length can be used to focus on a single chair at 3 m, but it can also be used to image a mountain that is 20 km away. A camera of this sort is called a framing camera because whatever is in the area framed by the lens is imaged at the same time. The resolution of this area is determined by the number of pixels in the charged couple device (CCD) of the camera. A typical video camera CCD may have a pixel array of 700 x 500. In this case, the total number of sampling points will be 350,000, and they will all be sampled at the same time. There is a limit to the number of frames this type of camera can shoot in a given amount of time--usually a second--and this is called its framing rate. A camera with a 30-per-second framing rate will shoot thirty 350,000 pixel frames in 1 second, each having a 700 x 500 pixel arrangement. If such a camera were used to image a pavement from a moving survey vehicle, its shutter would open and close 30 times each second, and each time a 700 x 500 pixel array would be sampled. Figure 49 shows that, depending on how fast the survey vehicle is moving, the frames will meet edge to edge, overlap, or have gaps between them. The actual measurement of the pavement imaged will depend on the focal length of the lens and its distance from the pavement. The larger the area covered, the lower the resolution of the image will be, because each pixel will cover a greater distance.



a. Speed of survey vehicle matches framing rate.



b. Speed of survey vehicle is slower than framing rate.



c. Speed of survey vehicle is faster than framing rate.

Figure 49. The relation of vehicle speed and framing rate. Note that in b the frames overlap, and in c the entire pavement is not imaged.

To determine how well suited video cameras would be to the task of surveying pavements from a moving vehicle, it is necessary first to specify the capacities of a specific camera or type of camera. As far as resolution is concerned, there are roughly three types of cameras: (1) standard video cameras, which have resolutions in the 640 x 480 range, (2) high-resolution video cameras, typically 1,024 x 1,024, and (3) very high resolution cameras, some of which have CCD pixel arrays greater than 4,000 x 4,000.

Resolution is not the only area of concern, the output or throughput of the camera and its framing rate are also important. If the survey vehicle is traveling at 20.12 m/sec and it is required that the full width of a lane be imaged at a .5 mm pixel resolution, then the width in pixels of the imaged area is 7,320 and the length (again in pixels) is 40,240. A standard RS-170 video camera with a resolution of 640 x 480 and a fixed framing rate of 30 frames per second has a data rate of 9,216,000 pixels per second. The camera would have to have a data rate of 294,556,800 pixels per second to meet the requirements above. First of all, the horizontal resolution is far short of 7,320. Second, the vertical resolution of 480 pixels is such that the distance covered downlane during 1 second is only 14,400 pixels (30 x 480), whereas what is needed is 40,240 pixels. Even if 12 cameras were mounted side-by-side to provide sufficient cross-lane coverage, although much more of the required area would be covered, the problem with the down-lane coverage would remain. The framing rate causes other problems. To get full coverage downlane at 20.12 m/sec, the framing rate would have to be 84 frames per second; at 30 frames per second, the speed of the survey vehicle could be no greater than 7.2 m/sec (16 mph). Any speed greater than this would mean that gaps will appear between the frames. The situation is not much improved if high-resolution cameras are used. Many of these cameras retail for $6,000 to $12,000, and at a resolution of 1,024 x 1,024, it would take 8 of them side-by-side to provide sufficient cross-lane resolution. The down-lane coverage at 30 frames per second is still far short of what is needed, and the maximum speed the vehicle could maintain would be 55 km/hr (34 mph). Framing cameras with pixel arrays in excess of 4,000 x 4,000 do exist. For example, one camera currently on the market has a pixel array of 4,096 x 4,096. The manufacturer's advertised retail price for this camera is about $70,000. Two of these cameras mounted side-by-side would provide more than enough horizontal resolution. The vertical resolution is so high that it would only be necessary to have a framing rate of 10 frames per second to cover the required distance down the lane, and the manufacturer advertises it as having "fast" frame rates; however, with a pixel array this large, *fast* means one frame every 9 seconds.

If one or more of the requirements are relaxed, more possibilities are opened up. If the resolution required is reduced to 2 mm for the image and thus 1 mm for the pixels, it would be possible to cover the horizontal resolution with four 1,024 x 1,024 high resolution cameras. At 30 frames per second, the down-lane coverage would be 30,720 mm; consequently, it would be possible to increase the speed of the survey vehicle to about 31 m/sec (69 mph). Under these circumstances, if the speed of the survey vehicle were reduced to 24.57 m/sec (55 mph), there would be a significant amount of overlapping of the frames. Only if the cameras had a variable framing rate would it be possible to reduce the speed of the vehicle and yet avoid overlapping. At a speed of 24.26 m/sec and a frame rate of 30 frames per second, it would be possible to survey at a pixel resolution of .79 mm, which would allow for the resolution of cracks 1.58 mm wide. However, in this case, a fifth camera would be required to achieve sufficient cross-lane resolution.

Digital still cameras are the electronic counterpart to the typical 35 mm SLR. Like video cameras, they contain a CCD. At least one digital camera on the market has a pixel resolution of 3,060 x 2,036; however, this is unusual. More commonly, the pixel array is approximately 1,200 x 1,000 or 1,500 x 1,200. The cameras themselves have a very limited capacity to store images. If the highest resolution is used, 40 to 60 images can usually be stored in the camera. At lower resolutions, well over a hundred can be stored. In many cases, it is possible to connect the camera directly to a PC, which would allow a much larger number of images to be stored. However, the framing rate is the most serious hindrance to using this type of camera for pavement surveys. The framing rate varies from camera to camera, but in all cases, they are inadequate. For example, the camera that has a resolution of 3,060 x 2,036 has a framing rate that is 12 seconds long. The first exposure can be taken 1/4 second after the camera is turned on, but the shortest interval at which *each* of the remaining exposures can be taken is 12 seconds. The situation is somewhat better with the cameras that have lower resolution, some of which can shoot "bursts" of 12 exposures in 4 seconds; however, this is still woefully inadequate for pavement surveys. At 20.12 m/sec, the survey vehicle traverses 80.48 m of the pavement in 4 seconds. If the camera has a vertical resolution of 1,200 pixels at a pixel resolution of .5 mm, then the coverage down lane for 12 frames in 4 seconds would be 7.2 m, which is only 8.9% coverage of the pavement, and this is assuming that there are a sufficient number of cameras to cover the entire width of the pavement.

Line scan cameras are better suited to the task of imaging moving objects. A line scan camera images one line at a time (see Figure 50). Framing cameras are designated by a two-dimensional array of pixels, for example 700 x 500. By contrast, line scan cameras are designated by one-dimensional arrays; however, if a line scan camera were designated with an array that paralleled that of a framing camera, it would look like this: 700 x 1. A line scan camera with a 700 pixel resolution would image the area covered by a 700 x 500 frame camera one line at a time. For this to take place, either the camera or the subject must move. Line scan cameras do not have shutters; they image continuously as the subject or the camera moves. Line scan cameras come in a wide variety of resolutions; at the upper end of the scale some have a resolution of 6,000 pixels. At a pixel resolution of 1 mm, this camera would cover 6 m (19.68 ft). If the camera were centered on the right lane, it would cover that entire lane, the shoulder, and a portion of the adjacent lane. At a pixel resolution of .5 mm, it would take two of these cameras to cover the same distance. Line scan cameras do not have framing rates because there is no two-dimensional frame, but they do have a data rate, which is an indication of the number of lines they can scan in 1 second. The camera mentioned above with a 6,000 pixel resolution has a data rate of 30 MHz; and it can scan 4,900 lines per second. At 20.12 m/sec, the number of lines with

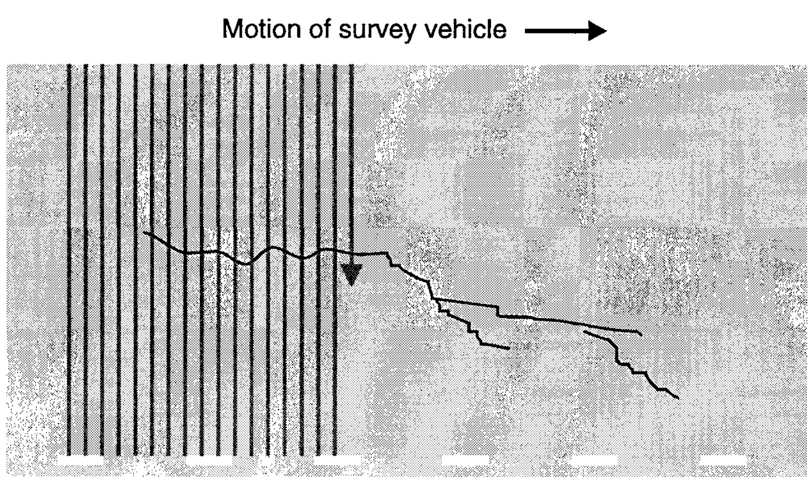


Figure 50. Building up an image using a line scan camera.

a .5 mm pixel resolution that the survey vehicle traverses in 1 second is 40,240. At 4,900 lines per second, the maximum speed of the survey vehicle would be 2.45 m/sec (5.48 mph). Despite the pixel resolution, which is more than adequate in many cases, the low data rate of many line scan cameras eliminate them from consideration for pavement surveys. There are, however, line scan cameras manufactured with more than one output, and even though they have resolutions considerably lower than 6,000 pixels, their data rate is so much higher that they are a viable alternative for imaging pavements. One model is available with a pixel resolution of 1,024 and a data rate of 60 MHz, which is capable of outputting 57,200 lines per second, which means that the maximum speed of the survey vehicle would be 28.6 m/sec (approximately 64 mph). At a 1 mm pixel resolution, four of these cameras would provide more than enough horizontal coverage; at .5 mm resolution, eight would be required. The maximum line scanning rate is determined by the data rate of a line scan camera, but below that maximum rate, these cameras vary the scan rate with the speed of the survey vehicle, so that even at speeds well below their maximum, they still scan at the same resolution. Another point in their favor is that they are considerably less expensive than high resolution framing cameras. The line scan camera mentioned above with a resolution of 1,024 pixels and a data rate of 60 Mhz retails for about $5,000.

Film Cameras

Like electronic cameras, cameras that use film can expose film in frames or in "lines." Actually, a line scan camera is an electronic version of a "slit" or "strip" camera. This type of camera uses 35 mm film on reels just as movie cameras do; however, these cameras do not contain shutters. They are ideal for photographing movement. Their use for pavement distress surveys is derived from their use in aerial photography. The motion of the survey vehicle causes an image of the pavement to move across a narrow slit in the focal plane of the lens. If the film moves continuously past this slit at the same velocity in the same direction so that it is synchronized with the image, a blur-free image of the pavement will be recorded in the direction of movement (Ray 1994). Two corporations have been using a slit cameras for pavement surveys for many years (Fundakowski et al. 1991). These systems are able to film more than a full lane width at highway speeds, and they are able to resolve cracks as small as 1 mm. As a storage medium, film has proved to be very efficient: images of about 56 km (35 mi) of pavement can be stored on a roll of film 305 m (1,000 ft) long. In many respects, this is the ideal system for doing pavement surveys. Further, there are no serious storage problems as there are with electronic cameras: for every 56 km of pavement, there will be one roll of film. The cost of developing and printing one 305 m roll of the film added to the cost of the film itself is about $600. However, although the film can be digitized for use by an automated image processing system, it is not a simple and straightforward process (Fundakowski, et al. 1991, A-4), and it is not practical to try to use it for the large number of rolls of film that would have to be processed every year. It would be ideal to have a system that could scan the developed 35 mm film automatically without losing any of the film's resolution. A slit camera and controller (the mechanism that synchronizes the speed of the film transport with the speed of the vehicle) cost between $125,000 and $150,000.

Still cameras that use film, whether they are 35 mm or medium format, have very high resolution, and they have faster frame rates than their electronic counterparts. The top-of-the-line Canon 35 mm SLR can expose 10 frames per second, and Nikon's top-of-the-line model can expose 8 frames per second. These framing rates are still inadequate for pavement surveys. However, there is another problem. The largest film back available for either of these cameras only holds film sufficient for 250 exposures. Shooting at 10 exposures per second, the film would be used up in 25 seconds.

The Connecticut Department of Transportation uses a 35 mm movie camera for photologging. This camera is a framing camera, not a slit camera. It is mounted inside the photologging van, and it is angled down the road for a full perspective view through the windshield. "The film is exposed at regular intervals of .01 mi to record pavement distress, shoulder conditions, and roadside inventory data" (Kalikiri, et al. 1994). That is a framing rate of 1 every 0.8 second. Attempts have been made to use the exposed film from the photologging as a basis for automated image processing. The Kalikiri paper documents these attempts. Since the film is exposed during the day without the use of artificial lighting, the portion of the image that is used for image processing (the lower 20% or so) has shadows and other forms of noise in it. Much of their effort has been devoted to eliminating these interferences to image analysis. They claim that the system they have developed has an overall accuracy of 90%; however, they do not mention the size of the smallest cracks they are able to resolve.

Other Hardware

Data Storage

Electronic cameras produce electronic signals, but they do not store them. A sufficient number of recorders are required to store the data coming from the camera or cameras. The collective bandwidths of all of the recorders must equal or exceed the camera data rate of all of the cameras. The total amount of storage needed on the survey vehicle will depend on how much surveying the vehicle is to perform before returning to its base of operations. For every hour the surveying vehicle surveys at 20.12 m/sec, with .5 mm pixel resolution, with full-width coverage of the lane (3.66 m), and 100% coverage down lane, there will have to be more than 987 gigabytes of storage, which is equivalent to more than 16 gigabytes per minute. A data rate this high is probably not manageable without the expenditure of a considerable sum of money. One possibility would be to use RAID disk arrays, which are essentially stacks of integrated hard disk drives that are very fast and have very large storage capacities. However, an array of these units capable of handling 987 gigabytes would cost several hundred thousand dollars.

There are a variety of ways of dealing with this problem. The requirements on which these data rates are based are the requirements of a very high resolution system. Each of the requirements directly affects the data rate. Obviously, one approach would be to relax one or more of the requirements. If the pixel resolution were set at 1 mm (rather than .5 mm), then the amount of storage needed for 1 hour of surveying would be 265,101,120,000 pixels--a reduction of 75%. If the pixel resolution were set at 1.5 mm, then the storage needed would be 117,819,790,000 pixels--a reduction of 89%. If the pixel resolution were 1 or 1.5 mm, then the width of the smallest crack that would be visible would be, respectively, 2 and 3 mm. The way the survey system is used will determine the choices here. For network level surveying, an image resolution of 3 mm may be adequate, and if an image resolution as low as 4 mm would be sufficient, then the data rate would drop to 16,568,820,000 pixels per hour--fewer than 16 gigabytes. For a network level system, it may be important to increase the speed at which the survey is taken. With this option, if the speed of the survey vehicle were increased to 22.35 m/sec (50 mph) and the image resolution were left at 4 mm, then the data rate would be 73,620,900,000 pixels (68 gigabytes). For project level work, the .5 mm pixel resolution could be maintained if the speed of the survey vehicle were reduced. A speed of 6.7 m/sec (15 mph) would put the data rate at 353,116,800,000 pixels (329 gigabytes). If the pixel resolution were also reduced to 1 mm, then the data rate would drop to 88,279,200,000 pixels (82 gigabytes), which is a reduction of 92% from the rate that resulted from meeting all of the requirements.

It seems certain that the resolution of this problem cannot be achieved by simply relaxing one or more of the requirements. There are three other approaches that can be taken: (1) the raw data could be compressed before it is stored, (2) it could be screened to eliminate that portion of it coming from pavement that is not sufficiently distressed to be of interest, or (3) the data could be converted from 8-bit to 1-bit. Each of these processes would require that some real-time image processing take place in the survey vehicle.

In an 8-bit system (gray scale, not color), each pixel has 1 of 256 gray levels. The designation for each pixel's gray level has 8-bits, i.e., a combination of eight 1s and/or 0s. The 8-bit binary representation of a gray scale of 50 is 00110010. For designations ("words") that are 8 bits in length, the memory must retain 8 bits of information. If an 8-bit data stream could be converted to a 1-bit data stream without the loss of essential information, then the amount of storage required would be reduced by 7/8ths. A 1-bit system is one in which each pixel is either black or white, i.e., the "words" consist of only 1 bit: either a 1 or a 0. Consequently, only 1 bit has to be stored rather than 8. The data stream in a pavement survey system could be converted from 256 gray levels to black and white by selecting a threshold, for example, 125, and then converting every pixel with a value of 125 or lower to black and all pixels with values above 125 to white. The difficulty here is to pick an appropriate threshold. Once the conversion to black and white takes place, there is no retrieving the gray-scale designations. If the threshold is set too low, then some of the distresses that are of interest will be included in the white background. It would be more appropriate to say that they will *disappear* into the background. If the threshold is set too high, far too much noise will be included with the real distresses, and this will make it much more difficult for the image processing system to analyze the image and accurately categorize the distresses. An appropriate threshold would have to be worked out. It is likely that different thresholds will have to be set for different types of pavements and for different surveying conditions. The data rate for a system with a 1 mm pixel resolution is 73,236,800 pixels per second. If this data stream is converted to a 1-bit stream, then the data rate would drop to 9,154,600 pixels per second or about 31 gigabytes per hour. Since the image will be black and white, there will be long sequences of pixels that have the same value; under these circumstances, it is possible to compress the data by using run-length encoding. A run-length code specifies a brightness level and the number of pixels in a sequence that have this same brightness level. So, rather than having to store a brightness designation for each pixel, it is possible to specify the brightness of a number of pixels with one code. How much compression can take place will depend on the nature of the image. In a binary image, there will be long sequences of white pixels and long sequences of black pixels; consequently, it is possible to achieve compression ratios from 4:1 to 10:1. At a minimum, it should be possible to compress the data rate of the option mentioned above to 7.75 gigabytes per hour. Finally, there is the possibility that screening algorithms could be developed that would screen the pavement data in real time and discard any data from pavement that did not show any distress, or, at any rate, did not show enough distress to be of interest. It is probably safe to say that at least 20% of the data could be discarded in this way. If so, the data rate would be down to 6.20 gigabytes per hour.

The problem of storage is a complex and difficult one. It will not be easily solved. Actually, providing sufficient storage on the survey vehicle is not as much of a problem as dealing with the data *rate*, which is extremely high. The use of standard bus architecture between the frame grabber and the storage hardware may not be possible. It may be necessary to use fiber cables to transfer data at these rates. Systems have been developed that are capable of transferring data at rates up to 400 megabytes per second using fiber optic cable.

Lighting

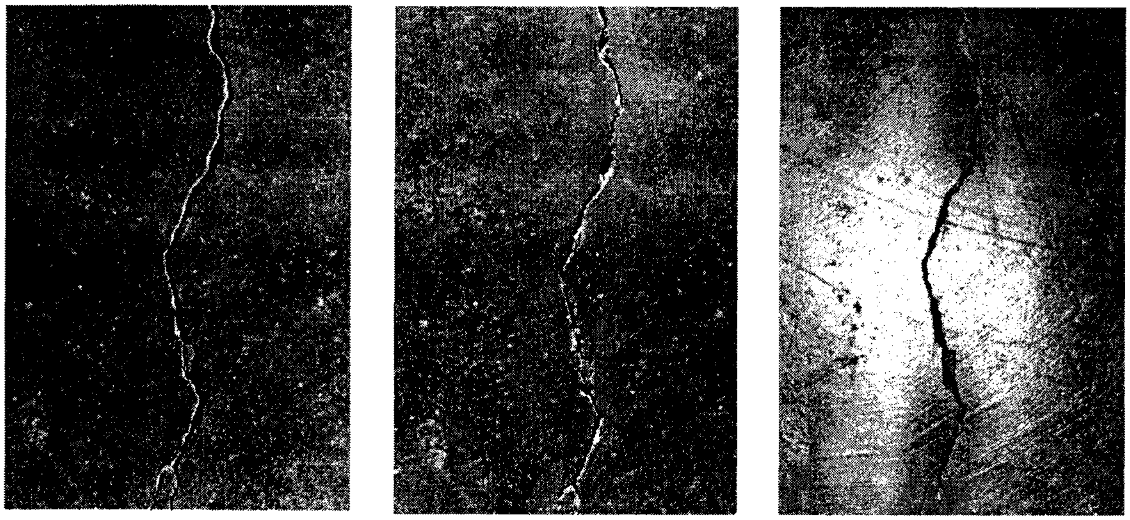
Lighting is an important determinant of the quality of the images coming from the survey vehicle. Cracks are only visible if they are darker or lighter than the pavement surface. Although cracks are usually darker than the pavement surface, in certain circumstances, they can be lighter (El-Korchi and Wittels 1990). The contrast between the crack and the pavement "depends on the depth and width of the crack, the reflectivity of the paving materials, the alignment of the crack with the light source, and the viewing direction" (El-Korchi and Wittels 1990, 75). An imaging system can use ambient light or artificial light to illuminate the pavement. Although a great deal of pavement surveying has been done under ambient light, there are problems connected with its use. If ambient light is the only source of light, the following problems will have to be dealt with or lived with:

- The brightness of the sunlight changes during the day and during the seasons.
- The angle and direction of the sunlight changes during the day and during the seasons.
- Shadows that are caused by power lines, light poles, telephone poles, signs, stop lights, trees, buildings, clouds, vehicles adjacent to the survey vehicle, the survey vehicle itself, etc. fall across the pavement. Further, the shadows will fall differently at different times of the day and at different seasons of the year, because of the differences in the angle of the sunlight.

The amount of light needed for exposure depends on the shutter speed and the aperture of the lens. If, for the moment, the aperture of the lens is assumed to be constant, then, if the shutter speed is too short for the light that is available, the image will be too dark; if the shutter speed is too long, the image will be too light. The lens controls the amount of light that is allowed to fall on the CCD or the film by adjustments in its aperture, which is simply how wide the lens opens. The shutter speed reflects how long the lens will remain open. The shutter speed and the aperture are inversely proportional for the same amount of exposure: if the proper exposure is achieved by an aperture of f2.8 and 1/60th of a second, then, if the aperture is closed down to f4 (half the light of f2.8), the shutter speed must be increased to 1/30th of a second (twice as long as 1/60th) to maintain the same exposure. Since there is only half as much of an opening for the light to enter, the light must be allowed to enter for twice as long. There is a limit to how much the aperture can control the amount of light falling on the CCD or film: *any* lens that is shot wide open for an hour in a dark room is not going to produce anything but a dark image, and a lens that is shot wide open on a pavement that has insufficient light will force the shutter speed to be so slow that the pavement image will be smeared by the motion of the camera as the survey vehicle moves down the road. If, in order to avoid smearing, the shutter speed is kept high, then there will be insufficient light for the exposure, and the image will be too dark. If the imaging of the survey vehicle is dependent on ambient light, then there is no guarantee that there will be sufficient light at all times. Of course, an image can be darker than is desired without for that reason being useless: there are stages between having the ideal amount of illumination on the pavement and having insufficient illumination for the purposes for which the image is intended. However, as the image becomes darker, the contrast between cracks and the undistressed surface of the pavement will be reduced, and this makes it more difficult for the image processing system to fulfill its objectives. Unfortunately, it would be no solution to limit the surveying to the brightest part of the day, because, apart from the other difficulties this would cause, the contrast between the part of the pavement in the bright sun and the part in shadow would be greater. The shadows created during this time of the day would be more of a problem because they would be darker. If the shadows were dark enough, it would not be possible to distinguish cracks in the shaded areas at all.

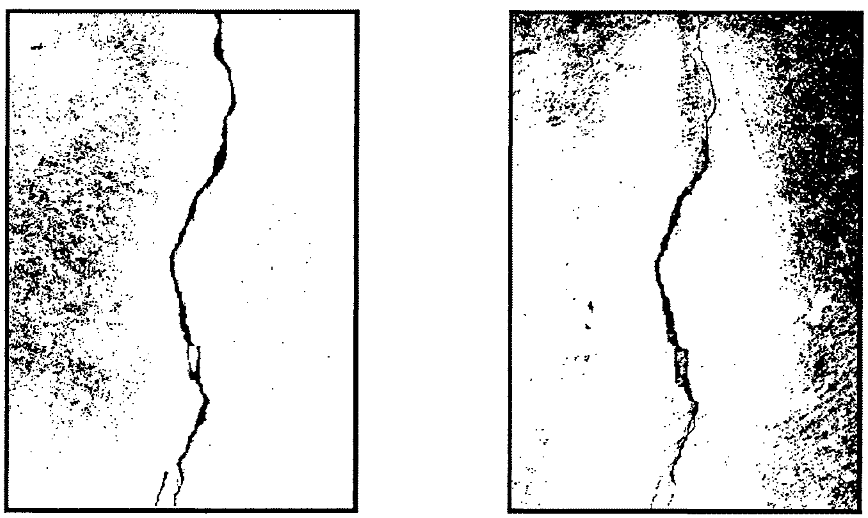
The angle that the light falls on the pavement surface as well as the direction from which it comes also have an effect on the way distresses appear (El-Korchi 1990). Figure 51 shows several images of the same crack illuminated from different directions. The angle of incidence of the light in 51a and 51b is 30 degrees, and in 51c it is 90 degrees. The light is coming from 3 o'clock in 51a, from 12 o'clock in 51b, and from directly above the crack in 51c. As can be seen, there are significant changes in its appearance. An image processing system is--at one level--going to operate off of differences in contrast, and it will "see" this crack differently depending on the lighting. Figure 52a is a binary image of 51a, and 52b is a binary image of 51c. A system that depends on ambient light for illumination will have to deal with many more changes in angle and direction than are represented by these simple images. At the very least, this will complicate the image processing, and in the worst cases, it will cause errors in identification and classification.

Shadows on the pavement are a serious problem for an image processing system. The Connecticut Department of Transportation (Kalikiri, Garrick, and Achenie 1994) and the Texas Department of Transportation (Chan, Rao, and Lytton 1992) have both had difficulties dealing with shadows in images created using ambient light. The darker the shadow, the more of a problem it is. As the shadows become darker, the more they have the same range of gray levels that cracks do. Consequently, when the gray scale image is converted to a binary image, the dark shadows will be black just like the cracks. Further processing will be required to distinguish them from the cracks. The situation is further complicated by the fact that the shadow from the same stop sign, for example, will look different if the surveying is done at a different time of day or during a different season because the angle of the sun and the direction from which it shines will be different.



a b c

Figure 51. A crack illuminated from different directions and angles. The source of light in a is at 3 o'clock, in b it is at 12 o'clock, and in c it is directly overhead.



a b

Figure 52. Binary images of a crack illuminated from different directions. The source of light in a is at 3 o'clock and in b it is at 12 o'clock.

The use of artificial lighting provides an imaging system with control over the lighting, and, to that extent, it goes a long way toward solving these problems. Studies have shown that controlled lighting is useful in minimizing the number of cracks missed by image processing systems (El-Korchi, et al. 1991). Artificial light will provide just about the same light throughout the day and night and throughout every season. Bright sun will still cause some faint shadows in some circumstances. It takes very bright artificial light to overwhelm bright sunlight and eliminate the shadows caused by it; however, light that is not strong enough to completely overwhelm sunlight can still lighten the shadows to the point that they are not a problem for the image processor. The angle, direction, and brightness of artificial light can be controlled, which means it will be very close to the same for every image whether it is taken during the day or at night. This will provide uniformity and consistency in the images and thus will contribute significantly to the consistency of the results of the image processing. Being able to control the brightness of the light also means that there will always be sufficient light to maintain the desired shutter speed.

Since framing cameras expose a two-dimensional array of CCDs while the shutter is open, it is necessary to illuminate the subject with strobes that are synchronized with the framing rate. In one respect, it is harder to provide illumination for framing cameras, because the entire area that is exposed each time the shutter trips must be illuminated at the same time. The larger the area, the harder it will be to provide lighting for it. Line scan cameras, on the other hand, only require that a narrow strip of the pavement be illuminated; however, since line scan cameras do not have shutters, the illumination must be provided continuously.

Monitors

Monitors are essential for monitoring the image acquisition process; however, they do not affect data acquisition or image processing. The resolution of monitors is not very high; consequently, the cracks that are below the resolution that a monitor can resolve will obviously not be visible. Some medical imaging monitors have resolutions as high as 2,000 x 2,000; they are, however, very expensive (approximately $30,000). High-resolution monitors that have resolutions in the range of 1,500 x 1,200 are much less expensive, and they are probably the best choice for the image acquisition system. If a single monitor with this horizontal resolution were used to monitor a pavement 3.66 m wide, the width of the smallest crack visible would be 2.44 mm. There are other options. Monitors with lower resolution could be used if a monitor were used for each camera. For example, if four monitors were used in a system that employed four cameras, and each of the monitors had a horizontal resolution of 1,024, it would be possible to resolve cracks smaller than 1 mm; however, in this case, only a fourth of the pavement would appear on any one monitor. If two monitors are used with a four-camera system, and each of the monitors has a horizontal resolution of 1,500, then it would be possible to resolve cracks as small as 1.2 mm; however, a system of this sort would require that the signals from two cameras be combined so that they could be displayed on one monitor.

CONCLUSIONS AND RECOMMENDATIONS

This report has provided an introduction to the promise--and some of the problems--of using an automated system for recognizing and classifying cracks in pavements. Developments in technology during the last 10 to 15 years have made it possible for VDOT to develop and implement such a system. Several states and several commercial operators have prototype systems in service now. These prototypes differ in many ways: there are differences in the algorithms they use, in the system hardware, in the placement of the cameras, in the types of cameras, in the resolution of the system, in the amount of coverage of the pavement, etc. Some of these systems are fully automated and some are hybrids that incorporate both automated and manual operations.

Probably the best way to describe the state-of-the-art at the present time is to say that the entire field is going through a period of intense development. The application of image processing techniques to the problem of recognizing, classifying, and measuring cracks is an area that appears to be wide open for development. During the recent past, the principal focus at research centers that have worked on this problem has been on the development of algorithms, although some have experienced serious difficulties with hardware that had to be resolved. However, many of the problems with hardware that taxed researchers in the 1970s and 1980s have been resolved by the rapid improvements in hardware since that time. Various individuals and organizations have been trying to come up with systems of algorithms that will accurately and consistently recognize and classify cracks in pavements; however, no single system of algorithms has achieved general acceptance as the best. Furthermore, there seems to be more than one way to achieve acceptable results. In many respects, however, it is difficult to assess the "success" of some of these systems because they have been tested only on a small number of images and only on a few types of cracks.

A system designed to recognize, classify, and measure cracks would consist of an image acquisition system and an image processing system. Despite the integral relation between these two components, to some extent, they could be developed separately. If an automated system were developed by VDOT, a long period of research and development would be required before pavement distress surveys could be turned over to it. One of the major failings of some of the attempts to develop algorithms is that the algorithms have been tested only on a small sample of pavement images, but one thing that can be ascertained from the literature is that the immense variability in the appearance of actual pavements would make them a far more difficult test for systems of algorithms than any sample. Before a system could be trusted to provide consistently accurate output, the algorithms would have to be tested on high resolution images of many kilometers of actual pavement. This would, of course, require that VDOT either have a suitable image acquisition vehicle in operation or at the very least that the developers of the algorithms would have access to the output of some survey vehicle even if it did not belong to VDOT. However, it would be possible to *begin* this process of development by using a large number of images of pavement in which the more important kinds of cracks are represented with as much variety as is practical.

VDOT has in its possession high resolution films of pavement sites that are a part of the long-term pavement performance study of SHRP. These films are made at regular intervals by the PASCO Corporation using a high resolution system that incorporates a 35 mm slit camera. They are of sufficiently high quality that they could be used at least in the initial stages of the development of a system of algorithms in lieu of outfitting an image acquisition vehicle to collect images, which would be expensive.

Using the PASCO films to provide a set of test images would involve the following:

- The PASCO films would be examined for suitable images of cracks. At the very beginning, the number of types of cracks would be limited: perhaps, only longitudinal and transverse cracks would be included. However, as many different variations of each of these types as is practical would be included. In other words, although many of the cracks that would be classified as transverse would be similar in shape, those that were chosen to represent the range of cracks included in the category *transverse* would include as wide a range of shapes as possible because there is no point in testing algorithms on a set of images of transverse cracks that look very much alike. It is important that the cracks look as different from each other as possible to provide the algorithms with a test that includes some at least of the variety of shapes of cracks certain to be found in a normal pavement. There should be no difficulty getting the required variety of shapes from the PASCO films, which cover many kilometers of pavement.

- Once suitable images of cracks were found on the PASCO films, those portions of the film would be digitized. At the end of this process, we would have several hundred digitized images of cracks. This collection of images would be the first set of test samples.

- This set of test samples would be used in the first stage of the development process to develop and refine algorithms that can adequately recognize and classify the longitudinal and transverse cracks in the samples. Once we had come up with a set of algorithms successful at recognizing and classifying longitudinal and transverse cracks, a second set of test samples could be culled from the PASCO films. This second set of samples would include images of different types of cracks. In this way, by continually increasing the demands on the algorithms, a set of algorithms could be developed that would be suitable for preliminary testing on the output from a survey vehicle, which would, of course, provide a much more difficult test.

There would be several advantages to proceeding in this way. First, the long process of developing a system of algorithms could commence without the need to purchase and outfit a survey vehicle. Second, the developers of the algorithms would gain experience with the image processing software and hardware on a fixed set of images, and this would be a good prelude to their application to the output of a survey vehicle. Finally, although the system of algorithms developed during this initial stage of the project would almost certainly have to be modified when it was applied to the output of a survey vehicle, there is no reason to believe that all or even a significant amount of the developmental work done during this time would not be of use in the later stages of the project. In fact, it would be better to have one or more alternative systems of algorithms ready to test on the output of a survey vehicle when it was finally outfitted and ready to undertake the distress surveys.

At some point, it would be necessary to test the algorithms on the output from a survey vehicle. There would be many options for outfitting a survey vehicle, which could also serve as a platform for a bank of lasers for determining roughness and rutting and for a standard video camera providing a windshield view of the road. The choice of a particular hardware configuration for imaging the pavement surface would depend on the needs of the user. A full specification for cameras, lighting, recorders, computer hardware, and computer software would be a complex matter, and it has not been attempted in this report. A general idea of the cost of the imaging hardware on a survey vehicle could be derived from the following. The costs listed below would be for a system capable of resolving 2 mm cracks, providing full coverage of the width of the lane and full coverage down lane, and operating at a speed of 72 km/hr (45 mph): (1) four line scan cameras: $20,000, (2) four dedicated frame grabbers: $17,000, (3) 3.66 m of sodium lighting specifically designed for line scan cameras (with a life of approximately 25,000 hr): $30,000, (4) on-board memory using a RAID system with a capacity of 90 gigabytes: $54,000, (5) Pentium PC: $3,000, and (6) two high resolution monitors (1,500 pixel horizontal resolution): $5,000. The total would be $129,000. The prices for mounting the cameras, for providing shock-absorbent stands in the vehicle for the computers, as well as other necessities for setting up the survey vehicle were not included, nor was the price for the technical assistance that will have to be purchased to get the system set up and running and to eliminate all of the bugs.

The images that the survey vehicle would make during its regular surveys would make an important contribution to the development of the algorithms, which will be a long incremental process of trial and error. It is important for these images of real pavements to be rated by human raters, otherwise it would not be possible to tell how well the image processing system was performing. Only by comparing the performance of the image processing algorithms with the ratings of human raters who had been trained to rate pavements using criteria endorsed by VDOT could the algorithms be tested. In this way, the algorithms could be evaluated by comparing the results of applying them to a set of images that had been examined for cracks by human raters. It would also be important to compare the images from the image acquisition system to the actual condition of the pavement surface. It would be essential to verify that the images accurately represented the actual condition of the pavement surface. The ultimate goal would be to provide a system that assessed pavement surfaces for cracks in the way that VDOT wanted them assessed, and this would not be possible without constantly comparing the ratings of the image processing system with the ratings of human raters trained by VDOT and by comparing the output of the image acquisition system with evaluations of actual pavements by inspectors.

An important benefit of the long-term process of developing the image processing system would be that it allowed everyone involved in the process to develop confidence in the performance of the system. Since everyone would be able to see the system's failures and the steps taken to overcome them, it would be possible to know what the system is doing and to have confidence in its performance. Incremental development of this sort would be essential to getting over uncertainties about the system's performance. Everyone involved in the development of this system in the Maintenance and Materials Divisions and at the Research Council would be kept informed of each step in the development of the algorithms, which would be essential to ensuring that VDOT would have confidence in the system when it is fully implemented.

Although it would be possible to develop a survey system at the Research Council, it may be more practical to contract for pavement distress surveys with one or more of the commercial concerns that have developed the capability to perform pavement surveys using image analysis. In this case, the Research Council could take on the role of evaluating the accuracy of the surveys and help to customize the algorithms of the contracted survey system so that they are more suitable to the pavements in Virginia. A decision with respect to further research in this area will depend on support from the Maintenance Division and the availability of resources.

REFERENCES

Baxes, Gregory A. 1994. *Digital Image Processing*. New York: John Wiley & Sons.

Burke, Michael W. 1992. The PAVUE real-time automated pavement distress analyzer. Paper presented in the pavements section of the 1992 TAC Annual Conference in Quebec City, Canada, September 14-17, 1992.

Castleman, Kenneth R. 1996. *Digital Image Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Chan, Paul; Ashok Rao; Lan Li; and Robert Lytton. 1992. Development of image algorithms for automated pavement distress evaluation system (Report 1189-2F). College Station, Texas: Texas Transportation Institute.

Chan, Paul; Ashok Rao; and Robert Lytton. 1992. System hardware for acquisition and automatic processing of pavement distress videolog (Report No. 1189-1). College Station, Texas: Texas Transportation Institute.

El-Korchi, Tahar and Norman Wittels. 1990. Visual appearance of surface distress in PCC pavements: I. Crack luminance. II. Crack modeling. Transportation Research Record No. 1260. Washington, D.C.: Transportation Research Board.

El-Korchi, Tahar; Michael A. Gennert; Matthew O. Ward; and Norman Wittels. 1991. Lighting design for automated pavement surface distress evaluation. Transportation Research Record No. 1311. Washington, D.C.: Transportation Research Board.

Fundakowski, R. A.; R. K. Graber; R. C. Fitch; E. L. Skok; and E. O. Lukanen. 1991. Video image processing for evaluating pavement surface distress. Washington, D.C.: Transportation Research Board: National Cooperative Highway Research Program.

Galbiati, Jr., Louis J. 1990. *Machine Vision and Digital Image Processing Fundamentals*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Haralick, Robert M. and Linda G. Shapiro. 1985. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing* 29:100--132.

James, Mike. 1988. *Pattern Recognition.* New York: John Wiley & Sons.

Kalikiri, Vinod K.; Norman W. Garrick; and Luke E. K. Achenie. 1994. Image-processing methods for automated distress evaluation. Transportation Research Record No. 1435. Washington, D.C.: Transportation Research Board.

Li, Lan; Paul Chan; and Robert Lytton. 1991. Detection of thin cracks on noisy pavement images. Transportation Research Record No. 1311. Washington, D.C.: Transportation Research Board.

Mahler, David S. and Zuhair B. Kharoufa. 1990. Image processing for a pavement crack monitor (Report number: SHRP-ID/FR-90-001). Washington, D.C.: National Research Council: Strategic Highway Research Program.

Ray, Sidney F. 1994. *Applied Photographic Optics.* Oxford, England: Focal Press.

Russ, John C. 1995. *The Image Processing Handbook.* Boca Raton, Florida: CRC Press.

Strategic Highway Research Program (SHRP). 1993. *Distress Identification Manual for the Long-Term Pavement Performance Project.* Washington, D. C.: Transportation Research Board.