An Investigation into the Coding of Halftone Pictures

by

Yao-Ming Chao

S.B., National Taiwan University
(1971)

S.M., Massachusetts Institute of Technology
(1976)

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May, 1982

Signature of Author ______________________
Department of Electrical Engineering
and Computer Science, May 18, 1982

Certified by ______________________
William F. Schreiber
Thesis Supervisor

Accepted by ______________________
Arthur C. Smith
Chairman, Departmental Committee
on Graduate Students

AN INVESTIGATION INTO THE CODING OF HALFTONE PICTURES

by

YAO-MING CHAO

Submitted to the Department of Electrical Engineering and Computer Science on May 21, 1982 in partial fulfillment of the requirements for the Degree of Doctor of Science.

ABSTRACT

Characteristics of different methods in the binary representation of contone pictures are investigated. A theoretical limitation for such representations is derived.

The difficult problem of rescanning a halftone picture is solved by implementing a dot size averaging technique. Moire patterns can be completely eliminated. Fine details, which otherwise might be attenuated in the dot size averaging process, are preserved by using an adaptive algorithm to locate and extract them.

The two approaches to encode a halftone picture are discussed in order to improve the efficiency of coding schemes for halftone pictures. The coding-before-screening approach requires the knowledge of a sampling density for the contone original which can then be used to reproduce halftone with the same quality as directly from the original. Experimental results show this sampling density to be dependent on the characteristics of the originals. A sampling density of 180 lpi is adequate for almost all pictures produced from screen of up to the 100 dots per inch. For lower screen densities and picture of low spatial frequency content, the sampling density can be as low as 90 lpi with little discernible differences.

Two coding-after-screening methods are developed to improve the statistics of the binary pictures such that run-length encoding can be applied more efficiently. The first one uses multi-pel predictors while the second one decomposes an image into two parts and then applies different schemes for each part. When short runs are skipped and line interpolation is used, compression ratios as high as 12 with respect to the original halftone can be achieved with little degradation in picture quality. Merits of the coding-before-screening and the coding-after-screening approaches are then compared.

Thesis Supervisor: Dr. William F. Schreiber
Title: Gordon Professor of Electrical Engineering

ACKNOWLEDGEMENT

I wish to express my profound gratitude to professor William F. Schreiber, my thesis supervisor. He not only introduced me to the filed of image processing, but also suggested the topic and guided me through the entire course of this research. Without his guidance and his patience in reading and correcting my draft, this thesis would never be done. His concern and consideration helped me through those days when life seemed so difficult to me.

Professor Donald Troxel and Professor Jae Lim served on my thesis committee as readers. Their careful corrections and comments on the draft are highly appreciated. Special thanks to Professor Troxel for his constant technical support while I was with CIPG. I must thank Professor Lim for his constant supply of new ideas from the digital signal processing point of view which provided a lot of stimulus.

I am indebted to Professor Richard Adler. In the early days while I was studying in MIT, he helped me a lot in the graduate program and so many times wrote recommendations on my behalf.

I also want to thank Don Levinstone and Ted Kuklinski. Their constant help in correcting my English and discussion on the ideas are extremely beneficial. I am also grateful to my officemate Malik Khan. During the time we have been together, he shared a lot of experiences with me, academical as well as cultural. There were many a valuable discussion concerning this work and other related issues.

Among those who provide the excellent computer facility support, I especially wish to thank Len Picard, Jason Sara, Mike McIlrath, Russ Walker, Webster Dove, and Tony Drogaris. They are a superb team and I had the pleasure to work with them. I must also thank Marilyn Pierce and Ginny Chambers for their help in taking care of the department bureaucracy.

Financially, this research is supported partially by the Providence Gravure Company, partially by the Chinese Ministry of Education and partially by the Grass Fellowship. I am indebted for their support.

To

My Parents

TABLE OF CONTENTS:

LIST OF FIGURES:

Introduction

The basic purpose of this thesis is to get a better understanding of the nature of halftone pictures and to develop a reasonable coding scheme for their transmission or storage. Chapter one gives a short history of the development and use of halftone pictures that inspired this research. Chapter two is dedicated to the characteristics of visual systems and the mechanism of bi-level representation of a contone picture, including methods of generation. An important subject in this chapter concerns the limitation of the halftone process.

In chapter three, a halftone picture suitable for printing purposes is discussed. Conversion between the two is treated in considerable detail. The frequency domain representation for generalized halftone is also developed and is included in the appendix. Currently popular coding schemes for both contone and bi-level pictures are reviewed briefly in chapter four. The special problems in coding a halftone picture are also discussed.

Chapter five and six discuss the method of coding halftone pictures. Two approaches of encoding a halftone picture are investigated. The first approach encodes the original contone and converts it to a halftone at the receiving site while the second approach screens the original first and then encodes it. Due to the nature of the screening process and also the perception of the human eyes, some features in a contone original probably can not be reproduced in a halftone. In chapter five, ways and means are found to set various limitations (number of grey levels, sampling density etc) in the preparation of contone pictures. Halftone pictures obtained from the contone so prepared are totally indistinguishable by human eyes when compared with halftones screened from the original contone. Information gathered in this chapter is used to compare with another approach in coding the halftone picture. In chapter six, two methods to arrange binary pels are used to improve the coding efficiency of run-length encoding by increasing the length of runs. Information about the screen, from which the halftone is made, is used to extract more redundancy and thus a better compression ratio can be achieved.

Results for the two approaches along with the merits of each method are discussed in chapter seven along with suggestions for further research on halftone pictures.

CHAPTER 1

A Short History and Statement of Objectives

Ever since the invention of the printing process, illustrations have been an increasingly important part of printing. Due to the special properties of the printing process, pictures always posed a big challenge. Simple drawings which use only fine lines (so called line art) can be reproduced rather satisfactorily. Intaglio was the early method of producing graphics on the printing materials where lines can be used to produce the perception of different shades of grey if the spacing of these lines is varied in appropriate ways. The halftone process has been used for more than a century. Basically, it converts a continuous tone picture into a regular pattern of dots that can then be printed. The size of each dot is related to the tone at that same place in the original picture being reproduced. Most of today's newspapers as well as books print using this process to produce pictures.

History of halftone process goes back to the last century. An English physicist named W.H. Fox Talbot was the first one to conceive the idea of using different sizes of minute dots to obtain the grey value for a contone pictures. He used a very primitive form of contact halftone screen by placing a screen of black gauze between the sensitized plate and the positive from which he printed and exposed the resulting sandwich. In 1855, A. J. Berthold was granted a patent for a practical screen, composed of a single set of parallel lines. George Meissenback successfully applied the technique to reproducing photographs, using a parallel-line screen. He turned the screen after one half of the exposure had been given, thereby gaining in principle at least the effect of a crossline screen. In 1885 in Philadelphia Frederick E. Ives produced the first mechanically ruled crossline screen. He scribed parallel lines in the opaque film of exposed and developed wet collodion plates. Two such plates were then cemented together with Canada balsam and the direction of ruled lines crossed at right angles to form transparent squares. Max Levy finally introduced ruled and etched screens for halftone photography. The first Levy screens were made in 1886, using the Ives scheme. The firm of Max Levy is still in business in Philadelphia making halftone screens and other products requiring extremely accurate rulings. Modern glass halftone screens are made from two pieces of clear, finely polished glass, each ruled with a set of very accurately spaced parallel lines. The two glass plates are cemented together with the ruled surfaces adjacent and with the rulings at exact right angles. Normally, the lines and spaces between them are of equal width. A competitive process used by one maker employed crossed rulings from a single piece of glass. The slightly round crossline intersections is claimed to produce smoother portrayal of midtones. The above mentioned screens have the problem of maintaining accurate screen distance. This difficulty was eliminated by the development of the contact screen, where the screen is in direct contact with the plate. The remaining difficulty is how to hold the screen and plate tightly together. High contrast film is placed behind the screen. An image is then projected onto the plate. The cross line of the screen creates a modulation effect on the image. The density of light on the screen has the average grey value of the image but with variation at the frequency of cross-lined screen. This fast varying image is then further enhanced by the high contrast film to create a binary picture. A printing plate can then be etched using the film as a mask.

Finally, in this modern electronic age, electronic screening without any physical contact screen is used. Specifically, an electronic memory holding the screen pattern that would be used in the contact or glass cases is implemented in order to do the thresholding and create a structured binary pictures similar to that produced by a contact screen and high contrast film. Electronic screening has the advantage of great flexibility. With the same hardware (i.e., memory, scanner etc), different densities and screen shapes can be implemented by simply changing the screen structure that is stored in the memory. The sampling density required is relatively high in order to obtain a sufficient number of apparent grey levels, especially when a high density screen is used. Other than the easiness in changing both size and shape of the screen pattern, the electronic screen is also easier to incorporate some signal processing into the screening, such as done by Kato [47] to enhance the picture and/or modify tone scale. Because of the flexibility and the availability of the Autokon electronic process camera, the electronic screen is used throughout this thesis.

Although the halftone process preserves most of the pictorial information in the continuous tone original, it does result in some degradation. Low spatial frequencies are usually retained. Some kinds of non-linear transformation of the grey level is performed during the process. Also the highest reproducible spatial frequency is not limited to the dot size. Inherently, the halftone pictures have only two levels. At the area where ink is applied a black patch is produced. On the other hand in the area that remains blank, a white patch is represented. During printing, ink is applied to the printing plate and then transferred to paper. The thickness of the ink layer, the force applied between plate and paper, and the surface smoothness limit the smallest white and black areas. A thin ink layer, light force and smooth paper surface result in a smaller area which can be printed, which in turn increases the fidelity of the halftone pictures.

The printing plate makes contact with the paper surface for each copy. Printing dots which are too small will eventually be worn out. Tollenar and Ernst [93] have the experimental data for this effect. The typical value of minimum dot size can be found accordingly.

Of increasing importance to the graphic arts and communications industries is the manipulation of large quantities of digitized graphic art images, such as line drawings and combined pages of text and art. Such images abound in printing and publishing and in facsimile operations. As the demands on communication channels increase and the costs of digital processing decreases dramatically with passing time, it becomes more and more desirable to represent this pictorial information efficiently. By efficiency, we mean both use of as few bits as is appropriate and specification of an information structure to which the reconstruction procedure may be easily referred. Encoding and decoding costs should be assumed to be relatively small, perhaps even negligible when compared to the transmission or storage-retrieval cost.

Automation of the newspaper printing process has been very successful recently. Text editing systems are now commercially available. Full page editing is also implemented except, in most cases, for the use of simple drawing forms instead of relatively complex pictures. The next step lies in the development of systems which can create a whole page plate ready for printing and furthermore to transmit this full page plate to distributed printing locations from a central editing office. As computer hardware becomes cheaper and its speed faster, the use of a computer to achieve this goal is a very logical approach. There are some obstacles yet to be overcome if the goal of full page editing with mixed text and pictures is to be achieved. For decentralized printing, the main body of the material is edited in the central office and then transmitted to local printing plants where plates are made and material is printed. Generally speaking, if many pictures are used in the pages, the storage space required will be increased tremendously due to inefficient representation of pictorial information. In turn, the cost and time of transmission could be excessive due to the cost of transmitting media and the large volume of information in the newspaper. The cost of on-line computer storage also becomes expensive if a large volume of information stays on line for a long time. As a result, only very simple forms of pictures can be used as in the case of Wall Street Journal, where a very limited number of simple line art images are used as illustrations.

To overcome this obstacle, appropriate data compression is required. For the text, the ASCII form can encode the text and transmit it efficiently. For typical newspaper body type, less than 200 bytes are sufficient to transmit a square inch of text. The typesetting machine can use the encoded data to produce the desired printed plate. The bottleneck lies in the pictorial material. To faithfully send a picture in uncoded form (usually called PCM), it would take one half to two million bytes per square inch depending on the quality of the original picture to be sent and the desired quality to be received. If data are transmitted by a channel with capacity of 56 kbits per second (this capacity channel is cheaper than video channel but more expensive than a direct dial phone line), a full page containing 15 % of pictures would take 2 minutes and 30 seconds to transmit, as compared to only a few seconds for a whole page of text. It can be easily seen that the small pictorial area contains a high volume of data. Even with the old saying that "a picture is worth a thousand words," this is still too high a price to pay and sometimes makes the implementation impractical.

At the present time, continuous tone pictures and bi-level facsimile have been extensively investigated both in space and spatial frequency domains. Many practical and efficient coding schemes have been developed and implemented. On the other hand the halftone pictures which form the major part of the the printed pictorial information have rarely been considered.

As a result, the transmission of pictures suitable for printing is still a problem due to the extremely large amount of information required. To improve the transmission time as well as requirement of on-line storage space for decentralized printing, efficient schemes for the coding of halftone pictures must be found.

Investigation of the properties of halftone pictures and use of this information to achieve a better coding scheme comprise the main body of this thesis. That includes the resolution and contrast compromise, the psychophysical effect of a halftone process, the information content, Moire pattern etc. Better coding schemes for halftone pictures are implemented that not only result in good picture quality but also a high compression ratio.

To store the picture information in the computer and retrieve it for later use, there are two aspects that must be taken into consideration. First, picture information should be converted in an efficient way such that it takes small space for storage and needs the least amount of time for transmission. Second is the achievement of this goal using as simple an algorithm as possible and yet retain as much useful information as possible. Certainly, there must be some trade-off between these two goals. In order to fulfill this purpose, experiments have been conducted to gather the desired information about bi-level pictures.

CHAPTER 2

Binary Representation of Contone Pictures

2.1. Characteristics of Visual Perception

Since the final objective of this thesis is to find an efficient coding scheme for halftone pictures, it is important to understand characteristics of the human visual system, i.e., how the binary image is perceived by eyes, what are the limits of binary representation etc. The performance of the human visual system (HVS) has been the subject of research for a long time. Basically, the HVS is comprised of two parts: the optical system such as lens, iris, etc. and the sensory nervous system. A more detailed description of the visual system can be found in Cornsweet [11]. As we know, the performance of the HVS is subjective, nonlinear and with large intrinsic deviation in measurement. One of the best ways to investigate such systems is through the use of describing function which gives the response curve for sinusoidal inputs. The optical system of eyes is similar to a typical optical system and has similar limitations. A common way to describe the performance of optical systems is through the measurement of the "modulation transfer function" or MTF in acronym. Some of the characteristics of the ordinary optical system can also be found in human eyes. For example, lenses usually have a limit on resolving fine detail. At a high spatial frequency, the image cannot be reproduced as well as at a low frequency. This is also the case with the visual system. As finer detail is presented in a picture, human eyes tend to fail to resolve them. In other words, we can not see the detail as well as in plain areas. This roll-off makes the eye act like an integrator at higher spatial frequencies. In addition to the optical system, there is also the psychophysical limitation due to the overall sensing and transmission of the visual information via the nervous system. This psychophysical effect essentially rolls off the sensitivity at low spatial frequencies and makes the eye act like a differentiator in that range.

There are various methods to obtain the MTF of human eyes. Cornsweet has listed the most commonly used methods, namely, the direct measurement by Davidson [13], the threshold measurement by Van Nes and Bouman [95], and the contrast match method by Davison [12]. The direct method asks the subjects to match the grey level of a small portion of an alternating pattern to a grey scale. The period of the alternating pattern is changed to get readings for different spatial frequencies. The threshold method simply adjusts the magnitude of the alternating pattern until subjects cannot tell whether it is a uniform grey scale or an alternating pattern. The contrast match method asks the subjects to change the magnitude of the alternating pattern until they see the same contrast as in the comparison pattern. These experiments are quite subjective and a linearized model of a highly nonlinear system - human eyes - seems to be simplistic. However, the measurements obtained from these different methods are quite similar in shape and give a good explanation for some visual effects such as the Mach band phenomenon.

The typical MTF of human eyes is shown in Figure 2.1. This characteristic of the visual system can be translated into the limitation of a perceptible number of grey levels versus spatial frequency. In an area with low frequency structure we can distinguish many more grey levels than in a fine detailed area. As the content of detail increases, the sensitivity of the eyes also drops and makes it harder to discern small changes in grey levels. A larger step change in grey value is required to enable the eyes to perceive the difference. Actually, in the plain area of a picture, the number of grey levels is important but in the highly detailed area (e.g. at a sharp edge) it is the texture rather than the exact grey level that is significant and must be reproduced as well as possible.

The MTF curve can be fit with a compact mathematical form. Dooley[15] has fit the MTF curve of human eyes into closed mathematical formula.

$$\# \text{ of levels} = 1010(\exp-[0.69f])(1 - \exp-[0.5f]) + 1 \quad (1)$$

where f is the spatial frequency measured in line per inch.

In this form, it is much easier to compute and compare results with experimental data. Roetling [77] adopted this mathematical form to calculate a theoretical limit on the number of bits needed to represent pictorial information, using the redundancy in human visual systems due to the roll-off at high spatial frequencies. At a sampling grids of 500 samples per inch and viewing distance of 12 inches, a theoretical limit of 2.8 bits per pel is obtained when compared to the typical 8 bits per pel contone representation. The same method can be used to find the sampling grid where one bit per pel date rate can be reached theoretically. This is found to be at a sampling density of 850 pels per inch when the picture is viewed from 12 inches away. These numbers are, of course, only theoretical values. There is no guarantee that an optimal coding scheme can be practically implemented to achieve the above rates. The same MTF can be translated into the number of grey levels versus spatial frequency curve and is used in the next section to derive the limit of a binary representation.

2.2. Limitation of Binary Representation

The type of picture we are more familiar with is the contone. In such pictures one can find different shades (or grey levels) in each picture element. It is often the case that we have to display these pictures on devices that can produce only two tones. For example, in a plasma panel display, controls are provided to turn each pel on or off. The brightness of each pel cannot be modulated. Also in most of the printing processes (with the exception of gravure), the apparent grey level is determined by the size of the area in which ink is applied rather than the thickness of the ink layer. Control in transfering the ink to the paper is done through etching of the plate. By arranging the binary pels carefully, the texture and shade of the picture can be reproduced quite accurately.

The reason why a binary picture can look like a contone is explained by Roetling, using the similarity of the visual performance curve and the obtainable grey level versus the spatial frequency curve of the binary pictures. A more convincing and intuitive explanation is from the point of view of the spatial frequency domain. By looking at a halftone picture from the spatial frequency viewpoint, one can find that the spectrum consists of a main lobe which contains the information of the original contone picture plus some repetitive peaks at the harmonics of a two-dimensional dot structure [50]. The visual system acts like a filter that attenuates the harmonics and perceives the main lobe that holds the information of the original contone. Human eyes are much more sensitive to abrupt changes in shape, brightness, etc. than the uniform area of a regular object. For example, regularly spaced dots tend to be perceived as uniform grey if we do not intentionally look for the dot structure. When this effect is combined with the fact that size and spacing of those dots which are small enough so that the harmonics of the dot structure are at a relativly high spatial frequency, they will be heavily attenuated by the visual system. This explains why a binary picture can look as good as a contone.

The first question which arises is "How small does the sampling grid need to be in order to render a contone picture perfectly by a bi-level display?" By perfect, we mean that there is no perceived difference if the pictures are viewed side by side, at the same distance. Although human visual systems cannot distinguish more than two grey levels of high resolution testing patterns, with more than 900 lines per inch at 12 inches of viewing distance, it does not mean that simply assigning one bit of data to each pel can produce a binary picture as good as its contone original at the same resolution. Actually, using output devices such as the Autokon, at 722 lpi resolution, by assigning one bit to each pel, the picture is not really as good as the the original when viewed at the distance of 15 inches. This is true even when optimized algorithms such as the contone screen (to be discussed in the next section) are used. One explanation can be obtained from the visual performance curve and the performance curve of the binary picture.

The performance curve of a binary picture is derived by Roetling [77]. It is shown to be:

$$\# \text{ levels} = \left(\frac{1}{2f}\right)^2\left(\frac{1}{\Delta^2}\right) + 1 \qquad (2)$$

where f is the spatial frequency and Δ is sampling grid size.

The relationship of these two curves can be seen in Figure 2.1. With a coarser sampling grid, the binary performance curve rolls off sharply at relatively low spatial frequencies. In the last section, we mentioned the 1 bit per pel coding efficiency for a theoretical limit at 860 samples per inch. It actually needs a complex coding scheme to get close to the above rate. If one bit is used to represent the state of each pel, much denser sampling is needed to produce the same quality as the original.

When the sampling grid in a digitized picture becomes smaller, it also becomes more difficult to see each individual pel. When such small pels are used to represent the grey level of a small area, it creates fewer artifacts. Theoretically, if this process continues, the binary pels will eventually be so small that they are no longer visible to the eye. Not only is the single binary pel not visible, but clusters of these binary pels that create the different shades will also be uniform to hide any defect due to this binary representation. Above that point, the binary picture is no different from its contone original as far as the eye perception is concerned. Human visual performance, as well as the characteristics of the binary picture determine this limit. The coarsest sampling grid to achieve the above effect will be found based on the visual model by Dooley [15] and the binary performance derived by Roetling [77].

The performance curve of binary representation has been mentioned before. The number of reproducible grey levels is shown in equation (2). For a coarse sampling grid, the performance curve will intercept the visual performance curve at two points, as indicated by A and B in Figure 2.1. Between the two interception points, the visual performance curve is above the binary performance curve. This indicates there are some signals visible to eyes which are unable to be reproduced by a binary representation. In other words, there will be differences between the original and the binary representation. If the sampling grid becomes smaller, the shape of the binary performance curve remains the same. However the entire curve is now moved toward the higher spatial frequency side. The interception points marked C and D are now closer to each other, indicating that fewer signals are visible but not reproducible. Finally, when the sampling grid is small enough, the two curves will be tangential to

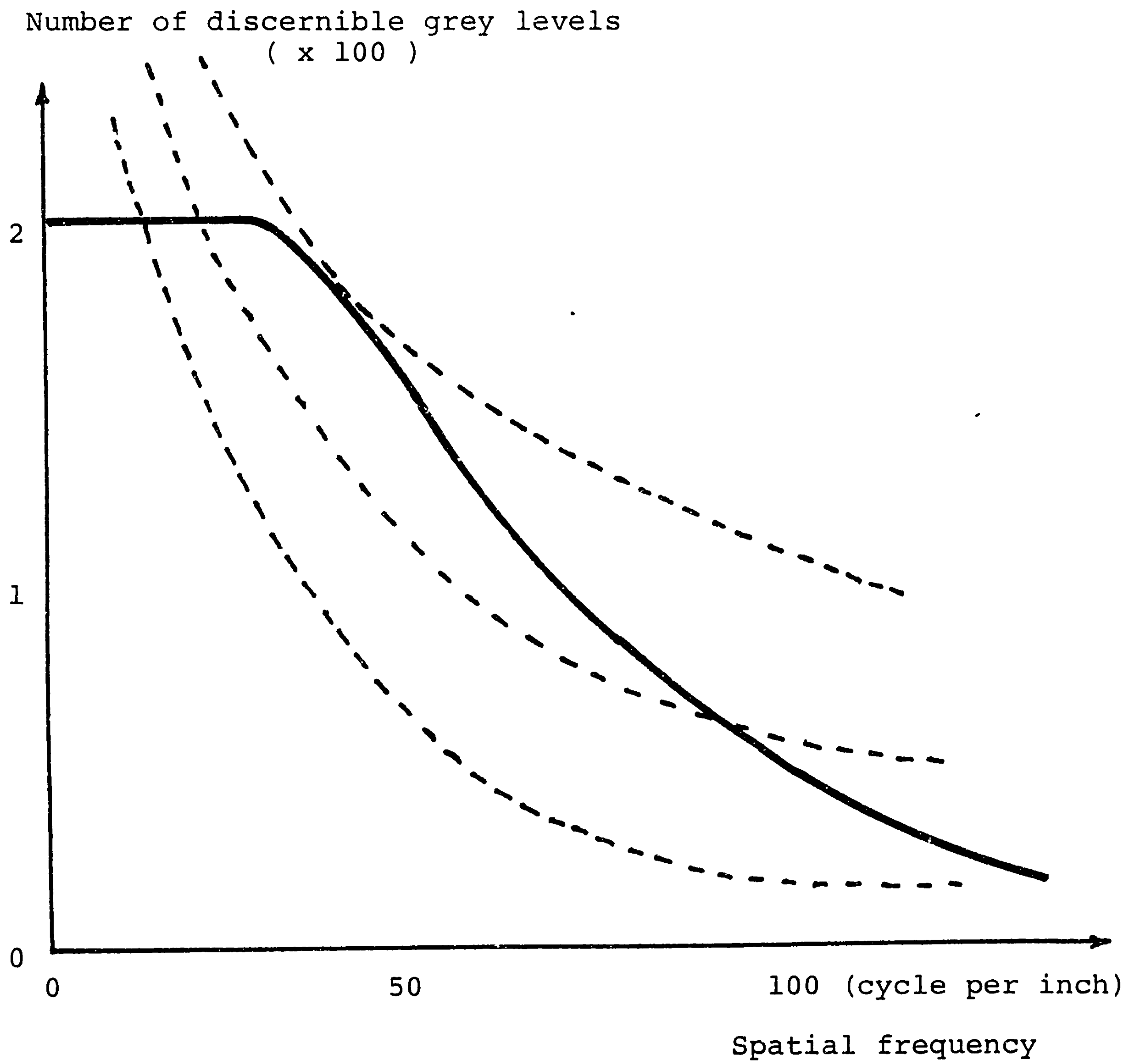


Figure 2.1 Performance of visual system and bi-level display.

each other. At that point, there will be no signal which is visible but not binary reproducible. The sampling grid at which this happens is the maximum grid that invariably generates a perfect binary picture. To find this value, the transcendental equation must be solved.

$$\frac{d^2}{df^2}\{(\frac{1}{2f})^2(\frac{1}{\Delta^2}) - 1010(\exp-0.69f)(1 - \exp-0.5f)\} = 0 \quad (3)$$

This value is found by solving equation (3) numerically. It is 1160 lpi at a 15 inch viewing distance. Note this value depends on the viewing distance. For a 12 inches viewing distance, the value can be scaled up to 1460 samples per inch.

2.3. Method of Representation

In the last section, the capability of binary display systems has been discussed. In this section, the actual implementation will be reviewed and compared. A more detailed discussion on halftones suitable for printing purpose will be included in the next chapter. The term "texture coding" is used to describe the representation of pictures based on only two levels (or one bit of information) for each picture element.

In principle, a binary picture is obtained by thresholding an original contone picture with a screen pattern. Depending on the grey level and the screen magnitude at a particular location, it produces either a black or a white pel for output. Due to the lack of grey levels, the binary representation has some inherent degradation. The next problem is to find a screen pattern such that the resulting binary picture has the desired characteristics. The objective is to minimize degradation due to the lack of grey levels.

There are some major concerns in reproducing pictures using binary output devices. The first one is whether a large enough number of apparent grey levels can be created. This factor determines the low frequency properties of a binary picture, i.e., areas without much detail. If the number of grey levels that can be produced is not large enough, then abrupt jumps between levels will be quite visible. This is the familiar contouring effect in the plain areas. The second concern is the capability to produce fine detail textures. A good thresholding method should produce fine detail as sharply as possible. The third one is printability. To display pictures with devices such as the Autokon and plasma panel, where each pel can be reproduced accurately in an "ON" and "OFF" fashion, the only concern is to arrange the pels to reproduce the best appearing picture. Unfortunately, this is not the case for most printing processes which are used to create illustrations in books or newspapers. It is natural to give special consideration to the capabilities and problems in the printing process. All methods used in printing, with the exception of gravure, have difficulties in accurately modulating the amount of ink to be transferred to each picture element on paper to create different shades. In order to achieve the apparent grey levels, the halftone process is used, which enables pictures to be printed with consistency. In letterpress printing, the amount of ink transferred to paper is controlled by the area where ink will be transferred. The smallest dot size is required to print a white patch. Were it not for these minimal dots, the paper would be in direct contact with printing plate and a black patch would be printed. To get higher dynamic range, smaller minimal dots are desired. But as the minimal dots become smaller, they are also easily worn out. The result is a change in tone scale as more copies are printed from the same plate. Ink running and wear on the printing plate which cause errors in dot size are the basic processes which create errors in reproducing the halftone dots that in turn will cause grey scale distortion. Yule [99] described the limiting screen frequency of halftone as being set by correct grey scale reproduction.

Another problem is the smoothness of the paper surface. For a smooth surface, ink can be transferred from plate to paper accurately. With rough surface, this will not be the case. When the dot is too small, ink on the dot may not be transferred at all, because it may not be in contact with the paper surface. Due to the characteristics of the ink transfer to paper and the wearing problem on the printing plate, a minimum dot size is usually specified in order to produce consistent halftone pictures. For a rough surface, a coarser screen should also be used. Usually, when a high density screen is used, it requires a smoother paper surface in order to print accurate and consistent tone scale. For an extremely fine screen, specially coated paper should be used. This extra constraint limits the binary display from obtaining its best possible visual quality. Other than the above major consideration, there may be some concern about the complexity of the implementation. But currently, as hardware becomes cheaper as compared with other development costs, this no longer seems of major concern.

There have been some recent developments in bi-level representation, both in display systems and in encoding schemes. Knowlton [52] has developed a regular noise pattern to avoid the low spatial frequency noise as introduced by the Roberts pseudo random noise scheme [74]. He also tried to preserve the high contrast detail at cell-size resolution. Techniques to generate halftone pictures on computer terminals have been implemented by Arnemann [4], Schroeder [86],[87], Phillips [70] and Rossol [80]. Arnemann was concerned with the display on a graphic terminal and he used run-length coding to reduce the memory requirement. Schroeder and Rossol used a microfilm plotter as the output device for halftone pictures and Phillip uses an ink jet machine. More recent developments can be found in the application of ordered dither to a plasma panel by Judice et al [43][44]. All literature mentioned above are mainly concerned with the elimination of contours due to the limited number of grey levels, in this case only two.

Accompanying the schemes mentioned above is the grainy appearance due to the introduction of pseudorandom noise. Dither has been used for a long time to eliminate contouring effects due to the lack of grey levels. Roberts [74] applied the technique of adding and subtracting pseudo-random noise that has a very long period of repetition. From the spatial frequency point of view, such pseudo-random noise has low frequency components as well as high frequency components (the actual spectrum is somewhat flat over the frequency of interest). The low frequency noise is visible because of this characteristic. On the other hand, ordered dither which has a shorter period of repetition and hence contains only high spatial frequency components, is much less visible to human eyes. Limb [55] has applied this technique and the result is quite satisfactory. Jarvis et al [40] have listed methods for the binary picture representation. More recently, Stoeffel [90] has reviewed those methods as well as some newer algorithms using the criteria mentioned above. In short, the current methods can be categorized in one of the seven ways listed below:

2.3.1. Fixed Level Thresholding

In this method, grey values of picture elements in the contone picture are compared with a fixed value. If the grey level is greater than the threshold value, a white pel is generated, otherwise a black pel is created. This method can produce high contrast fine detail but lacks the capability of reproducing grey scales. The contouring effect is the most severe among all algorithms.

2.3.2. Adaptive Thresholding

Instead of fixing the threshold value at a number, this method applies a constantly adjusted value for thresholding according to the grey value of the local area [61][39]. Compared to the last method, more grey levels can be reproduced due to this constantly changing threshold value. The fine detail is not as good however. Also, depending on how the adaptive algorithm is used, the picture quality can vary substantially. The method of Morrin is one way to represent a continuous tone picture, where dynamic thresholding and edge detection are used. Text pages and data like finger prints can be successfully represented, but so far as an ordinary picture is concerned, the quality is no match to that of the conventional halftone. Some features of the picture are conserved and enhanced making this method useful for certain special purposes. It consists of some preprocessing (edge enhancement) and data compression. Although serving very well as a tool for pattern recognition, it seems ruled out for printing applications.

2.3.3. Orthographic Tone Scale Creation

This method measures the average grey value of a contone picture in a small area, and then uses this average grey level to assign the polarity of binary pels in an nXn area. A pseudo halftone system has also been developed at IBM for a facsimile machines by Jarvis [38] and for a printers by Grad [24] where the brightness of an area is simply represented by pels arranged in some order. A similar method has been patented by Behane [7]. There is a pattern for each corresponding grey level. The contouring effect can be eliminated by using a large enough area for each grey value assignment. Since an average grey value is used over the entire nXn area, small variations in grey level between pels can not be fully accounted for, resulting in loss of fine detail in this method.

2.3.4. Pseudorandom Thresholding

The Roberts pseudo-random noise method is effective in getting rid of the contouring effect due to limited number of grey levels. When this method is applied to a 1 bit per pel case, we have pseudo-random thresholding, which is sometimes called a random screen. The random screen is a sequence of random numbers, more exactly a pseudo-random number sequence with a very long period. To get rid of the contouring effect due to the lack of grey levels, pseudo-random noise can be added and then the resulting signals are thresholded with a fixed value. We can also think of this process as comparing the signal to a two-dimensional pseudo-random screen pattern. The polarity of each pel output is thus determined in a probabilistic fashion. The larger the grey value is, the more probable the pel will be turned on and vice versa. The disadvantage of using the long sequence pseudorandom threshold screen is the noise effect, especially at low spatial frequencies. Human eyes are very sensitive to this kind of noise pattern.

2.3.5. Ordered Dither - Contone Screen

This is basically a repetition of a short sequence pseudo-random pattern in the two dimensional grid. Bayer [6] has developed the optimum ordered dither to represent a picture using only two levels. He uses the criterion that the lowest spatial frequency component of the dither should have the smallest intensity. A recursive formula has been derived to specify the screen pattern with nxm in size, where n, m are power of 2. The contone screen uses such repetitive screens to minimize the magnitude of the lowest spatial frequency noise. The overall effect of the status of these pels combines to give the desired apparent grey level. This noise problem mentioned in last method is improved in this method due to the "regular repetitive" structure of the screen pattern. The disadvantage of ordered dither is artifacts in plain areas at some particular grey level. The binary pels will create some pattern that is quite visible.

2.3.6. Error Diffusion

The error diffusion algorithm by Floyd and Streinberg [18] use a simple threshold at the mid-tone but distribute the error generated by the thresholding to its neighboring pels. The error produced when thresholding a pel is accumulated and the distributed to its neighboring pels in a predetermined weight. Over the entire picture, it alternates to create an error distribution that is the least visible. This method produces the best picture so far as grey levels and details are concerned. Its major disadvantage is the complexity of implementation and occasional artifacts in plain highlight and shadow areas.

2.3.7. Electronic Screening

Another way to represent continuous tone picture on the two-level basis is that used in the printing industry i.e. the halftone process, where the grey level is represented by areas of ordered dots. This method is the only one that is subject to the constraint of minimal printable dots. The threshold function is a two dimensional repetitive pattern. The pattern is so designed such that for any given uniform grey level it produces a dot that can be reproduced fairly well in the printing process. Sometimes the pattern is also designed to compensate for some difficulties in printing. This method can produce very good grey levels and can render high contrast fine details. It can not produce detail with low contrast. In that case, almost uniform dots will be created and the texture will not be seen.

CHAPTER 3

Conversion between Contone and Halftone

In the last chapter, various bi-level representations of a contone picture have been discussed. Although electronic screening does not produce the best visual quality as compared to some other methods like the contone screen or error diffusion algorithm, it does produce good pictures suitable for printing purposes. The major concern in this thesis is to find characteristics for pictures that can be transmitted and eventually printed on paper. For this reason, the process of electronic screening is to be discussed in more detail in this chapter. Topics such as the screen structure, the spatial frequency domain representation and conversions from one form to the other are investigated. The information obtained will be used later in finding efficient coding schemes for pictures of this nature.

3.1. Structure of Halftone Screen

The halftone process is nonlinear. Basically it thresholds a contone picture with a two-dimensional periodic screen structure and produces one bit per pel output. As an example, the one dimensional case is shown in Figure 3.1 , where Figure 3.1a shows the original contone picture, Figure 3.1b, the periodic screen and Figure 3.1c the resulting halftone picture. By varying the screen period, different screen sizes can be obtained. Similarly, if the shape of the screen function is changed, the shape and size of dots corresponding to various grey levels will also be different. The net result is a tone scale transformation in the process. That is sometimes desirable to achieve some special filtering. An example can be found in Kato and Goodman [47]. The two dimensional screen is similar to the one dimensional case mentioned above. The screen pattern can now be visualized as a periodic structure of mountains and valleys and the contone image as a surface over the screen structure. In areas where the surface representing the grey level is above the screen function, white pels are produced. On the other hand when the surface is lower than the screen function, black pels are created. Figure 3.2 shows one of the screen patterns.

Theoretically, the spacing between dots should be as small as possible in order to achieve better rendition of detail and create the least visible dot pattern. In practice, there is a limit as to how fine a screen can be used. Due to the limitation on the printing process such as wear on the printing plates, running of ink, and smoothness of paper surfaces, the design of the halftone screen is always compromised in order to print enough dynamic range and minimize the problems intrinsic to the printing process. In order to render the optimal grey levels and reduce the noise

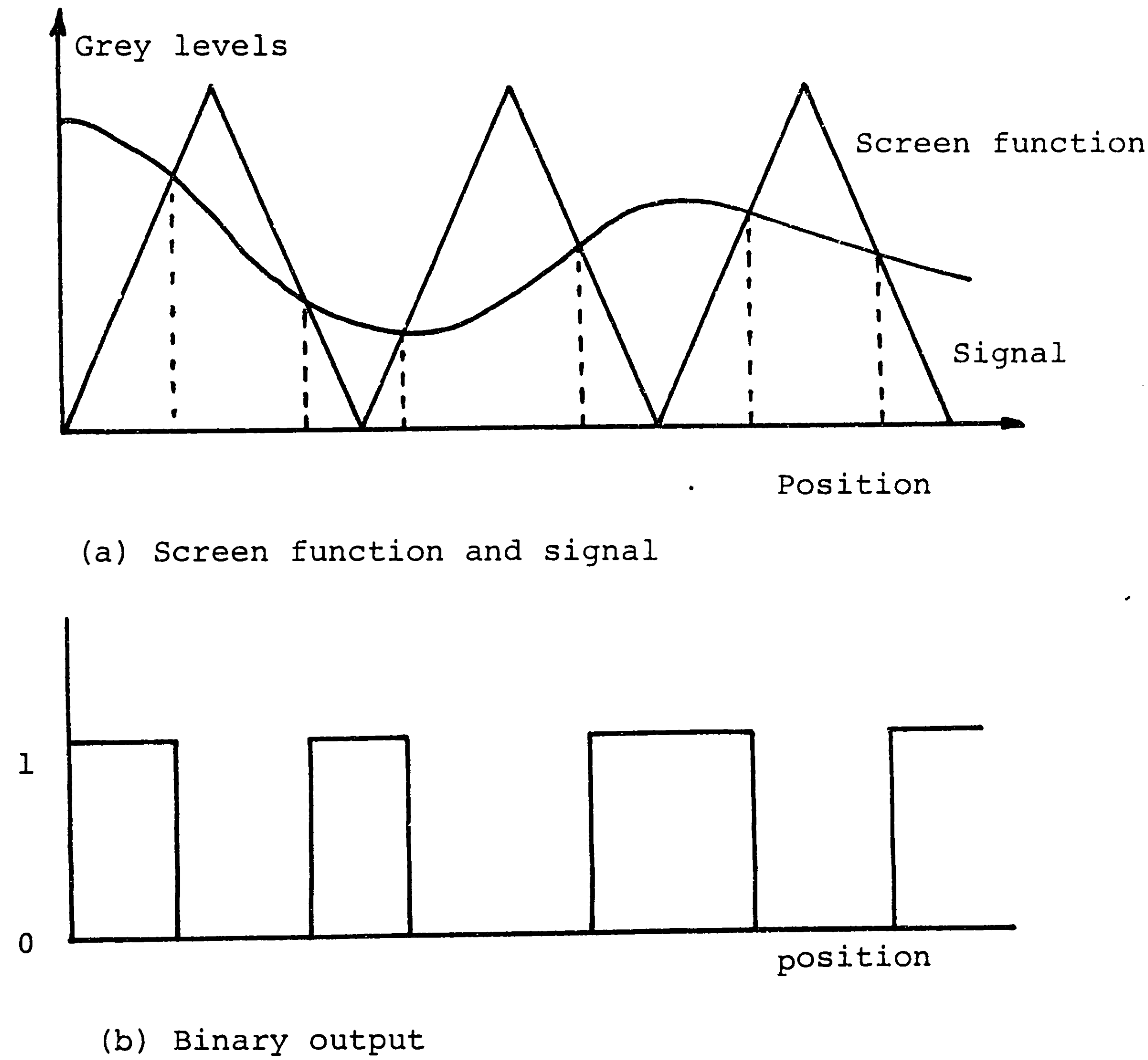


Figure 3.1 One dimensional thresholding process.

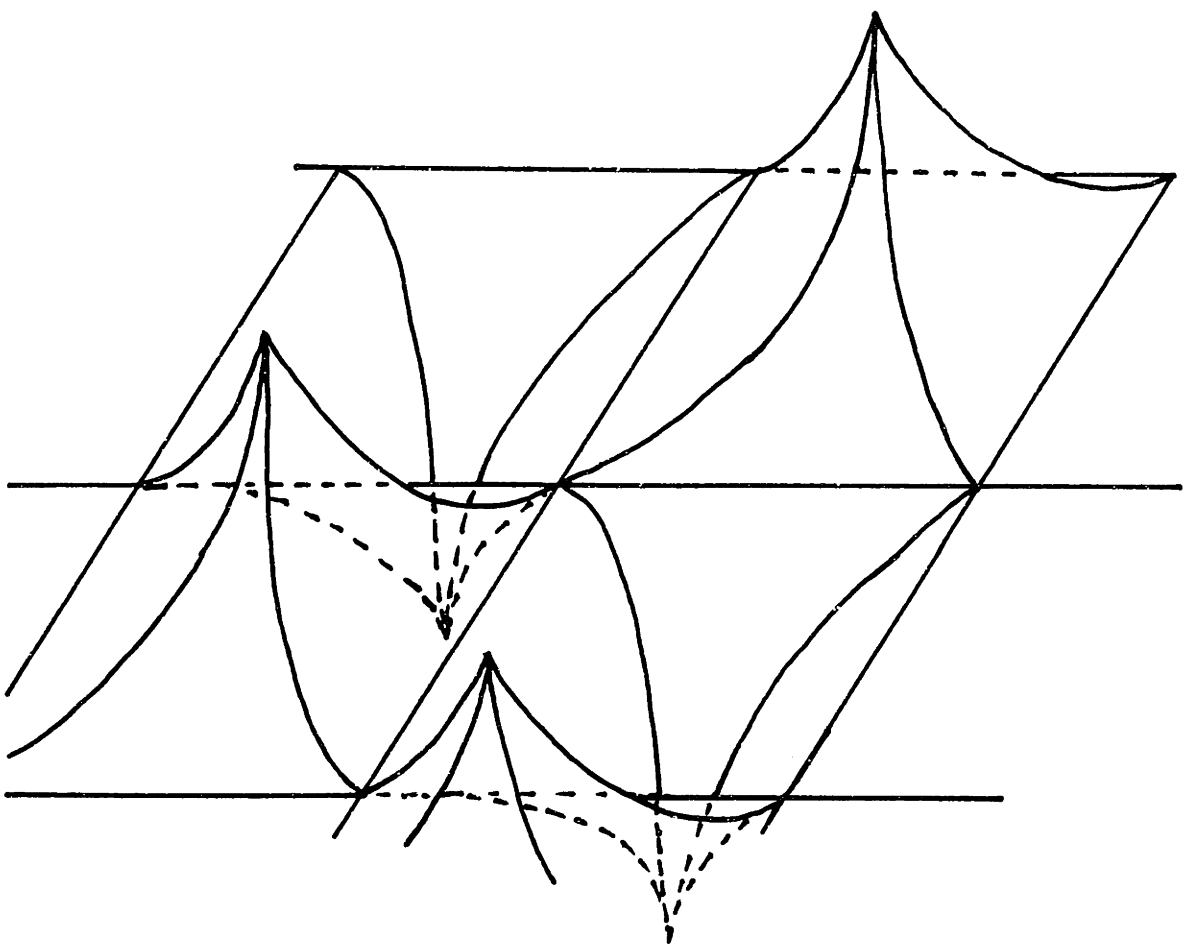


Figure 3.2 Two dimensional screen function.

effect, screens are often so chosen such that small dots are round and hence have smaller perimeter/area ratio in both highlight and shadow areas and gradually change their shape into elliptical or square dots in the midtone.

The following criteria have been followed to create a two dimensional halftone screen to do the thresholding and create halftone pictures.

(1) The screen is repetitive in two directions.

(2) Minimal size of dots created must be large enough for printing purpose.

(3) The size of dots depends on the average local grey shade. When averaged out over a local area, the white areas should be equal to the average brightness around the same area of the contone original.

(4) The shape of the dots for different grey shade should produce minimal distortion when ink is transfered from plate to paper. In some special cases, when a tone scale transformation is to be incorporated into the screen process, the shape corresponding to a grey level should be designed accordingly.

Various shapes of the halftone screen have been simulated on the Autokon electronic process camera [5]. Especially interesting is the hexagonal screen. Theoretically, a hexagonal array renders the best results in the spatial frequency domain. This is due to the reason that among different shapes of the sampling grids with the same interval between sampling pulses, a hexagonal grid can render the highest spatial frequency components [69]. One would not be surprised to find that a hexagonal screen can produce better halftones than the ordinary diagonal screens. Another advantage of the hexagonal screen is the orientation of the screen with respect to the picture to be reproduced. The optimal orientation for rectangular screens is such that the two axes of the dot array make a 45 degree inclination to the vertical. This has been empirically proved to be the least sensitive to the eyes so far as the dot pattern is concerned. Hexagonal screens have three axes of symmetry as compared to two in the traditional rectangular screen. The orientation of the hexagonal screen is thus less critical than the rectangular.

In order to test the benefit of the hexagonal screen some aspects of it have been investigated. Three different sizes of screen patterns have been implemented. They consist of 61, 91, and 127 pels for each dot respectively. The patterns are shown in Figure 3.3 through Figure 3.5. All three patterns are simulated by software. The hexagonal halftone is then sent to the Autokon to make a hard copy. It is found that the hexagonal screen produces rather interesting patterns which

29 35 44 56 61

32 26 20 47 52 55

38 23 9 8 19 43 40

53 41 10 2 7 18 25 34

59 50 11 3 1 6 17 28 31

57 48 12 4 5 16 22 37

45 21 13 14 15 49 46

36 27 24 42 51 58

30 33 39 54 60

Figure 3.3 93 dots per inch hexagonal screen for Autokon

44 50 56 81 96 101

47 41 35 72 87 92 95

53 38 32 23 20 77 80 83

81 59 26 9 8 19 31 71 55

93 78 29 10 2 7 18 28 40 49

99 90 75 11 3 1 6 17 34 43 46

97 88 21 12 4 5 16 25 37 52

85 73 24 13 14 15 22 74 58

57 36 33 27 30 76 89 86

51 42 39 70 79 91 98

45 48 54 82 94 100

Figure. 3.4 76 dots per inch hexagonal screen pattern

62 68 74 100 115 124 127

65 59 56 80 96 112 121 118

71 53 50 44 38 94 109 106 103

83 77 47 22 21 20 37 91 88 85

101 86 39 23 9 8 19 36 43 79 73

116 104 89 24 10 2 7 18 35 49 55 67

105 119 107 25 11 3 1 6 17 34 52 61 64

122 110 92 26 12 4 5 16 33 46 58 70

113 95 40 27 13 14 15 32 42 82 76

98 81 45 28 29 30 31 93 96 99

75 57 51 48 41 90 108 111 114

69 60 54 78 87 105 120 123

63 66 72 84 102 117 126

Figure 3.5 64 dots per inch hexagonal screen appears to have better detail rendering with the same size of dots. Unfortunately, the Autokon can only output picture at a fixed rectangular grid of 722 pels per inch in both directions. With this restriction, output of the hexagonal halftone is not exactly the shape we want. Even with this shortcoming, it can be seen that pictures still have very good visual quality. The output would be better if a true hexagonal device were available i.e., with the pels interlaced between lines and the distance between lines $\frac{\sqrt{3}}{2}$ the distance between any two adjacent pels. The only annoying effect in the hexagonal halftone occurs in the border of the picture, where the border and the axes of dots are not parallel thus create a sawtooth-looking edge.

3.2. Spatial Frequency Domain Representation

To help understand the properties of halftone pictures, it is useful to investigate the halftone picture both in the spatial domain and the spatial frequency domain.

Mathematical models of a contact screens and a ruled screens have been developed by Engeldrum [17] and Streifer et al [91]. Robinson [76] also has formulated representations in the two-dimensional spatial frequency domain for a contact screens. A conventional screen pattern in the rectangular coordinates has been derived by Kermisch and Roetling [50]. Their prediction of aliasing due to the screening is confirmed by experimental results. Their representation also explains why a halftone can produce the appearance of a contone picture. Intuitively and by mathematical analysis the Fourier spectrum of a halftone picture is composed of the desired image signal plus the modulated versions around the two dimensional harmonics of the screen density. The low frequency fundamental gives the appearance of the picture, and the high frequency harmonics are the artifacts. The reason for this structure is the semi-periodic property of the halftone. This result will be extended to a generalized coordinate system i.e., where dots of a halftone picture are repetitive in two arbitrary directions and at distances that may not be equal in the two repetitive directions. This extension is via the generalized reciprocal bases to be discussed in the appendix. It will be specially focused on the hexagonal halftone. Note that the hexagonal raster is nothing but a grid with its two basis vectors equal in magnitude and with an angle of displacement of 60 degrees.

3.3. Conversion from Halftone to Contone

In this section, problems asociated with the halftone picture when it is required to be rescanned are discussed. Some new approaches have been investigated. The dot size averaging algorithm is found to be the most effective and renders much better quality with no guess work as compared to the traditional methods and other approaches in this section.

3.3.1. Problems in Halftone-contone Conversion

Sometimes it is necessary to scan a halftone picture and store the image either in halftone or contone format. In the case of halftone format, the problem is not so serious if no screen conversion is necessary i.e., the space between dots is the same in the input and output pictures. In order to keep enough apparent grey levels in the output binary picture, the input halftone is usually scanned in with relatively small sampling grid. When this sampling grid is 6 to 8 times smaller than the line density of the input halftone then Moire will not be a problem [31]. A more difficult problem is to scan the halftone picture and then convert it to a contone form or convert the original halftone to another screen density for other printing purposes. As we learn from Kermisch and Roetling [50], the halftone picture has very strong harmonics around the screen frequency in its spatial frequency domain representation. Due to the roll-off characteristics of the visual system at high spatial frequency, those strong peaks do not cause serious problems in viewing such halftone pictures. When a halftone picture is scanned in as a contone picture, problems arise however.

The sampling process can be modelled as multiplication by the sampling function which is an impulse train at the sampling grid as mentioned in the last section. When two functions are multiplied in the space domain, that correspondins to a convolution in the spatial frequency domain. The sampling function is periodic and has an impulse at the fundamental as well as harmonics of the sampling frequency. For ordinary contone pictures, this process does not cause a problem because a contone picture does not have significant content in the high spatial frequency region. Optical and chemical processes always attenuate the high spatial frequency component tending to band-limit the signal. On the other hand, in the case of a signal with strong content at high frequencies such as a halftone image, the resulting beat signal does cause a serious problem. When the angle of these two functions are not adjusted properly or the sampling grid is not chosen correctly, the beat frequency can fall in the most sensitive range in the spatial frequency domain. This is the well known Moire pattern. In the halftone-contone conversion, this creates very objectionable low frequency interference.

An example of a Moire pattern is shown in Figure 3.6 which is obtained from scanning a 65 line halftone with a 253 and 95 line per inch device. Clearly the first beat frequency occurs at a horizontal frequency of ($95 - 65^* \sqrt{2} = 4$) and vertical frequency of 0. Figure 3.6 is the image obtained in such process. Note that the vertical Moire pattern is slightly tilted due to the relative position of the original halftone picture and the scan direction of the laser beam. Also the Moire pattern has a slightly higher spatial frequency, due to the fact the the original halftone picture is actually near 64 line per inch and the sampling density is slightly higher than 95 per inch.

To avoid this problem, several methods can be used. First, the sampling grid can be chosen to be small enough compared to the halftone screen density. This is the case when a halftone is to be scanned in and stored as a binary image. The scanning rate should be at least 6 times that of the halftone screen density with 8 times or more preferred. Second, the angle of the scanning direction with respect to the screen direction can be carefully arranged such that the the two-dimensional beat frequency falls outside the sensitive region or has a magnitude small enough not to be noticed. Third, the halftone picture can be preprocessed to get rid of the high frequency harmonics. After preprocessing, only the main lobe of the original spectrum remains. The high frequency side lobes are attenuated to a negligible level making the signal close to band-limited. Traditionally, this procedure is done by defocusing the original halftone until the screen pattern disappears completely.

Figure 3.6 Image with Moire Pattern.

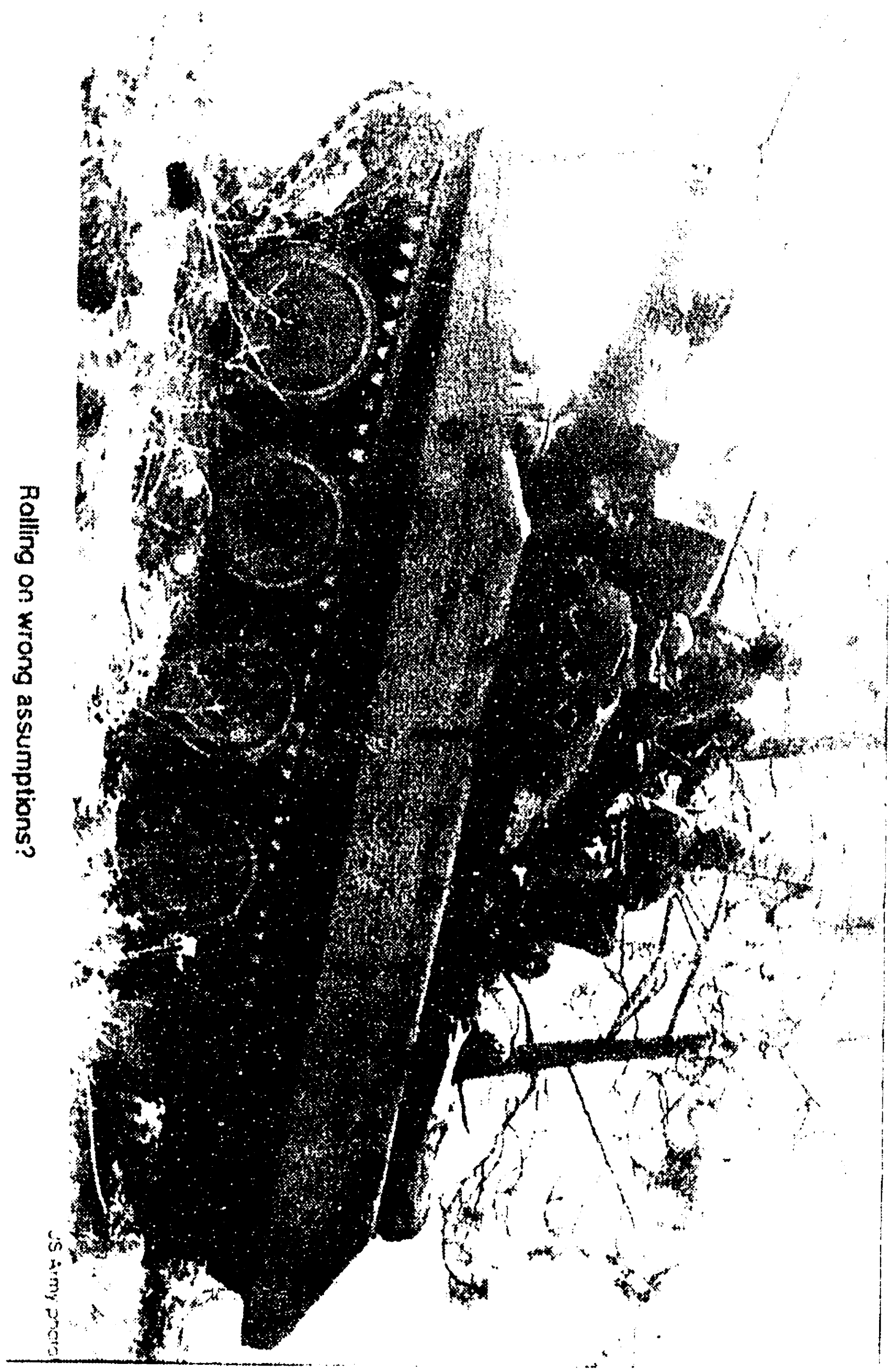


Figure 3.6 Image with Moire Pattern.

All the above approaches have their shortcomings. In the first one, there is not much freedom to select the sampling density. A picture is usually scanned at a particular density to suit some particular output device. In the second, the angle of scan is very critical so far as the Moire pattern is concerned. In practice, just a few degrees of difference can produce pictures of very different characteristics. In the third approach, the quality of the picture is severely degraded. The reason is that halftone pictures can retain a great deal of information in the frequency range higher than the screen density. This is especially true when high contrast details are present in the picture, in which case the dot structure is highly distorted and even broken into smaller dots called partial dots. In the spatial frequency domain, such signals are located above the fundamental screen frequency. When a picture is low-pass filtered to attenuate the screen structure, it loses much of its fine detail. Part of the reason why the defocusing produces very blurry pictures is that a substantial degree of low pass filtering is necessary to suppress the harmonics of dot structure significantly.

In the next few sections, many methods are used and their results compared.

3.3.2. End Point Estimation and Interpolation

In this approach, the boundary of a dot is assumed to be the thresholding point where the grey level of the contone picture is equal to the screen function at that particular location. If the screen pattern is completely known i.e., both magnitude and phase (or the starting and ending location), then the grey value of the boundary can be estimated. After the boundary pels have been given grey values, other pels can then be interpolated by a two dimensional algorithm. Theoretically, if the estimation of the end points is accurate enough, then errors occur due only to the interpolation and those are small. There are two difficulties in this approach however. First is the error in estimation of grey levels of the boundary pels. The accuracy of estimation depends on the number of pels in a dot. For continuous systems, there are an infinite number of pels in a dot, hence the grey value at the end point can be determined exactly. This is unfortunately not true in a discrete system. In most input/output devices, especially those of digital design, discrete systems are used and a finite number of pels are specified for the screen structure in a dot area. For example, in the design of the Autokon, 128 pels are specified for the dot structure of a 65 line screen. For 85 and 100 line screen, they are 72 and 50 respectively. This poses a big problem in estimating the grey values at the end points.

In the continuous system, the location and the grey value at the end point is exact. In the discrete system, the location of the end point is only approximate. The grey values for the corresponding pel only reflects a threshold value. It does not tell the exact grey value of the original contone at that point. For example, for two adjacent pels having different polarity, all we can conclude is that there is a transition somewhere in between the two pels. The grey value is somewhere in between the screen value of the two corresponding positions. The way the screen pattern is designed, two adjacent pels of the screen can differ by a large value resulting in a large estimation error.

A second problem is the difficulty in finding the screen structure from the halftone itself. As mentioned before, this approach requires both the screen structure as well as the relative position of the screen with respect to the picture itself. Any deviation from the true phase value creates errors in the estimation and worst of all, this error propagates when grey levels of the rest of the pels are found by interpolation. In reality, halftone pictures, after many steps of reproduction, almost always have some small skew. All dots are not uniformly spaced, either in distance or in orientation. The precision requirement in order to make a good restoration to the contone is extremely high. It is in principle a noisy approach because errors will be spread over the neighboring pels instead of being averaged out.

3.3.3. Randomized Average

Methods to eliminate Moire patterns in obtaining a Moire-free halftone have been investigated by Allebach and Liu [1] where the local screen is scrambled to produce a pseudo-random pattern. Similar methods can be used to scramble a halftone picture and then use an averaging process to convert it to contone.

The randomized process can be achieved in three ways.

(1) Pels in a small area can be redistributed randomly and then the randomized picture can be averaged over the desired area.

(2) The scanning position can be randomized. Instead of a fixed sampling distance, a randomized sampling distance is used. A small random number is generated each time a contone pel is to be generated. This random number is the offset from the regular sampling grid. After this offset is added to the regular sampling location, the periodicity is broken and Moire patterns generated by the beat frequencies are decreased. Depending on how random this distance is, Moire can be eliminated at the price of more added noise.

(3) The mask inside which the number of white (or black) pels are counted and normalized to a contone grey value can be randomized. For any converted contone pel, it may derived from a larger mask and its neighboring pel may average over a smaller mask. The area of the mask can be randomly generated.

All three approaches have been implemented. It is found that Moire can be avoided if enough randomization is added to the picture during the conversion. At the same time, as the degree of randomization is increased, the noise in the picture becomes worse. Overall, this approach is not as good as the dot size averaging to be discussed in the next subsection.

3.3.4. Dot Size Averaging Approach

The dot size average approach is intended to simulate the human eye's performance with respect to halftone pictures, namely, the averaging effect. The point is to find the best area over which to do the averaging. If a large area is used, more grey levels can be obtained and the risk of contouring decreases. On the other hand, the larger the area chosen to be averaged to get a contone grey value, the more detail will be lost due to this low pass filtering. The problem is to find a mask that can effectively attenuate the undesired harmonics of the screen pattern and also keep the high frequency detail information other than that generated by the screen structure. A good mask is one exactly the size and shape of the screen dot.

We can obtain the information about the halftone screen from the the distance between dots, the orientation of dots (i.e., how the dot pattern repeats itself in areas with no detail content).

If the halftone screen density and the direction of the dots are found then we can create a mask with a size equal to that of a dot and a shape corresponding to the orientation of the halftone screen structure. The mask is then used as a space domain two-dimensional filter. Convolving the low pass filter with the original halftone picture, we obtain a contone picture with only slight degradation. Moire patterns are suppressed so as to be negligible if the mask is properly created. The dot size averaging algorithm needs just the size and shape of the dot structure to do the conversion. It is, in principle, a noise-free operation. An error occurring in one pel is likely to be averaged out during the process.

In order to create a mask for dot size averaging, the first task is to examine the binary picture and from its structure to get the best estimate for the mask in both size and shape. One of the problems is to find a small area where dots are nice and clean. An averaging mask derived from such an areas has a small error so far as the elimination of Moire patterns is concerned.

In a halftone, a clear, regular dot structure is not to be found everywhere in the picture. In a plain area, where the picture consists mainly of low spatial frequency components, there is a slow variation in grey from pel to pel. Dots are arranged regularly and they are spaced evenly. In this case, the vector found by drawing a line between two adjacent dot centers is very close to the true direction and spacing of the screen structure. On the other hand, in an areas with high contrast detail, the dot structure is not so distinct. The shape of dots is distorted to reflect the detail in the picture. In extreme cases, the dots are even broken into partial dots. Clearly, areas of high contrast detail are not ideal for locating dots and finding the screen structure. The spacing between the center of the dots is thus a good criterion to determine whether a small area is proper to find the dot structure. Once this criterion is selected, the next problem is to find the center of a dot.

Areas in which low spatial frequencies prevail are a better choice. An interactive system could be designed to locate such areas. Such a system would be too time-consuming and add an extra burden to the operator. A better approach is to find an algorithm to pick up an area with the desired properties and let the computer do all the work. In this thesis a random search algorithm is used. Small areas are randomly selected to find dots. Criteria have been established to determine whether the mask so found is good. A few such masks are found and compared. The process keeps going until masks which pass all tests can be found.

The first step in finding a mask is to find a dot. In order to find a dot, the polarity of a dot in the area must be determined. In the highlight, we have black dots in a white background. Similarly, in the shadow, we have white dots in a black background. The runlength of both black and white runs can be used for this purpose. We can locate a small area in the halftone. Inside that specified small area, lengths of both black and white runs are tabulated. If the area has black dots, the minimal length of black runs can be as small as one pel. The maximum of the black runs is likely to be the width of a black dot. On the other hand, the minimal length of white runs is the distance between two black dots and the maximal length can be as long as the length of the entire picture line. This maximal runlength of black runs is thus shorter than the maximal runlength of the white runs. This relation is better illustrated in Figure 3.7. The polarity of runs that produce, the smaller maximal length is taken to be

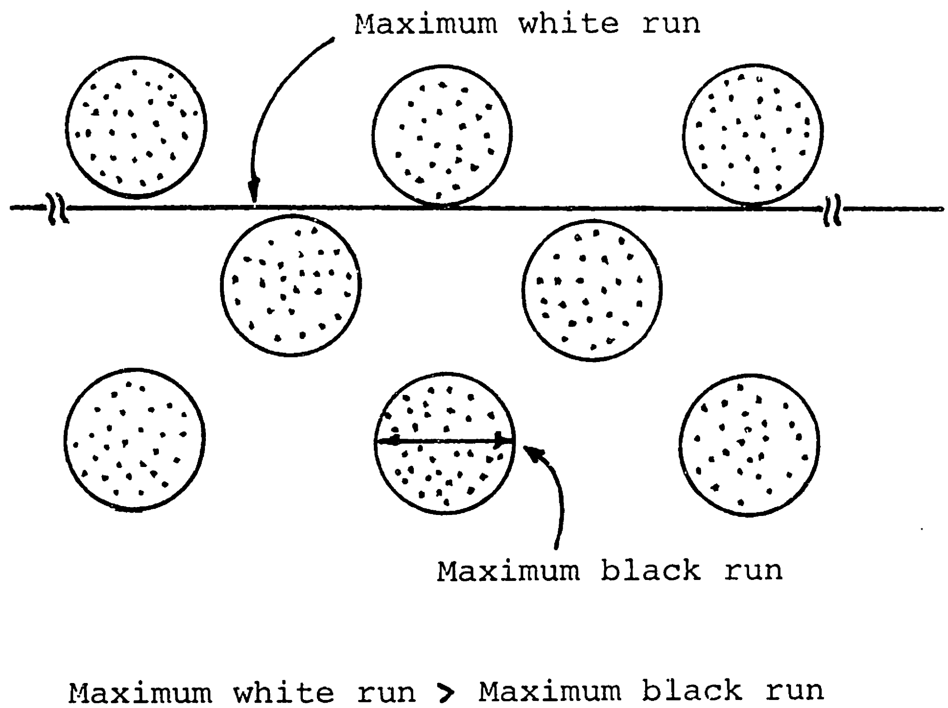


Figure 3.7 Determine the polarity of dot by length of runs.

the polarity of the dot in that small area. Once the polarity is determined, the search can be started from where the maximum length has been found because it must be close to the center of a dot. The two ends of a run are used to locate the left and right coordinate of a dot. Lines above and below the starting location are checked to see if these lines contain pels for the current dot. If the answer is yes, then coordinates of the left and right end points are compared with that of the current leftmost and rightmost pel of the dot and updated if necessary. If the line is found not to contain any pel from the current dot then the the line number is taken to be either the upper or lower boundary of the current dot. The process continues until the top and bottom of the dot are reached. Once the four extrema of the dot are found, it is easy to estimate the center of the dot by simply averaging those coordinates.

After a dot is located, the next thing is to find the adjacent two dots. This can be accomplished by searching for dots at the lower left and lower right of the dot that has just been found. The same dot-finding algorithm can be used. In principle, the coordinate of the fourth dot can be determined from those of the other three. Actually, the coordinates of the fourth dot is found in the same way as the other three and then compared with the calculated value to make sure the mask is correct.

When four adjacent dots are found, the mask can be obtained by drawing lines between the centers of these four adjacent dots. For typical halftones, the mask is a square. Its orientation is generally about 45 degree with respect to the vertical and horizontal axes. When more exotic screens are used, a different shapes of mask will be created. For example a hexagonal screen creates a rhombus at 60 degree and an oblique screen creates a parallelogram.

After the mask has been determined, it is then put on top of the halftone picture and the averaging process started. The number of white (or black) pels inside the mask area is counted. The sum is then normalized by the area of the mask (total number of pels in the mask) and converted to a contone grey value with the desired number of bits.

Although the dot size averaging algorithm does attenuate the high spatial frequency components, it differs from low pass filtering. In the case of low pass filtering such as the defocusing process, all high frequency signals above the fundamental screen frequency are attenuated severely in order to prevent the screen pattern from generating beats with the sampling function and creating Moire patterns. In the dot size averaging process, only those signals in the vicinity of halftone screen harmonics are attenuated. The shape of the filter is a two-dimensional notch filter. It rolls off sharply at the undesired screen frequency and its harmonics.

For signals that are above the fundamental frequency of screen pattern but not in the immediate vicinity of harmonics of the screen structure, the attenuation is smaller than in the case of a low pass filter. Such signals usually come from the high contrast detail and their rendition usually significantly improves the picture quality in detailed areas.

The effect of the dot averaging process is explained as follows:

a. The mask derived from the screen structure is in general a parallelogram and hence symmetric with respect to its geometric center. If this parallelogram is flipped twice, once around the horizontal axis once around the vertical axis, then the shape and orientation of the new parallelogram is identical to itself.

b. The dot size averaging process can be written as:

$$G(m,n) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} M(i,j)H(i+m,j+n)$$

where M is the mask function. It has the value 1 for pels inside the mask and 0 otherwise. H is the halftone signal and G is the total number of white pels to be counted in the mask which will be normalized to a grey value. The effect of counting the number of white (or black) pels within the mask area is equivalent to multiplying the picture with a mask that has a value of 1 inside the mask and 0 outside.

c. Using the symmetric property of the mask, the output now can be rewritten as

$$G(m,n) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} M(-i,-j)H(i+m,j+n)$$

by substitution of i to m-i and j to m-j, we have

$$G(m,n)= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} M(m-i,n-j)H(i,j)$$

That is the familiar convolution form. In the spatial frequency domain, it is easier to deal with the one dimensional case first. Referring to Figure 3.1, it is clear that when the period of the 1-D halftone screen is T, the mask is 1 from -T/2 to T/2. From the basic Fourier transform, a square wave transforms into the sinc function $\left(\frac{\sin 2\pi\frac{T}{2}f}{2\pi f}\right)$, where f is the spatial frequency in cycles per unit length. Zero-crossing occurs at frequencies which are multiples of the inverse of the width of the square wave. These zero-crossing points are located exactly at the harmonics of the screen structure. Therefore after this averaging, we can expect the screen structure to be completely removed because of those zeroes in the frequency domain. The two-dimensional case is similarly an extension in dimensionality. If the mask is a square, then in the spatial frequency domain, it is simply the product of two sinc function in the x and y directions. It is illustrated in Figure 3.8, where we find that the zero crossing at multiples of dot frequency in the two orthogonal

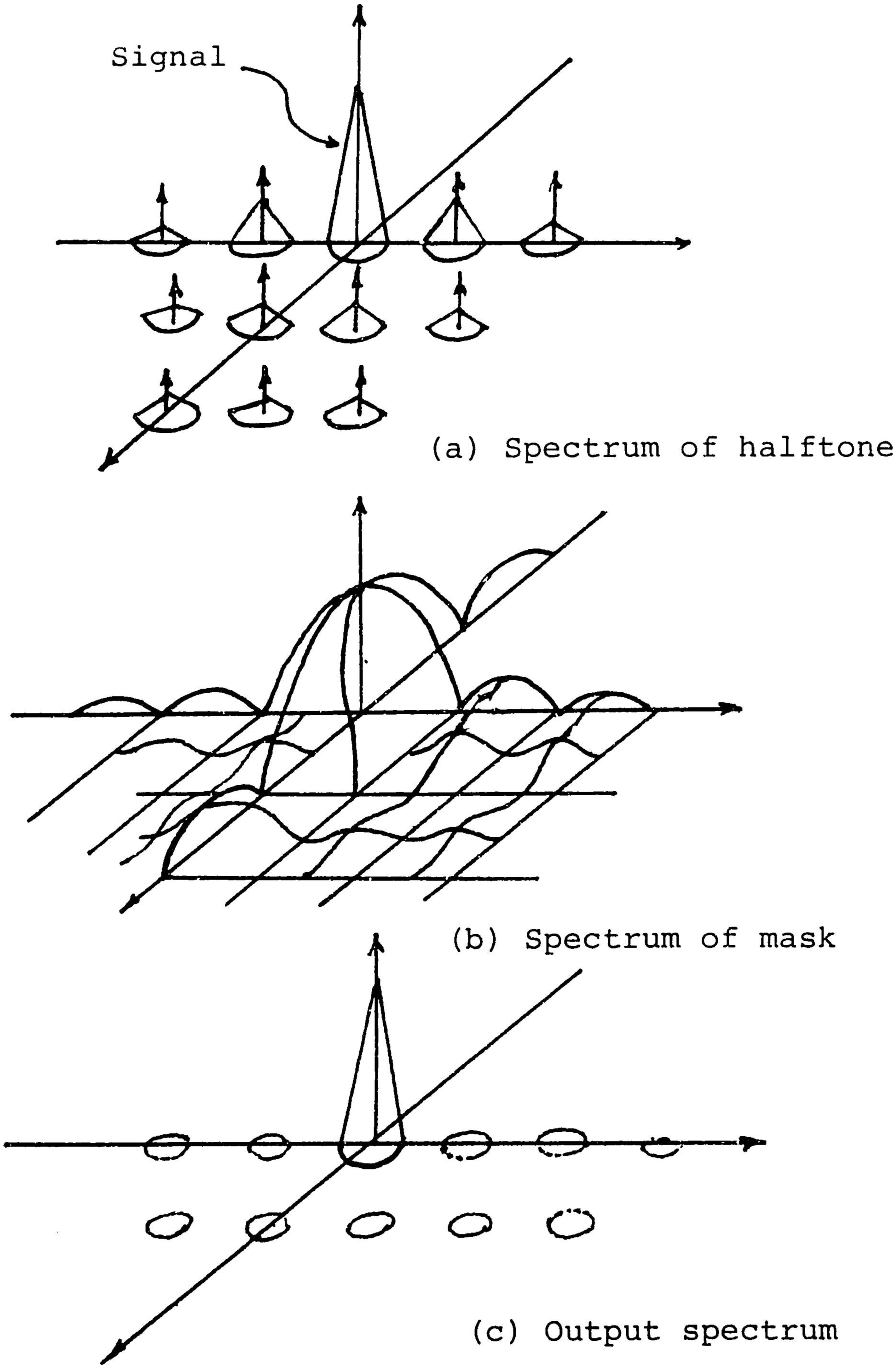
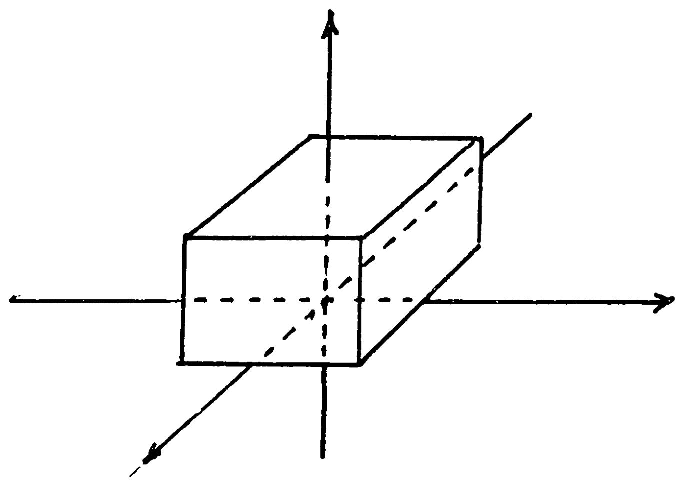


Figure 3.8 Convolution of two functions.

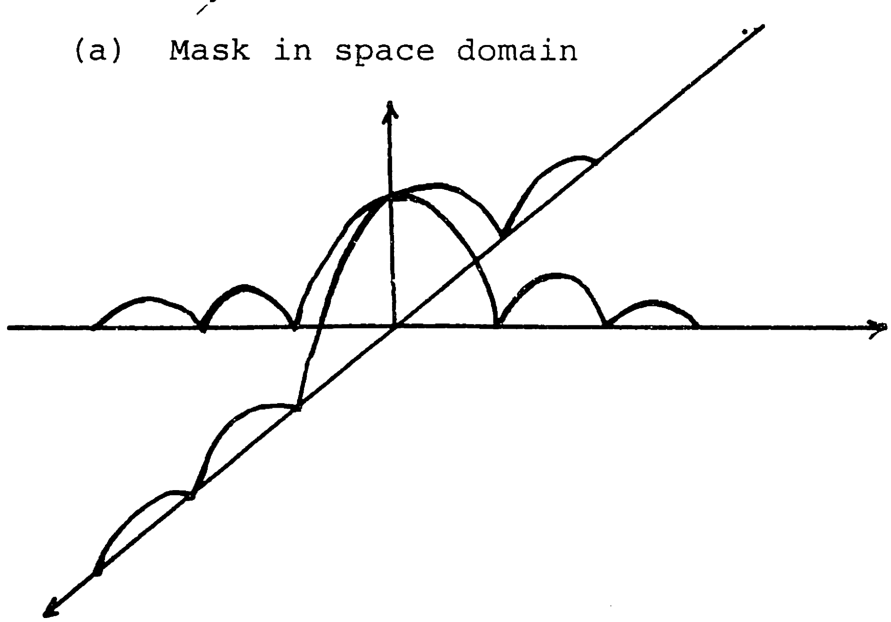
directions. If the screen is not of the traditional rectangular shape, we can still derive a similar result by using generalized coordinates discussed previously. This is illutrated in Figure 3.9, where Figure 3.9a is the mask and Figure 3.9b is its Fourier spectrum. Notice that the mask is created from vectors that characterize the screen pattern. There will be zeros located exactly on the top of the impulses generated by the repetitive screen pattern. It is this comb filtering characteristic that attenuates the strong periodic signal in halftones generated by the repetitive screen structure, thus maintaining some high frequency energy and the corresponding high contrast details.

The dot size averaging algorithm has been implemented and works very well, especially for halftones with high line density such as above 100 line. An example can be seen in Figure 3.10. There is almost no lose of detail. The only defect comes from the "wash out" in highlight or dark area when small dots can not be precisely reproduced in the scanning process. Although the high frequency components are somewhat attenuated, the detail rendition is still rather good.

On the other hand if the original halftone uses a coarse screen (eg. a 65 line screen or coarser), then due to the large mask and hence a much lower cut off frequency, more



(a) Mask in space domain



(b) Mask in spatial frequency domain

Figure 3.9 Mask for generalized dot structure.

Figure 3.10 Original 65 line halftone

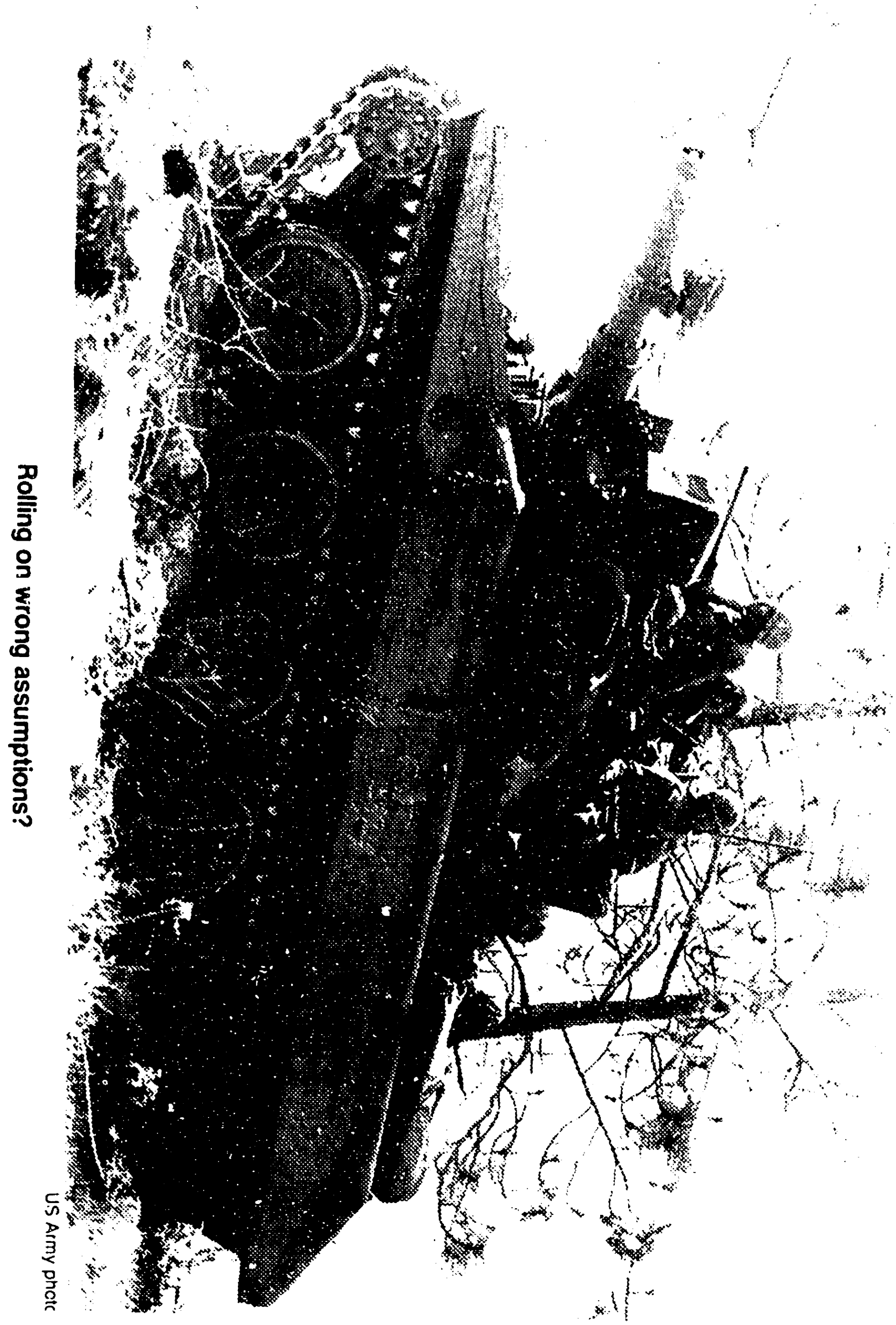


Figure 3.10 Original 65 line halftone

Figure 3.11 Converted contone

Figure 3.11 Converted contone

severe attenuation of the important high spatial frequency signals will occur. In the case of pictures with a large amount of high contrast fine detail such as test patterns, the attenuation of fine detail becomes visible. In these cases, some method to extract the information on the detailed area is necessary in order to render the desired visual quality.

One logical approach to this problem is by adaptive averaging. From experience, we know that Moire patterns are not a problem in areas where the contents of the picture are highly detailed. They are troublesome only in plain areas. Picture quality can be improved if the averaging mask can be adapted to the local contents of the halftone, i.e., depending on the local pictorial content. A different algorithm can be chosen either to enhance the detail or to avoid Moire patterns. In areas where low frequency components prevail, dot size averaging effectively eliminates Moire patterns. On the other hand, where high contrast fine detail prevails, then Moire will not be the major problem. In this case, instead of using dot size averaging, small area averaging can be used to simulate the ordinary scanning process.

The scheme works as follows: in the area where only grey levels prevail dot size average is used. On the other hand for areas with fine detail, Moire is not a problem. In the frequency domain, there is little or no peaking at the harmonics of screen frequency. In such areas, a mask with much smaller size can be used to simulate the ordinary scanning process. The smaller mask is in effect a filter with much higher pass band width. There is less attenuation of the high spatial frequency content. Of course, due to the smaller mask size, the number of reproducible grey levels is also decreased, but that is not a problem either because the contouring effect is unlikely to occur in such areas.

Fortunately, the structure of the halftone itself can give us the information needed to determine the detail content. Methods can be devised to determine the regularity of the dot.

The method used in this thesis employs the symmetry of regular dots. As mentioned earlier and in the previous chapter, the shape of dots tends to be regular in plain areas. These regular dot shapes tend to be round for both dark and highlight areas. They tend to be square in the mid-tones. In both cases, the dot created will be more or less 2-d symmetric with respect to more than 2 axes passing through the center of a dot. On the other hand, in areas where high spatial frequency components prevail the dots are usually distorted. The high contrast fine details usually distort the dots to render the texture. The dots are distorted at least in some direction. In most cases the degree of symmetry is decreased.

For example, if a dot has been elongated from a circle to an ellipse, it has two axis of symmetry - the major and minor axes instead of any axis through the center. If the dot is even further elongated or broken into partial dots, chances are that it loses its symmetry entirely.

The second problem is the determination of whether the dot is distorted. The algorithm employs the projections of perimeters in the two orthogonal axes. If we shift the dot in any direction and then compare it with the original dot, the difference obtained by an "exclusive or" operation will indicate the projection of the perimeter in the direction perpendicular to the direction of dot movements. For example, we move the dot to the left or right by a unit and then use the "exclusive or" operation with the original dot to find the difference of these two pictures. This difference generated gives the projection of the perimeter of a dot in the vertical position. When a dot is elongated vertically, we can expect that such operation will generate a large projection. Similarly, the projection of the perimeter in the horizontal direction can be obtained by shifting the picture up (or down). This is illustrated in Figure 3.12. For dots with regular shape, the two numbers obtained will be very close to each other. On the other hand, if the dot has been elongated horizontally, its vertical perimeter projection would be much smaller than the horizontal perimeter projection. The same

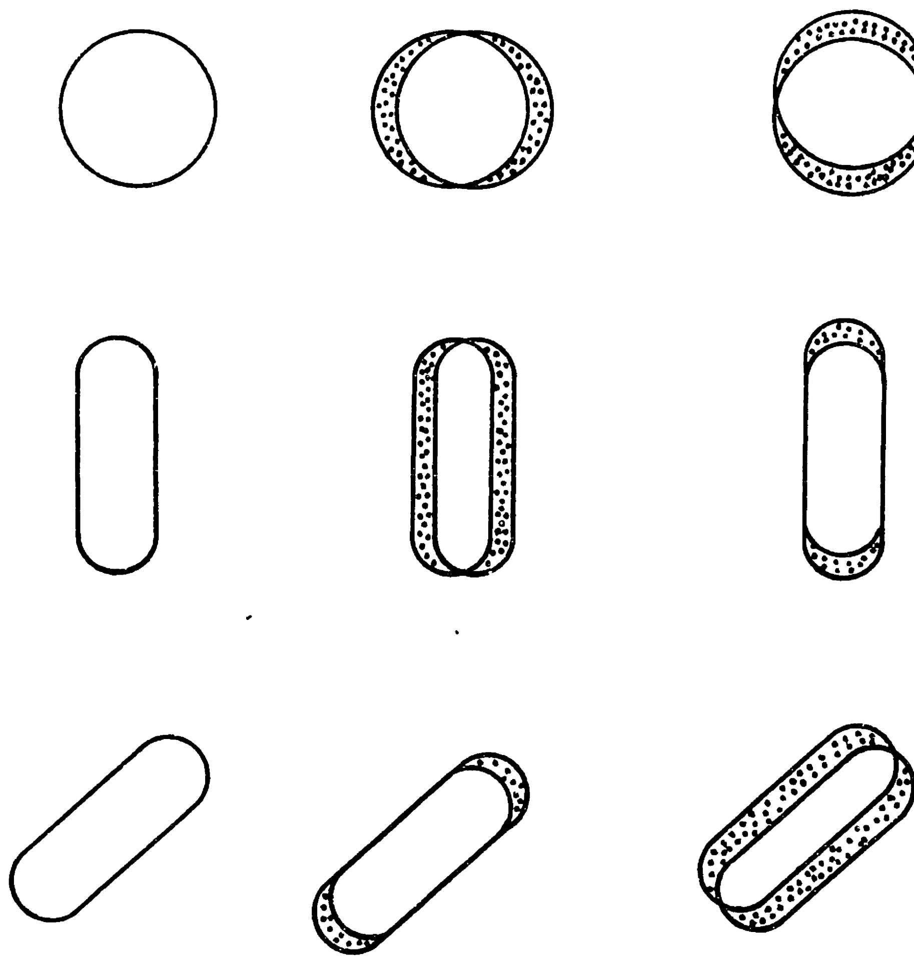


Figure 3.12 Detection of dot symmetry.

situation is true for dots that have been elongated vertically when the vertical perimeter projection will be larger than the horizontal. The ratio of these two numbers forms a good indication of the shape of dots. This method can be extended to detect the elongation of dots in any direction. If the dot is distorted in a direction other than either horizontal or vertical axis, we can apply the same operation to that particular direction it has been twisted. Theoretically, we need such parameters in every direction in order to determine how much and in which direction the dot has been elongated. In practice, it has been found that two such operations are enough in deciding the shape of dot and hence the texture of the picture in that particular area. One of the operations detects the elongation in either the horizontal or vertical direction. The other determines the extent of elongation in the axis that are inclined at 45 degree with respect to horizontal axis.

Ratios for these parameters of elongation in four directions are then added up as the indicator of detail content in a halftone picture. In areas where this indicator is small, it means the halftone does not have high spatial frequency content. Dot size averaging is then used for contone conversion to eliminate the Moire patterns. When the indicator is above some threshold, it is assumed that the area in rich in fine detail, and hence detail rendition is more important than the Moire suppression. A small mask is used then to preserve the detail.

The result of this detail enhancement method is shown in Figure 3.13.

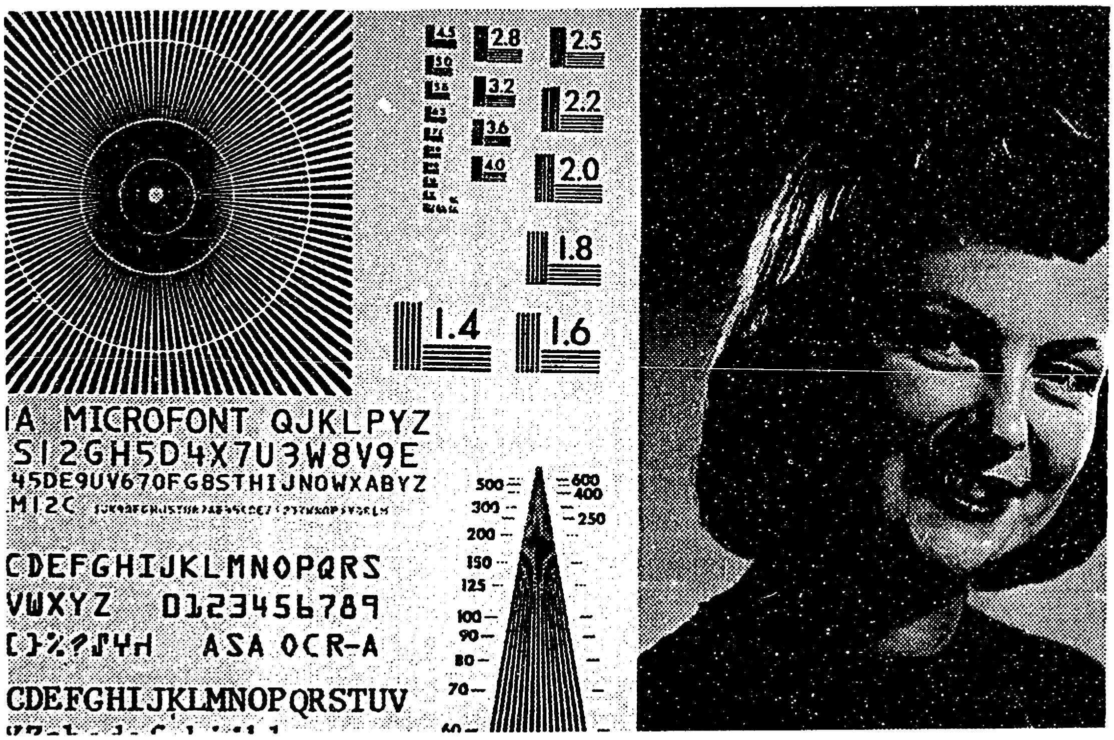


Figure 3.13 Original halftone

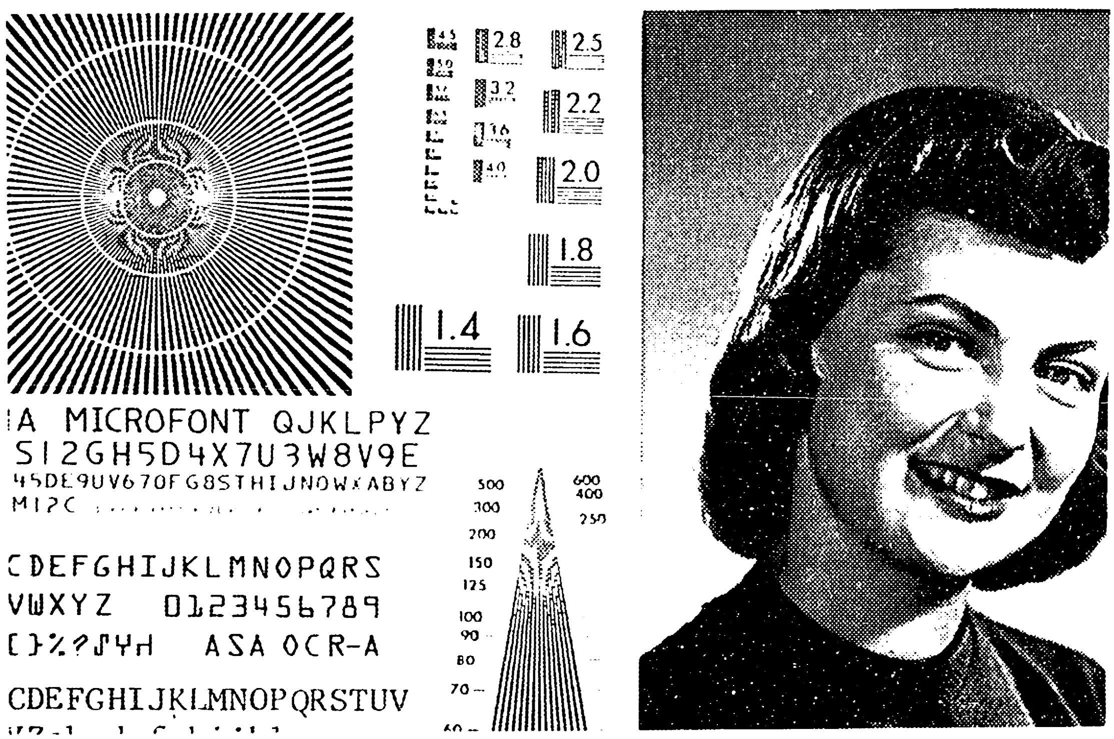


Figure 3.13 Original halftone

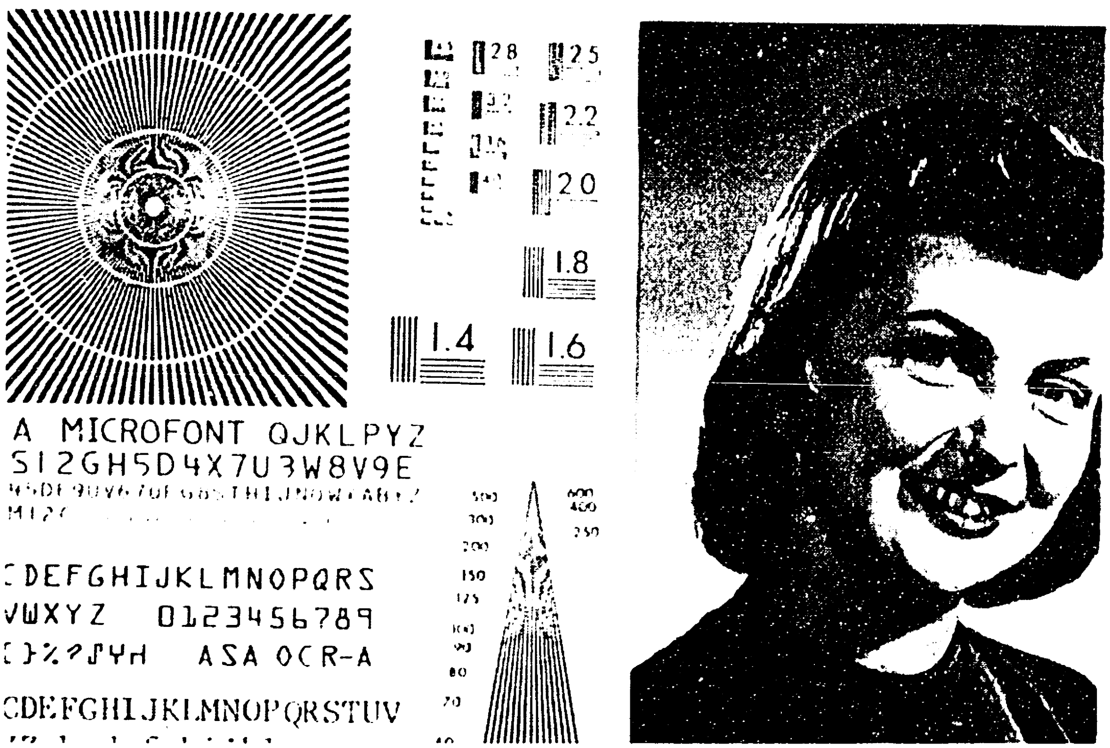


Figure 3.14 Conversion using dot size averaging only.

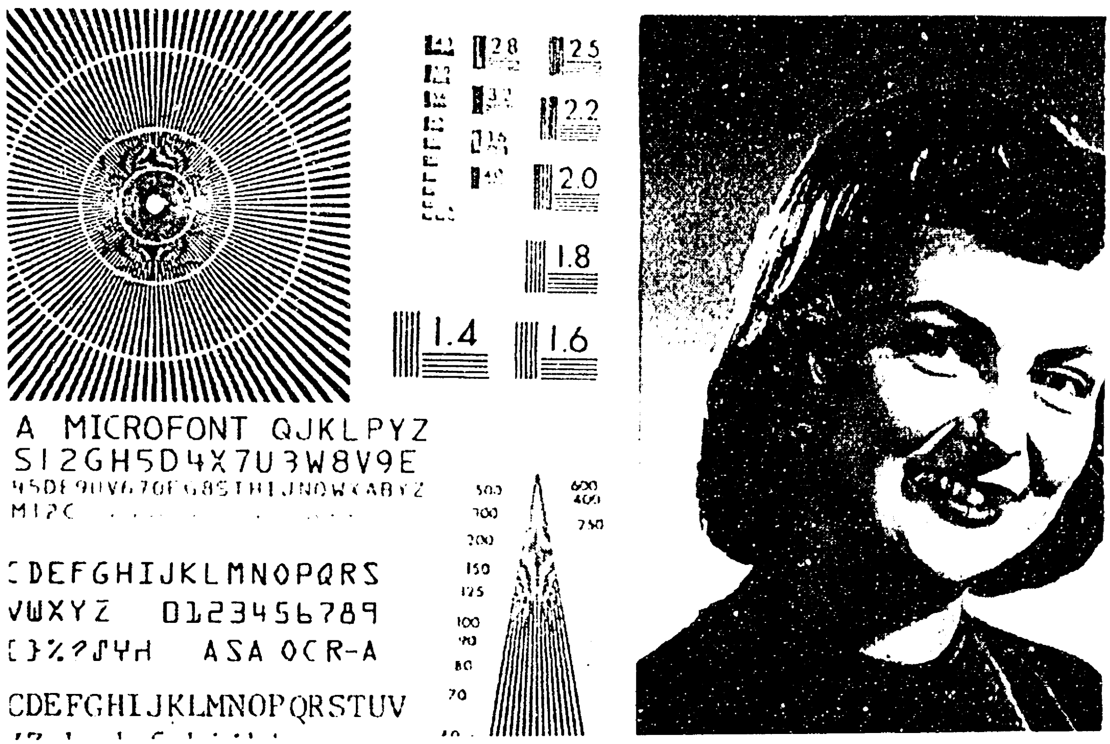


Figure 3.15 Conversion using adaptive detail extractor.

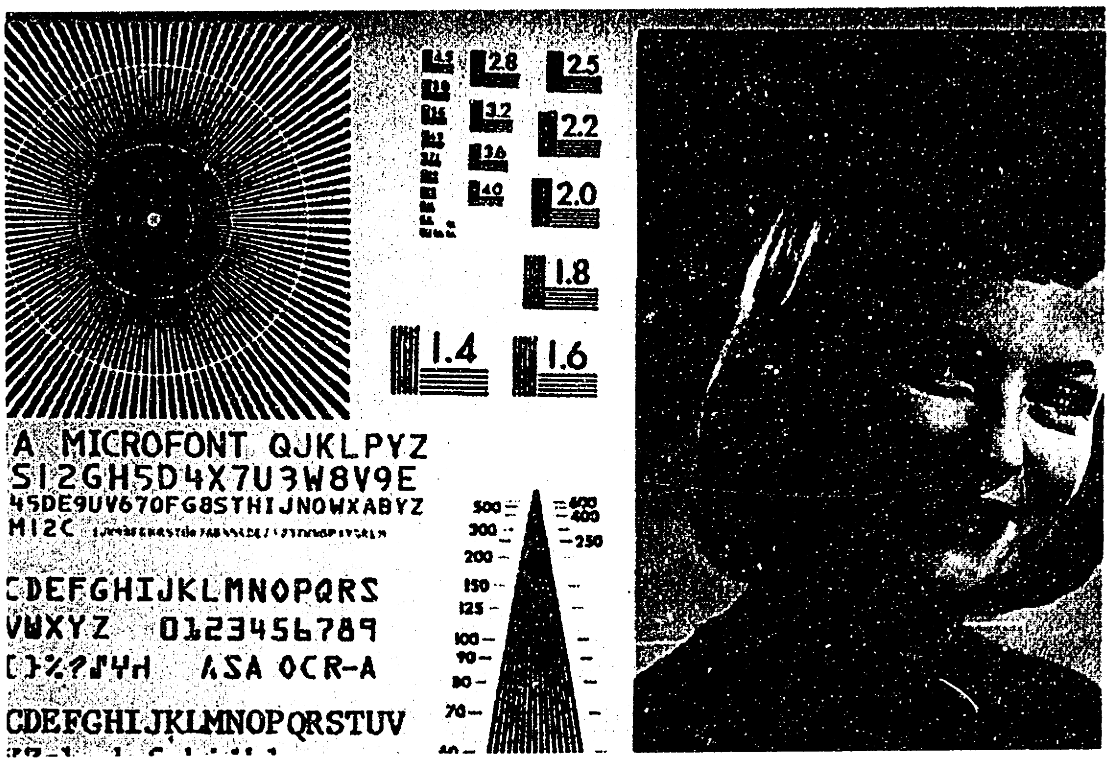


Figure 3.15 Conversion using adaptive detail extractor.

CHAPTER 4

Problems in the Coding of Halftone Pictures

In this chapter, commonly used techniques for picture coding are reviewed. Special problems which arise when these methods are applied to halftone encoding, and some relevant aspect of coding halftone images are discussed.

4.1. Source of Redundancy

In a discrete system, in order to represent a picture faithfully, there are two rules that must be closely followed. First, the sampling grid must be sufficiently small. This will ensure the detail to be rendered clearly. From the sampling theorem, if a bandlimited signal is to be represented perfectly in a discrete system, the sampling frequency must be at least twice the one sided bandwidth of the signal. Second, the number of representable grey levels must be large enough so that the difference from two levels can not be perceived. Usually, it takes from 6 to 8 bits (64 to 256 grey levels), to represent the grey levels in a pel, in order to eliminate the undesirable contouring effect.

In the uncoded form, called PCM, there is redundancy in the representation. The redundancy comes from two sources. First, there is the psychophysical redundancy of the human visual system. The two constraints mentioned above almost always do not occur simultaneously. In other words, where the picture is rich in fine detail, there is no need to use so many bits to represent the grey levels. That is because in detailed areas, the human visual system cannot discern the differences in grey levels as well as in plain areas. On the other hand, in areas where the visual system is most sensitive to the changes are those that contain mostly low spatial frequency components and, consequently, there is no need to have a very fine sampling grid. This effect has been discussed in chapter 2.

The second source of redundancy comes from the high correlation of grey values between neighboring pels. In most cases, the grey level of any pel can be accurately estimated based on the knowledge of its neighboring pels. This is especially true in plain areas where the grey levels of adjacent pels are highly correlated. As the texture of the picture becomes more complex, the correlation is reduced and it becomes difficult to predict the levels of adjacent pels. Thus the magnitude of the prediction error becomes large. In other words, there is less redundancy in fine detail areas and it takes more data to represent a picture faithfully thus only small data compression ratios can be achieved.

In the case of halftone pictures, we have a somewhat similar situation with two major differences. First they have a periodic dot structure and second their individual pels can have only two values. From the periodicity of the screen pattern, we can expect that the autocorrelation of the halftone picture will have a similar periodic characteristic. In reality, the autocorrelation of a halftone picture is strongly influenced by both the screen structure and the pictorial material itself. Autocorrelations of a typical halftone pictures is shown in Figure 4.1. It can be clearly seen from this picture that the autocorrelations have the periodic structure of the halftone dots with the same period and orientation as the original digital halftone. The autocorrelation peaks at the origin and tapers off in the radial directions but it has repetitive local maxima and minima. This peaking of the autocorrelation indicates that the adjacent pels are not the only source of redundancy. Because of this correlation with pels a dot period away, coding schemes that reduce only the redundancy of adjacent pels can not provide high coding efficiency.

4.2. Existing Coding Techniques

A number of different data compression techniques have been developed for encoding images. The basic goal of all of them is to reduce the amount of data that is required to reproduce image faithfully.

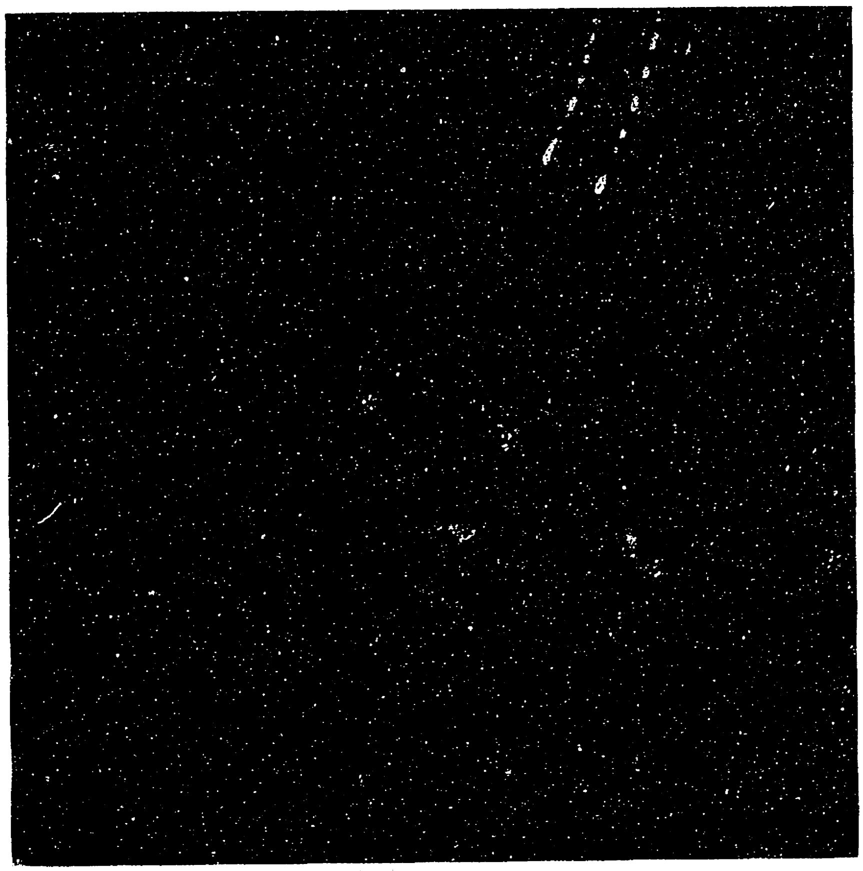


Figure 4.1 Autocorrelation of a halftone picture

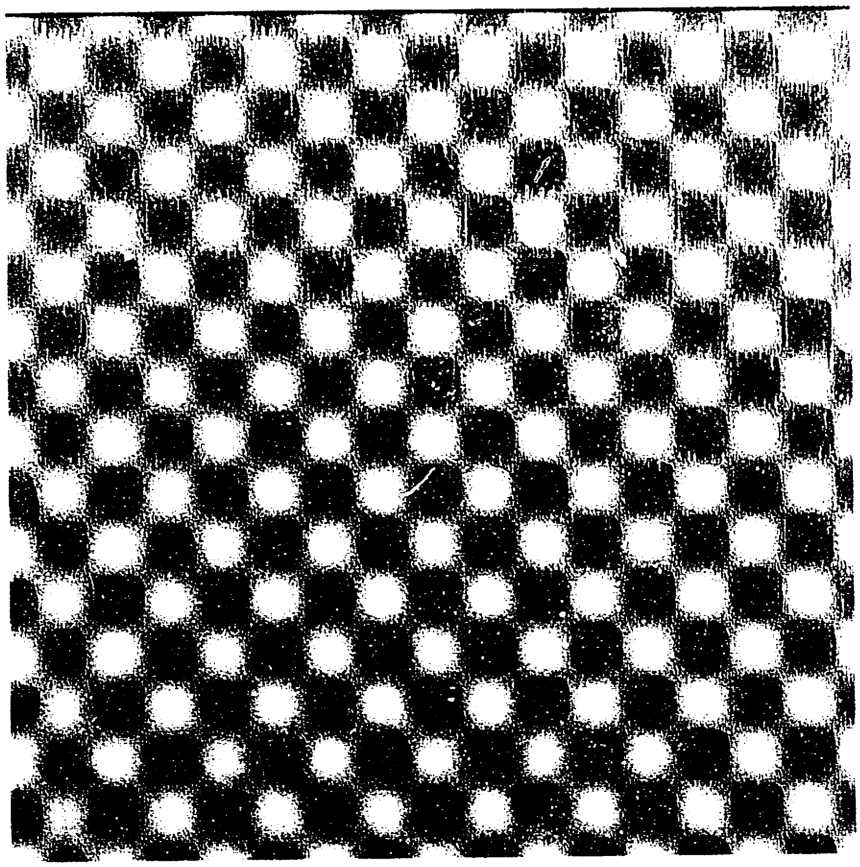


Figure 4.1 Autocorrelation of a halftone picture

Working in the contone domain, we have DPCM [67], delta modulation [42][83], contour coding [25], transform coding [27][97], bit-plane encoding [16][89] and the two channel coding scheme developed in MIT CIPG [85]. In the bi-level picture category, there is runlength coding, bit interleaving encoding [45], and pattern matching coding [46].

Among the compression techniques, transform coding can achieve the highest compression ratio with the least measured mean square distortion. Its disadvantage is mainly the large number of computations required for both encoding and decoding. A consequence of this is that computation time is long or, if speed is important, the system is substantially more complex than other coding methods. Transform coding achieves its high efficiency by assigning bits appropriately according to the frequency of the signal component. More bits are assigned for each signal at low spatial frequency, fewer bits for the high frequency components, and signals above a certain frequency are thrown away completely. This distribution of data bits reflects not only the intrinsic roll-off characteristics of an image, but also the decreasing of sensitivity of human visual systems at high spatial frequencies. In other words, both sources of redundancy are decreased. Due to this property, transform coding is the most efficient coding scheme. With the same number of bits, it can achieve the smallest coding error. Although this is not necessarily translated into good picture quality.

DPCM is a straightforward and effective method for the coding of contone pictures. Its simplicity of implementation is its main merit although the compression ratio is not as good as some transform coding techniques. It has been adopted for many real time applications. Effectiveness of this technique depends on the close correlation of grey levels among adjacent picture elements. Usually only small differences in grey values are expected for neighboring pels. This small difference permits the use of fewer bits to transmit the information needed to reconstruct a picture. Performance of DPCM can be improved by increasing its order, i.e., to predict the grey level of a pel by more than one neighboring pels. N-th order DPCM is found almost as effective as transform coding when N is greater than 3 [27] and the predictor is tuned to the characteristic of the picture to be encoded. When picture characteristic changes, performance can be degraded. Transform coding in general is less sensitive to the characteristics of pictures.

The two channel coding scheme decomposes an image into a lows channel and a highs channel. In each channel, The number of levels and the sampling density are assigned to optimize the coding efficiency and the visual quality.

So far there are many coding schemes developed for binary pictures with different characteristics, most of them in facsimile for text and graphics. Among them, run length encoding is the most popular due to its simplicity and efficiency. There are various techniques to code bi-level pictures, especially for facsimile. Among them run length coding seems to be the most popular [9] [30][33][35][36][60]. Statistics of facsimile have been investigated by Frank et al [20], and Kunt [53][54]. In their work, they have been mainly concerned with images containing text and line art, and not the conventional halftones. Some statistical models have been derived for text and simple line art. If further encoding schemes such as bit interleaving [45] or 2-D pattern matching [46] are used, then another 3 to 1 compression ratio is achievable. In the case of text or line art, the distributions of black and white pels are usually quite different. This makes run length coding rather efficient.

4.3. Problems in Coding Halftone Pictures

There are some differences between halftone and contone pictures that limit the application of most of the current contone coding techniques in direct halftone encoding.

First, halftone pictures cannot be approximated as a bandlimited signal. Due to the dot structure, the halftone image is rich in high frequency components as analyzed by Kermisch and Roetling [50]. These high frequency components severely degrade the coding efficiency if transform coding is used. As mentioned before, the effectiveness of transform coding lies in the fact that in contone images, signal energy is concentrated at low frequency, rolling off quickly as the frequency increases. With this property, a smaller number of bits can be assigned to represent the upper range of the signal spectrum. A halftone picture does not have this distribution in its frequency representation. On the contrary, the repetitive dot structure modulates the signal and creates strong components around the screen frequency and its harmonics. Lacking the roll-off characteristic, transform coding can no longer be efficiently applied because a large proportion of data would have to be allocated to represent high frequency components.

Second, the adjacent pels have either the same or different grey level. It takes one bit to indicate the difference of two adjacent pels. There is really not any gain unless this is only an intermediate step for arranging pels such that another coding techniques can be applied to the image as discussed in chapter 6. This renders direct DPCM useless in binary picture coding.

Among bi-level picture coding schemes, run length coding is efficient only when the length of each run is long enough. By the same token, contour encoding is useful only when large and simple contours dominate in the picture. Chain coding and polygon encoding are very useful techniques in the reduction of data bits as has been done by Ellis in his bit plane encoding scheme [16]. Unfortunately, the effectiveness of this technique depends on bit plane. When a picture is composed of a large number of small areas with short perimeters, as in the case of a halftone, then this technique can not be applied effectively.

To transmit a halftone pictures, we have two choices: either we can screen the contone picture and transmit the resulting halftone or we can transmit the original picture and screen it to a halftone after it has been received. Each method has its own advantages and disadvantages. The first approach is to investigate the kind of pictorial information that is important in reproducing a halftone from a contone. This vital information is then coded and transmitted. The screening process is performed at the receiving site to obtain a halftone. The other approach is to screen the contone first to get a halftone and then encode and transmit the halftone. These two approaches will be discussed in the next chapters separately.

CHAPTER 5

Coding Before Screening Approach

There is at least one inherent advantage in the approach of coding-before-screening approach. Since a contone picture is sent, it is easy to select the appropriate screen to threshold the contone into a halftone with the desired screen density. On the coding-before-screening approach, the coding scheme is not the major concern, any efficient contone coding technique can be applied.

From the nature of screening process, we know that some pictorial information is bound to be lost when image is converted from a contone to a halftone. If only the reproducible pictorial information can be extracted and transmitted, then the overall data rate can be reduced substantially. In other words, the problem becomes: "Given an original contone picture, what kind of information needs to be sent in order to produce a halftone with the equivalent quality as if directly screened from the contone original?" The most decisive factors in the amount of information content are the resolution and the contrast of the detail which usually takes a great percentage of total data volume in order to be reproduced properly. The amount of data is inversely proportional to the area of the sampling grid. By doubling the grid dimension in both the horizontal and vertical direction, the amount of data in the sampled contone can be reduced by a factor of 4.

In this chapter, we shall try to find the sampling grid required to produce a sampled contone from the original which in turn can be screened to a halftone with the same quality as directly from the original. Contone pictures obtained at equal or higher sampling density are then considered to be of full quality so far as halftone reproduction is concerned. This information is useful in comparing the coding efficiency for different approaches.

5.1. Maximal Sampling Interval

We must remember the psychophysical paradox that extremely fine sampling and quantization do not improve the quality of a picture nearly in proportion to the number of bits needed to represent this collection of numbers. As a matter of fact there are generally thresholds of sampling and quantization beyond which no further improvement is noted for any given picture.

Since a halftone picture needs only one bit to represent the grey values of each pel while contone uses between 6 and 8 bits, it is easy to jump to the conclusion that there is an instant eight to one compression ratio when a halftone is used instead of the 8 bit PCM. This actually is not true. The reason is that there is a trade-off between the number of grey levels and the sampling density. Within limits, when the number of grey levels is increased, the sampling density can be reduced while keeping the same visual quality.

The halftone process is one of the methods that trade texture for apparent grey levels. In a digital format, the number of pels in the size of a dot area is the number of grey levels that can be represented. In order to represent each dot of a halftone picture faithfully and to generate enough grey levels, a relatively high resolution is therefore required to scan a halftone. In the case of the Autokon electronic process camera, a 722 lpi resolution is used. To produce a 65 line halftone, a screen size of 8 by 16 pels is used. That creates 128 different grey levels for the dot. For 85 and 100 line halftones the screen sizes are 6 by 12 pels and 5 by 10 pels, having 72 and 50 grey levels respectively. Pictures are produced with quality comparable to the traditional optical process camera. If full quality is to be preserved, the halftone should be digitized with a comparable grid size. By using this high resolution, we can end up with more data to be transmitted than a contone picture with comparable quality in the uncoded form, even though we use only one bit to transmit the information for each pel instead of 7 or 8 bits in the contone case. For example, it takes 65 Kbytes of data to store one square inch of halftone generated by the Autokon as compared to 40 kbytes for the contone sampled at 200 lpi.

The problem in the coding-before-screening approach thus is to find the size of the sampling grid such that the quality of halftones produced from such pictures is as good as directly screening the original. We want to use a finer grid to preserve the quality while using a coarser grid to reduce the data volume. Obviously, this sampling density must be higher than the screen density. As to how fine the sampling grid needs to be, that is the subject in this chapter. An apparent limit is the pel density of the Autokon.

The amount of information and hence the volume of data in a halftone picture depends on the following factors:

(1) Screen density: The higher the screen density, the better the halftone can resolve. When a higher density screen is used, the sampling density is also increased.

(2) Screen orientation: According to experiment, human eyes are least susceptible to a screen oriented at 45 degrees with respect to either the horizontal or vertical axis when a rectangular screen is used. Instead of eight fold symmetry as for the rectangular screen, the hexagonal screen is 12 fold symmetric. Theoretically, the results of the hexagonal screen should be less sensitive than the

rectangular halftone to the screen orientation. This factor does not affect the required sampling density.

(3) Screen shape: provided it has the same distance between dots, a hexagonal screen has the most uniformly distributed frequency spectrum in any direction. This could translate into better detail reproduction, for a given sampling density. In this thesis, only the traditional rectangular screen is considered.

(4) Nature of the contone picture to be converted into halftone. Details with high contrast tend to be reproduced better than those with low contrast, thus requiring high sampling density.

5.1.1. Experimental Method and Procedure

In order to find the sampling grid required to produce pictures undiscernible from the originals, this experiment was conducted. Pictures with different characteristics are chosen. The first is the GIRL face, which consists mainly of plain texture with some high frequency detail in the hair and eyebrow area. The second is the CROWD picture that is rich in medium to high spatial frequency components. And then the resolution test pattern and type font from the standard IEEE facsimile test picture. This consists mainly of sharp, high contrast edges as well as high resolution patterns.

First, pictures are scanned by the Autokon with the desired sampling density. The digital binary pictures are stored on disk, and then sent back to the Autokon for halftone output by the hardware screen. The output of the Autokon uses 722 pels per inch resolution, so pictures sampled at lower density are scaled by either the sample-and-hold or by linear interpolation. For sample-and-hold, pictures can be sent directly to the Autokon using the repeating pel, repeating line feature. For this reason, integer fractions of the 722 lpi resolution are used in the input sampling density. Those numbers are 361, 241, 181, 144, 120, 104, and 90 lpi. When the input sampling resolution is too low (less than 120 lpi), then linear interpolation is needed in order to get rid of the square structure. The quality of the output pictures is then compared. For each picture and screen density, a subjective threshold is chosen, above which the sampled contone is thought to produce the same quality as the original. The threshold values are then plotted as a function of screen density and sampling density.

5.1.2. Experimental Result for Resolution Test

(1) For the GIRL face, representing pictures consisting mostly of low spatial frequency components, coarse sampling does not seem to affect the quality of the resulting halftone. There is gradual but barely visible

degradation in picture quality as the sampling density decreases. When compared to the halftone made by directly screening the original, the difference is relatively small even at sampling density as low as 90 lpi. This picture is not sensitive to different screens. Using a higher density screen does not considerably improve the quality.

(2) The CROWD picture needs higher sampling density in order to reproduce the full quality of the direct screening. This is due to the fact that it has a large amount of information in the high and medium spatial frequency range. Even so, with a 120 lpi sampling density, the resulting halftone almost cannot be distinguished from the direct screening of the original. There is almost no quality improvement for sampling above that density.

(3) For the resolution test pattern and type fonts, the deterioration of picture quality is gradual. There is no clear-cut threshold. The improvement in picture quality is more visible as the sampling density is increased. Also because of the nature of higher contrast as well as more regular pattern, it takes a higher sampling density to achieve the best achievable quality.

(4) Type fonts also require a higher sampling density. That is because of the high contrast sharp edges. Even so, at around 180 lpi, it can be reproduced almost as well as

directly coming from the Autokon. Also the improvement by using high density screens is more obvious in this case.

(5) In general for halftones with finer screens, it requires a higher sampling density in order to reproduce the full quality. This is quite intuitive, as the finer the screen, the more information that can be preserved during the halftone process.

(6) A sampling density of 300 lpi is necessary for the case of reproducing the test pattern from the IEEE facsimile test picture if we really want to produce original quality including the Moire patterns.

In summary, the a sampling density required to produce a full quality halftone is shown in Figure 5.1.

5.2. Resolution versus Contrast

In the last section, we dealt with the pictorial information that is reproducible by the halftone process and thus should be preserved in the sampling process. This is accomplished by using an appropriate sampling density for different types of pictures. In this section, an experiment is devised to find the contrast required, as a function of resolution in a detail area, in order to be properly reproduced by the halftone process.

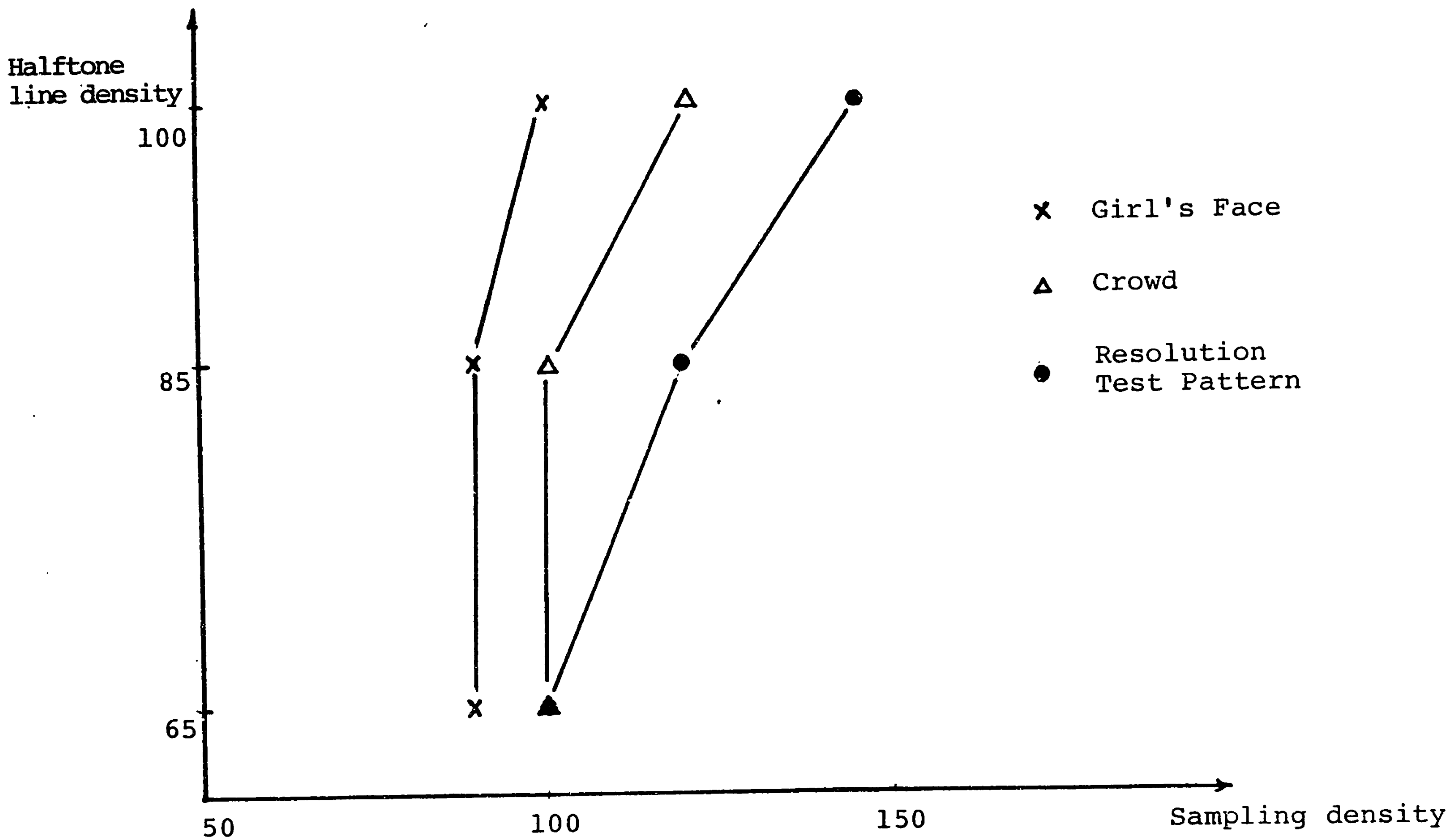


Figure 5.1 Sampling density for full quality halftone reproduction.

5.2.1. Experimental Method and Procedure

The resolution wedge of the IEEE facsimile test chart is selected for this experiment which is calibrated in lines per inch.

The test pattern is first scanned by the Autokon and stored on the disk at full resolution of 722 lpi. A tone scale transformation is then performed, whose purpose is to create an image with only two grey levels at the desired value. By selecting the appropriate tone scale table for transformation, contone test patterns with different contrast at different background average grey levels are created.

The resulting test pattern is then sent back to the Autokon for halftone output with different screen densities. The halftone representations of the test patterns are included in Figure 5.2 to Figure 5.5. These halftones are then carefully inspected. A threshold value for each test pattern is stated, above which the test bar pattern is thought to be unresolvable. This selection of this threshold is, of course, subjective.

5.2.2. Experiment Results

The relationship between the reproducible resolution and the contrast for different backgrounds is plotted in Figure 5.5.

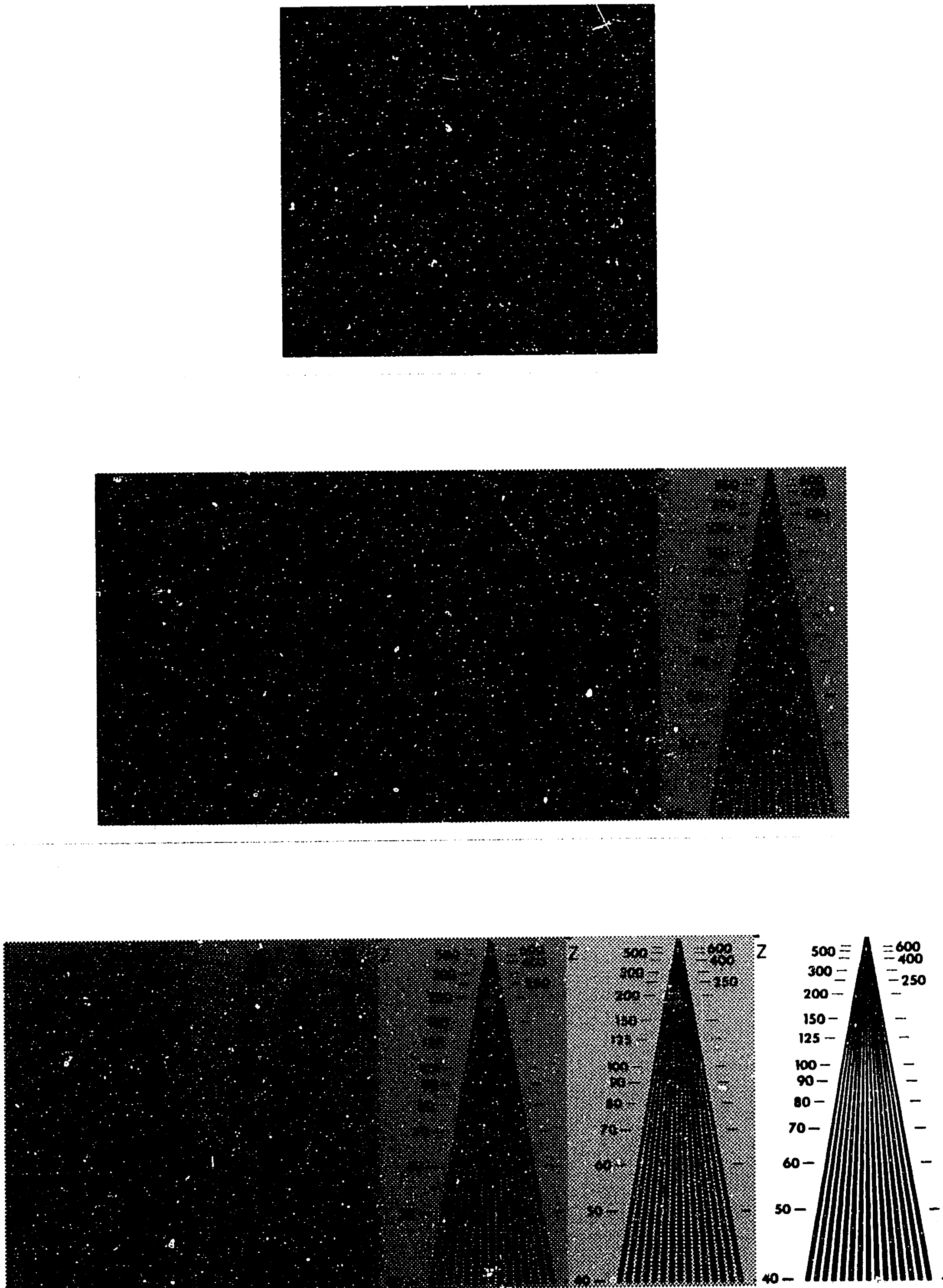


Figure 5.2 Test pattern for resolution of 65 line halftone.

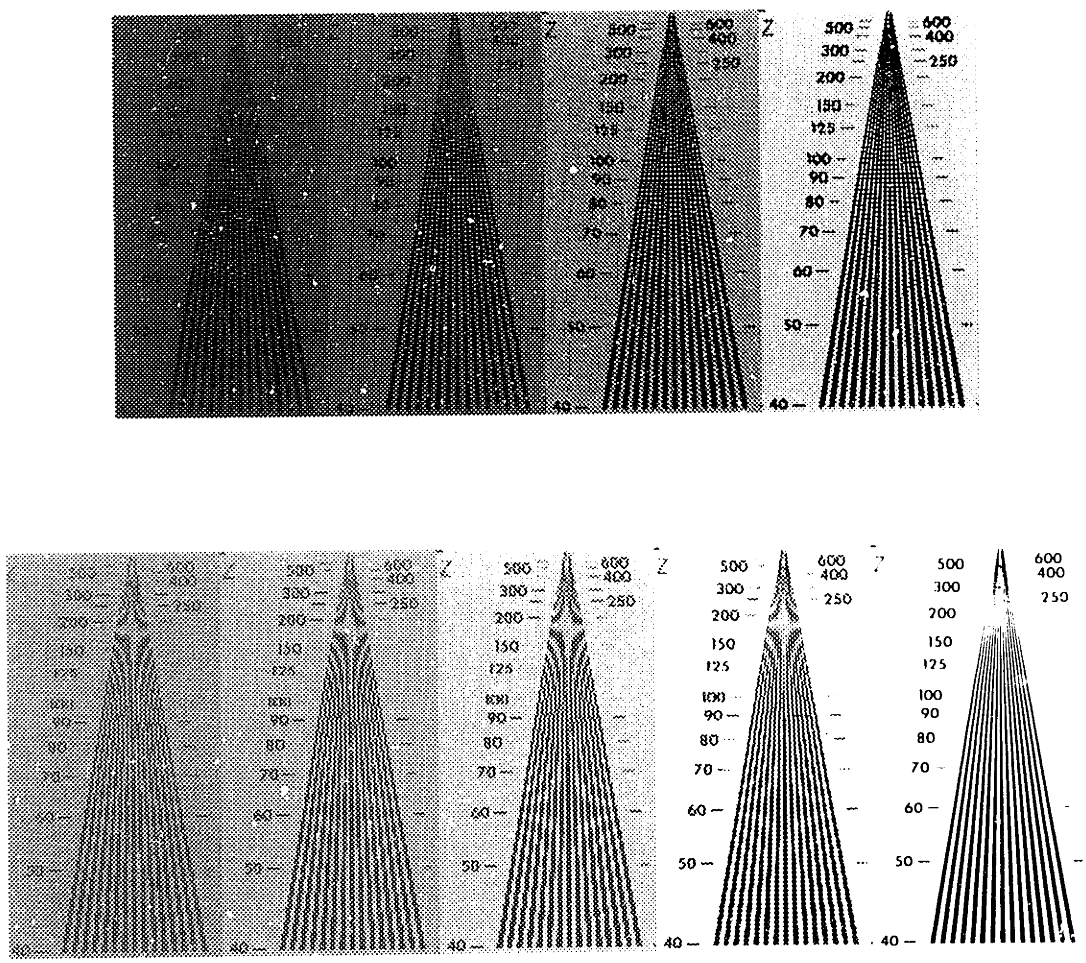


Figure 5.2 Test pattern for resolution of 65 line halftone.

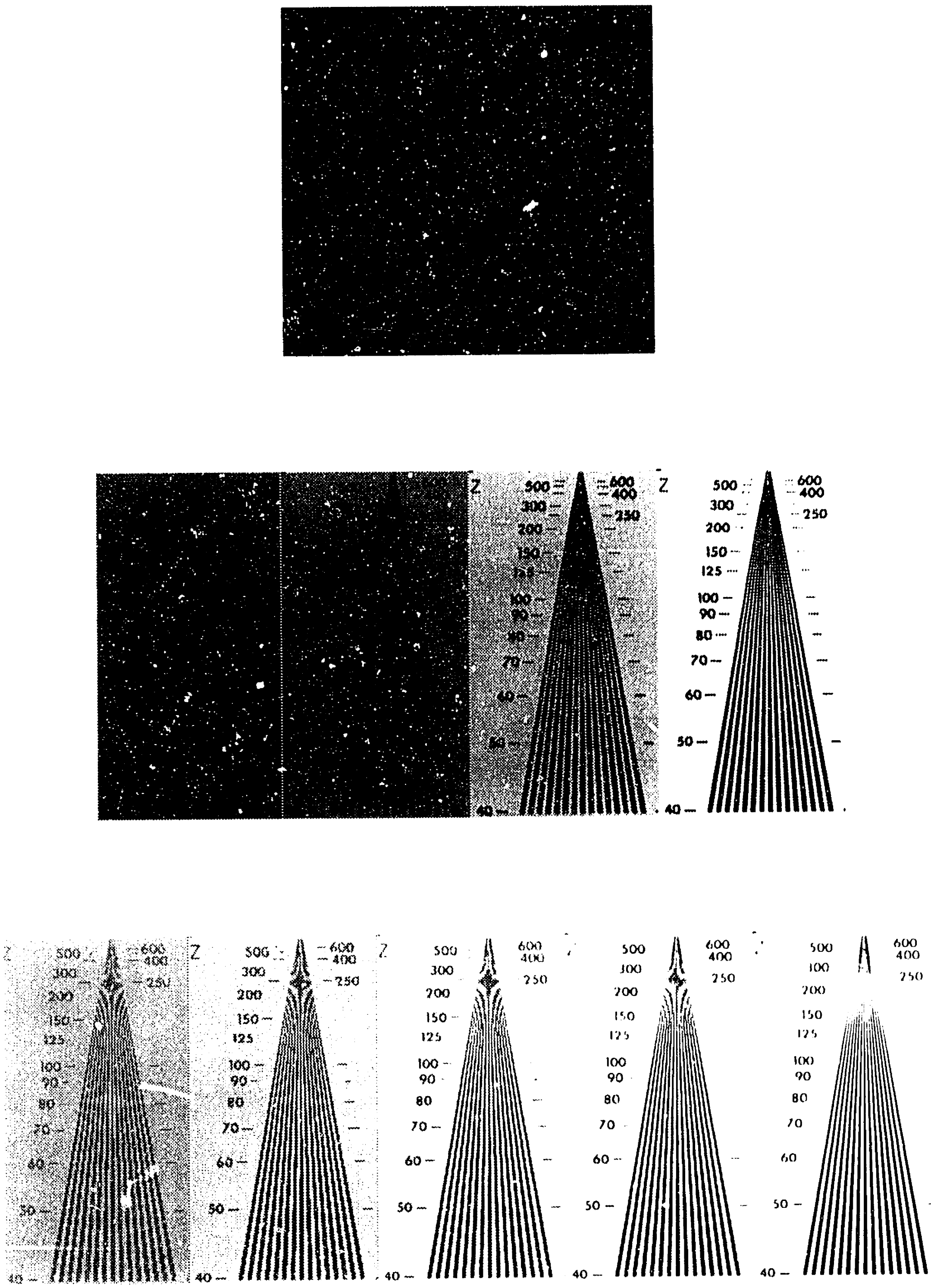


Figure 5.3 Test pattern for resolution of 85 line halftone.

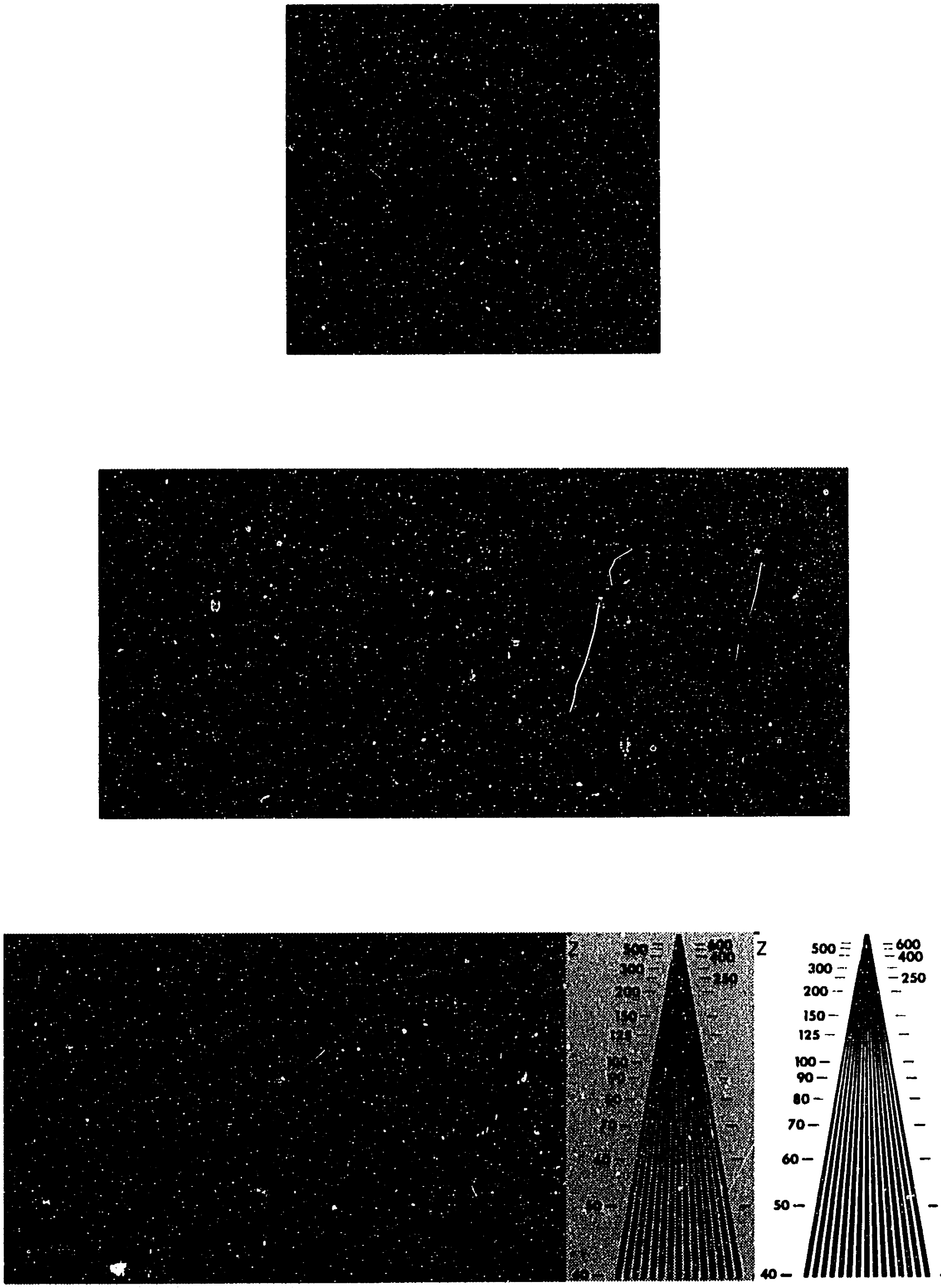


Figure 5.3 Test pattern for resolution of 85 line halftone.

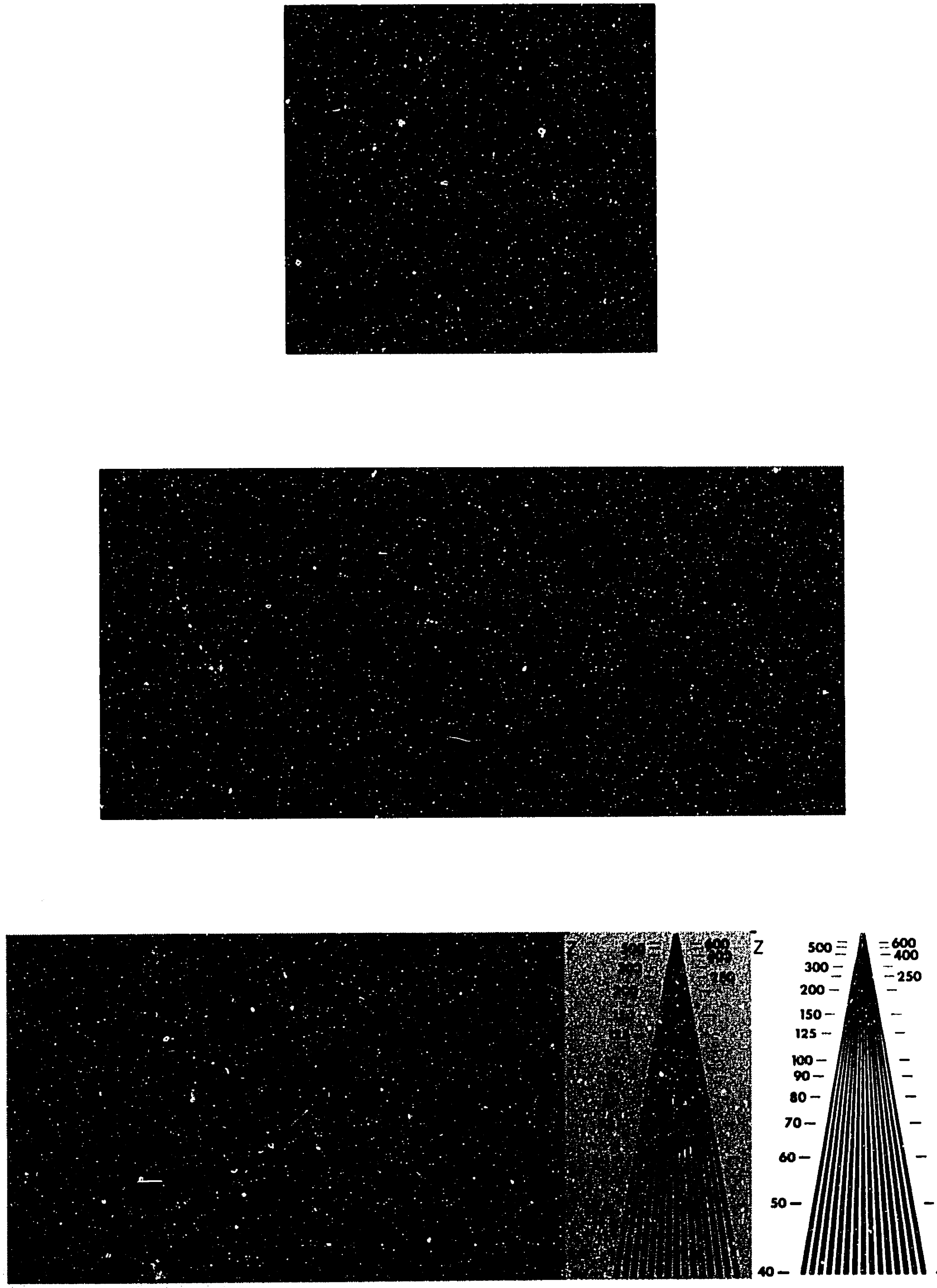


Figure 5.4 Test pattern for resolution of 100 line halftone.

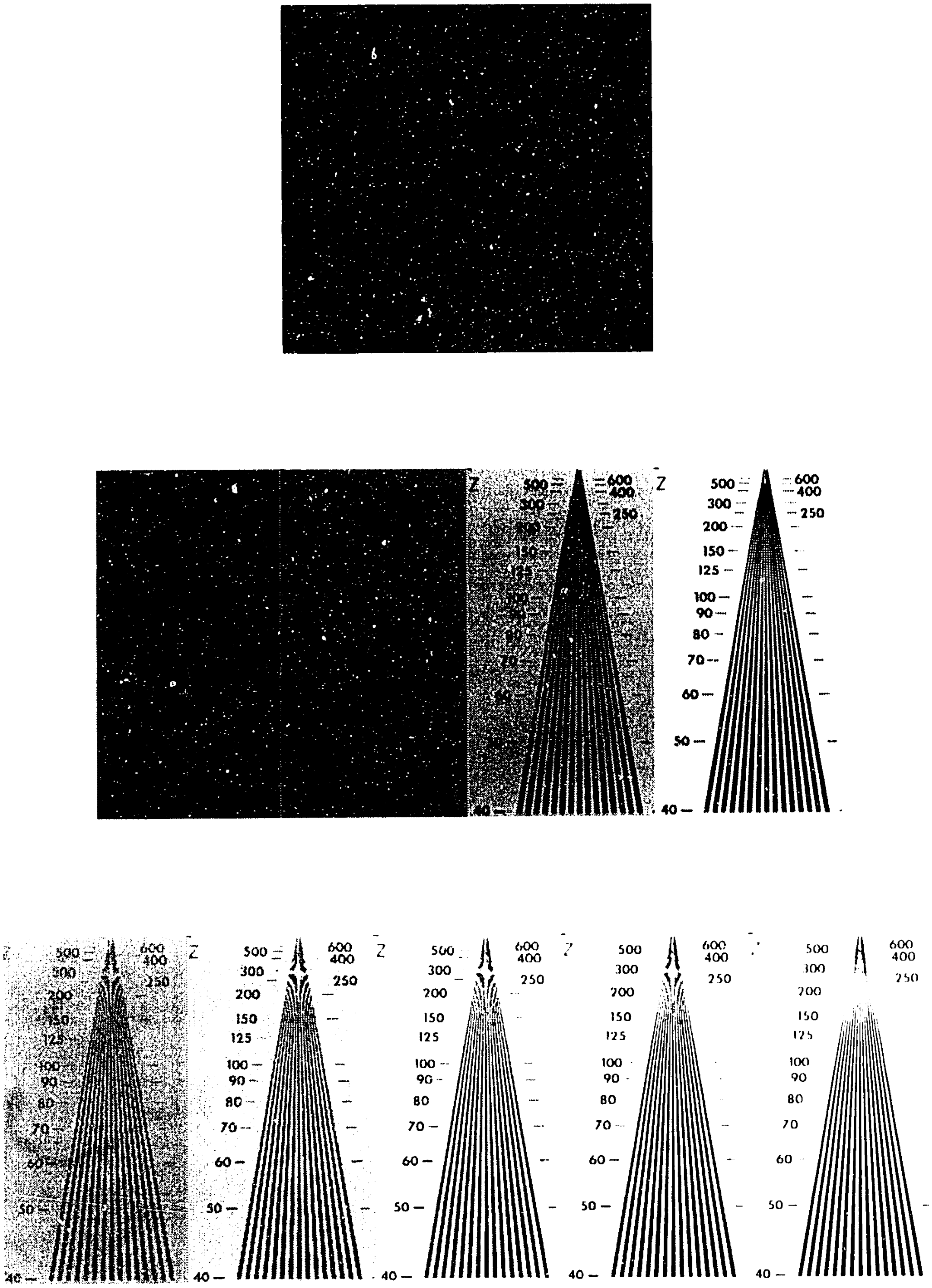
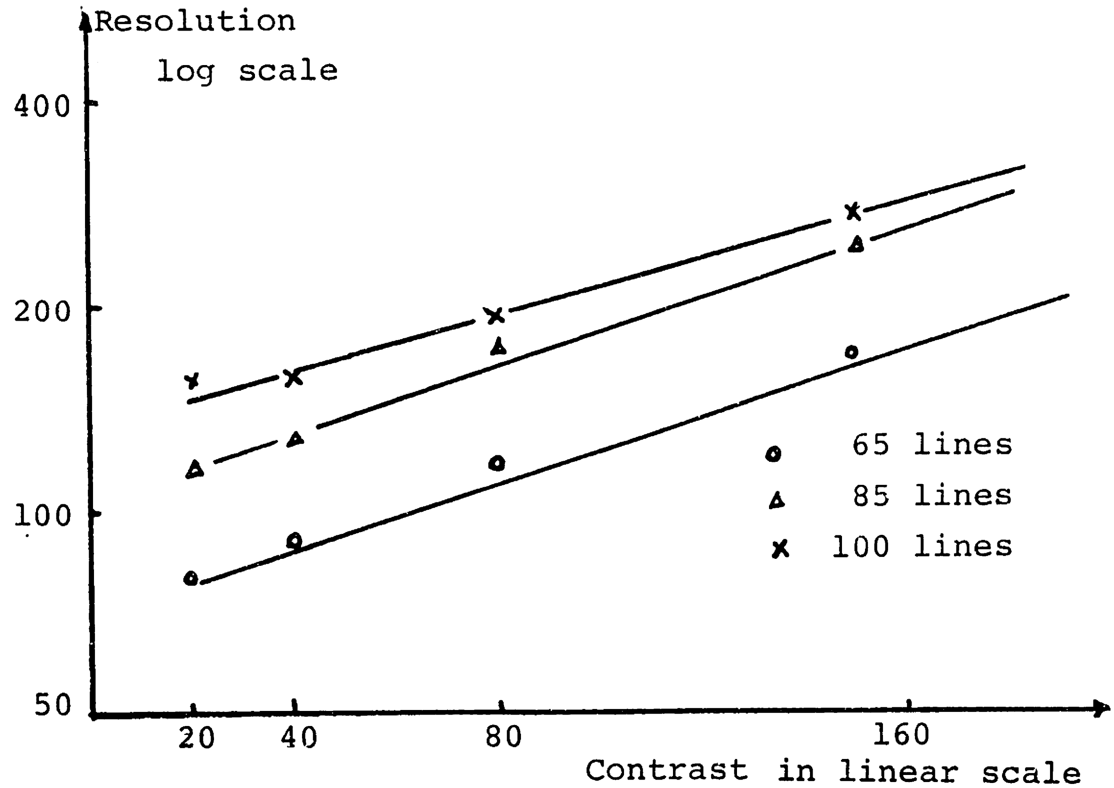


Figure 5.4 Test pattern for resolution of 100 line halftone.



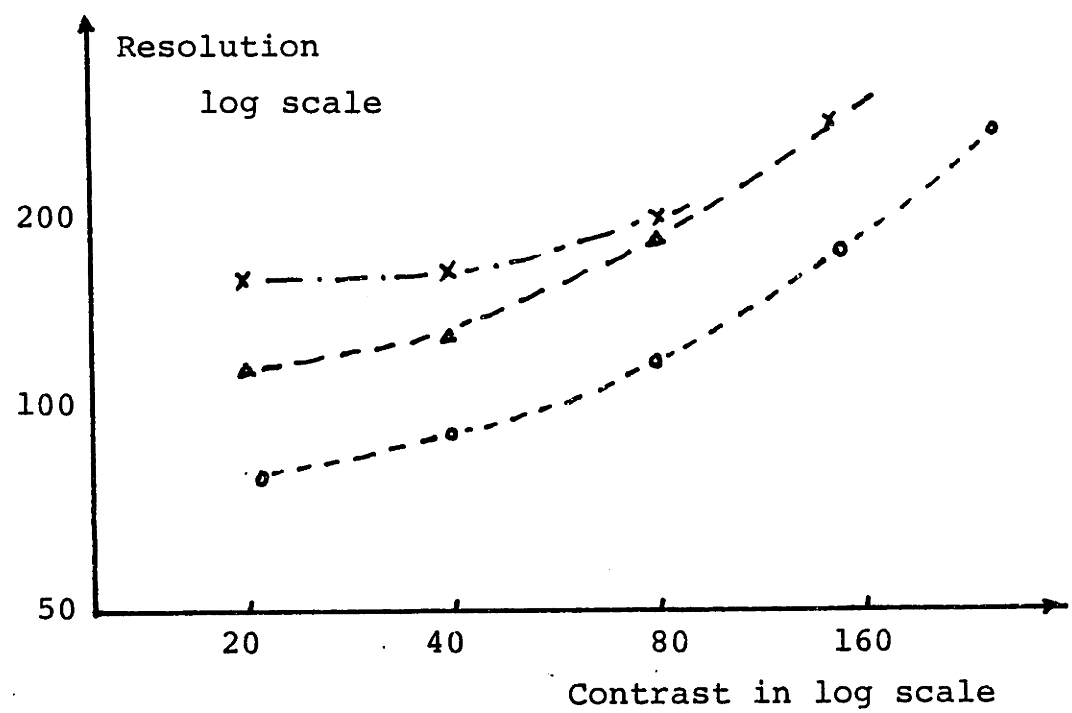


Figure 5.5 Result for resolution vs. contrast test. (a) in semi-log scale, (b) in log-log scale.

If the relationship of log of resolution versus contrast is fit with a line, the following equation can be obtained:

$$R = f(S) \exp [\ k(S)\ C] \tag{1}$$

where R is the subjective resolution, S is the screen size, and C is the contrast.

From the experiment, we have

a. for 65 line screen

$$R = 65.6 \exp (C/153) \tag{2}$$

b. for 85 line screen

$$R = 112 \exp (C/275) \tag{3}$$

c. for 100 line screen

$$R = 138 \exp (C/323) \tag{4}$$

The physical meaning of this experiment can be summarized as follows:

(1) With the same contrast, a slight improvement in resolution can be achieved if the background grey level is either in the highlight or the shadow instead of midtone. This is quite reasonable, because in the midtone the screen tends to create the regular dots instead of the more drastic shape as in highlight or shadow. The dot structure is the most visible in the midtone as the halftones have alternating black and white

dots. Signals are more easily masked by this strong dot structure. On the other hand, in the highlight and shadow area the dots tend to be small and change their shape according to the texture of the picture.

(2) Increasing contrast results in a better resolved test pattern just as expected. Thus when a picture has high contrast detail it requires more data to reproduce it properly than if it had low contrast.

(3) When a higher density screen is used, the resulting halftone can also resolve a better test pattern. This implies that when a high density screen is used, more information will be preserved and a higher sampling rate is required in order to keep this pictorial information as indicated in the last experiment.

(4) The relationship between resolution and contrast can be formularized in semi-log coordinates (log resolution vs. contrast increment) as shown in equations (2) through (4). Improvement in resolution by increasing the contrast is small when the contrast is low. Below a threshold value of 80 in a 0 - 255 scale there is little or no discernible effect with increasing contrast. Above 80, the improvement increases sharply with the increase in contrast. This value may therefore be used as a measure to decide what kind of pictorial information must be preserved.

5.3. Results Using a Two-Channel Coding Scheme

To test the efficiency of the coding-before-screening approach, pictures are sampled at 180 lpi, which is sufficient for up to a 100 line screen in almost all cases. The sampled contone picture can then be compressed by any coding scheme described in chapter 4. As a comparison, the two channel coding scheme developed in MIT CIPG is used. In the highs channel, both 2 bits and 3 bits have been tried, both with and without the adaptive technique. A 5x5 subsampling area is used for the lows channel. When only two bits are used in the highs channel, it is found that without the adaptive encoding, the pictures are too noisy. This is especially visible in the IEEE test picture. When 3 bits for each pel are used in the highs channel, the quality is as good as the original. When only two bits are used for the highs channel, the effect generated by the pseudo-random noise becomes visible in some area. The noise problem can be substantially reduced with slightly degraded coding efficiency. A compression ratio about 5 can be achieved with respect to the halftone obtained by scanning the original contone with the Autokon.

Overall, this approach uses 722/4 = 181 pel per inch sampling density for the original, which is sufficient for most pictures.

When three bits are used in the highs channel, we need on the per pel basis:

3.00 bits	for highs channel		
.28 bit	for adaptation	(7 bits over 5X5 area)	
.32 bit	for lows channel	(8 bits over 5X5 area)	

3.60 bits total for a contone pel

Thus it takes

$$3.60x180.5x180.5=117288bits=14661bytes$$

to encode one square inch of equivalent halftone pictures. The compression ratio is 2.22 (8.0/3.6) with respect to the contone PCM. When compared with the halftone original from the Autokon, the compression ratio is 4.44, i.e.,

$$\frac{(722)^2}{8} x \frac{1}{14661} = 4.44$$

If only two bits is used for highs channel instead of three, we have:

2.00 bits	for highs channel	
.44 bit	for adaptation	(4 bits over 3X3 area)
.32 bit	for lows channel	(8 bits over 5X5 area)

2.76 bits total for a contone pel

or 2.76x180.5x180.5 = 90064 bits = 11258 bytes per square inch.

The compression is thus 2.9 (8/2.76) with respect to the sampled PCM contone. Or when compared to halftone original from the Autokon, it is 5.79.

$$\frac{(722)^2}{8} x \frac{1}{11258} = 5.79$$

If the picture is to be sent to a different sites with various printing facilities, this can be a good approach because of the flexibility.

There is a drawback in this two channel coding scheme however. When high contrast edges are present in the picture, the subsampled contone after interpolation and clamping, cannot produce sharp edges. The dynamic range is also somewhat decreased. This makes the picture look much fuzzier than the original in the high contrast detail area.

CHAPTER 6

Coding after Screening Approach

While the coding-before-screening approaches are discussed in the last chapter, the coding-after-screening technique will be discussed in this chapter. According to a data processing theorem in information theory [21], the more processing a picture is subject to, the more information about the original can be lost. When the processing is not invertible, there is always a decrease in information content. The halftone screening process is non-linear and non-invertible, and, as a result, there is always some information loss during the screening process, With this fact in mind, it seems logical to suggest that it would be more economical to code the halftone in its final binary format.

In this chapter, the binary picture is run-length encoded. Before the run-length encoding can be used efficiently, the picture must be preprocessed to get better statistics and make the run-length encoding more effective. Two preprocessing methods are investigated for this purpose. In the first section, we discuss the use of a multi-pel predictor, while a new method based on separation of low and high detail signals is discussed in section two.

6.1. Multi-Pel Predictor

Run length coding, when directly applied to a halftone picture, usually results in a limited compression ratio. The reason is that information about the screen structure used in creating the halftone is not incorporated in the coding scheme at all. Generally speaking, the periodic dot structure in a halftone breaks the picture lines into a large number of short runs. This is especially true near the midtones. This results in shorter average runs and thus hampers the efficiency of the scheme. One approach to this problem is to rearrange the black and white pel, according to both the average grey level in the local area and the structure of the screen.

In DPCM coding of still pictures, the grey value of a pel is predicted based on its neighboring pels. The error is transmitted using fewer bits partly due to the smaller expected error. Essentially, it utilizes the redundancy between adjacent pels to achieve data compression.

In halftone pictures, a similar technique can also be applied. It is possible to predict the polarity of a pel based on the neighboring pels and then send the error.

A halftone picture is thus transformed into another binary picture with a statistical distribution more suitable for data encoding. By using the correlation between pels, there is high probability of the polarity of a pel being correctly predicted. If a "0" indicates a correct prediction and "1" a wrong prediction, the resulting picture clearly will have quite uneven distribution of 0's and 1's if a predictor is properly designed. With the better distribution, the new binary picture can then be encoded by any run length coding scheme to get a better coding efficiency.

Just as in the case of higher order DPCM, where the redundancy of pels close but not adjacent to each other is considered in the coding scheme, a binary predictor can be based not only on adjacent but on nearby pels as well. An especially important factor in designing a predictor is the periodic structure of the halftone picture. Due to this periodic screen structure, the polarity of a pel is closely related, not only to its neighboring pels, but also to those pels that are one dot period away and subject to the same threshold screen value when the halftone was made. Such pels have a strong tendency to be of the same polarity while pels that are a half period apart tend to be inversely correlated and pels with quater screen period apart tend to be much less correlated even though they are geometrically closer. This has been illustrated in Figure 4.1, where the autocorrelation of a halftone picture is shown. The strong peaks and valleys at the screen period and half period can be clearly seen.

Usubuchi et al [93] have applied this technique to the adaptive transmission of newspaper pages. An optimal predictor of this type is used to rearrange the binary pels and thus achieve a better compression ratio.

Another method to achieve better coding efficiency is by rearranging the pels according to their statistical property. This method has been suggested by Netravali [64] and adopted by Usubuchi. If a pel is predicted according to these two parameters, an error bit will be produced to indicate a correct or a wrong prediction. For pels that are in the middle of a black or white dot, the prediction tends to be more exact, which is termed "good prediction". On the other hand pels that are on the edges of either black or white dots (i.e., pels that are close to the thresholding point in the dot) tend to have a larger prediction error, termed "bad prediction". Netravali et al [66] have used the same terms for binary facsimile pictures. If the bad and good prediction is now redistributed in such a way that good pels are put on the left hand side of a line and bad pels are put on the right hand side, then the average run length can be increased. This approach utilizes the redundancy of the screen pattern, but not the redundancy of the statistical dependence of adjacent pels in grey value.

Predictors based on more than one neighboring pels have been implemented to produce a map of erroneous prediction. As we can expect, these error pictures have different distributions for black and white pels. If the predictor is appropriately built, the number of white pels (which indicate a prediction error) is much smaller than that of black pels (which indicate correct prediction). Furthermore, this prediction error depends more on the characteristics of pictures such as the complexity of texture, number of sharp edges etc., rather than the overall apparent grey value of the picture. This uneven distribution can add efficiency to any binary coding scheme such as runlength coding.

In this thesis, run length coding is used to do such block encoding using the standardized one dimemsional CCITT code.

Predictors used in this thesis include the following:

(1) Predictor based on 1 to 12 close neighboring pels. Pels that are one screen period away are not included, although pels from different lines may be included. The arrangement of pels is shown in Figure 6.1.

(2) Predictor based on close neighboring pels as well as pels that are one screen period away. The number of total pels used for prediction varies from 2 to 12. The geometric relation of the predicting pels is shown in

```
                       X                       XX
XP                    XP                      XP

(1)                   (2)                     (3)


 XX                    XXX                    XXXX
XXP                   XXP                     XXP

(4)                   (5)                     (6)


                         X                       XX
 XXXX                  XXXX                    XXXX
XXXP                  XXXP                    XXXP

(7)                   (8)                     (9)


 XXX                    XXX                     XXXX
XXXXX                  XXXXX                   XXXXX
XXP                   XXXP                    XXXP

(1Ø)                   (11)                    (12)
```

Figure 6.1 Predictor using only the neighboring pels.

```
                                 X                       X
X       XP              X       XP              XX      XP
|<- T  ->|              |<- T  ->|               |<- T ->|

   (2)                     (3)                     (4)


        XX               X      XX               X      XXX
XX      XP              XX      XP              XX      XP
 |<- T ->|               |<- T ->|               |<- T ->|

   (5)                     (6)                     (7)


 X      XXX              XX     XXX             XXX     XXX
XXX     XP              XXX     XP              XXX     XP
 |<- T ->|               |<- T ->|               |<- T ->|

   (8)                     (9)                     (10)


         X               X       X
XXX     XXX             XXX     XXX
XXX     XP              XXX     XP
 |<- T ->|               |<- T ->|

   (11)                    (12)
```

T = screen period

Figure 6.2 Predictors using pels a dot period away.

Figure 6.3. Pels from one to three pictures lines are required.

The predictor is obtained by a program that checks a sample picture and tabulates all the possible distributions. After the entire picture is checked, the polarity of higher occurrence frequency is chosen for prediction. For example, if in the final tally for pattern A, there are more occurrence of A0 than A1, then a 0 is taken to be the polarity when pattern A is encountered.

Numerical data for the performance is listed in Figure 6.3. The result is somewhat surprising.

In summary:

(1) The predictor is sensitive to the characteristics of the picture itself. Two sets of data are obtained using predictors derived from (a) part of a picture and (b) a grey wedge distribution for generalized purpose as shown. For data obtained from (a), as expected, as the number of neighboring pels is increased, fewer prediction errors are made, thus resulting in fewer error bits, a better distribution of run length, and improved compression ratio. A predictor using pels one dot period away is proved to be better than one using only pels surrounding the current pel. The correlation of neighboring pels is better utilized in this approach. As for the predictor

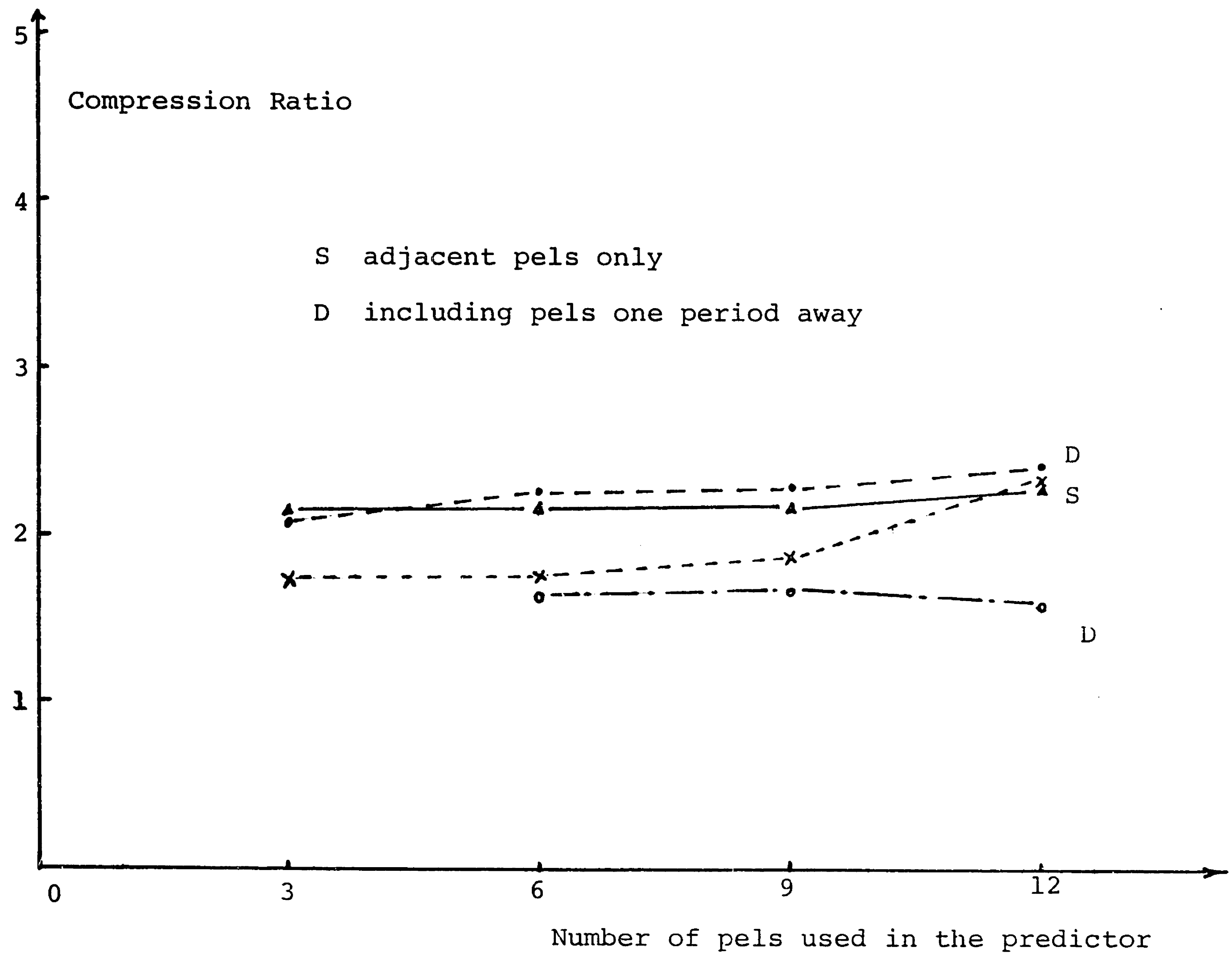


Figure 6.3 Performance of different predictor.

in (b), the result is surprising. It is found that by using the generalized predictor, the result does not improve as the complexity of prediction increases. Rather there is a peaking effect, above which, a further increase in the complexity does not improve the performance. The reason for this effect is that the predictor is a generalized and not matches the distritution of the particular picture to be coded. The correlation does not increase above a certain number of pels as more complex pictures are used. Another reason is that the grey wedge is not truly representative of normal pictures due to the simplicity of it distribution.

(2) For the same number of previous pels used in a predictor, those utilizing pels one period away seems to have a better result than those which use only close neighboring pels. This is especially true for simple pictures. For very complex pictures, the difference between the two sets of pels is smaller and in some cases, even the reverse is true. The physical meaning of this result is that as the texture of the picture becomes more complex, the dependence of a pel value on that of its immediate neighbor is greater then on pels a period away but corresponding to similar screen phase. The screen pattern becomes less important to the correlation as the complexity of picture becomes greater. This is similar

to the contone case, in which, as the picture becomes more complex, its autocorrelation becomes narrower than in a simple picture.

(3) For pictures rich in high contrast fine detail, the compression ratio is generally lower than pictures of simple structure. As the complexity of the predictor increases (i.e., more pels being used), improvement in performance does not increase noticeably, peaking in the saturation value rather quickly. In some cases, the compression ratio is not improved at all as compared to straight runlength encoding without any of the above transformations to improve the uneven black/white pel distribution.

5. For simple pictures, the compression ratio is usually improved more noticeably as the complexity of predictors is increased.

The rearrangement of pels according to adjacent pels as well as those one screen period away essentially takes care of both aspects of redundancy, that of the picture itself due to the close correlation of neighboring pels, as well as that due to the screen structure.

6.2. A New Approach

In the coding of contone pictures, Yan and Sakrison [97] have implemented a coding scheme in which signal components are first broken into two parts. Two different coding techniques, each tuned to the characteristics of these two signals, are then used to encode them. The picture is divided into low contrast and high contrast components. The two channel coding technique [84] uses a similar principle but a different technique. The signal is decomposed into a low spatial frequency part and a high spatial frequency part. The low spatial frequency part is then subsampled coarsely but finely quantized. The high spatial frequency part is sent at full sampling rate but its grey level coarsely quantized. A similar principle is applied in the approach discussed in this section. A halftone can be decomposed into a low frequency and a high frequency component. Each can then be sent using an efficient coding scheme. In the low frequency channel, only the average grey level is transmitted. At the receiving end, a replica of a low pass version of the halftone is reproduced using the same screen structure as at the sending site. For pictures of simple texture this information can be used to reconstruct the entire halftone with satisfactory results. On the other hand, if fine detail is abundant in the original halftone, extra information must be transmitted. The process can be implemented by transmitting the difference between the original and the reproduced low halftone. An exclusive or is performed to get the "difference picture". In plain areas, there is almost no difference. It is only in the fine detail high contrast areas, that the difference is substantial. In those areas, the average grey level is no longer enough, and extra information is transmitted. The block diagram for this coding scheme is shown in Figure 6.4.

6.2.1. Construction of Two Channel Signals

The low channel picture is obtained by combining the local average of the halftone. In this method, the picture is first converted to a low resolution contone using the dot size averaging technique discussed in chapter 3. The averaging mask for each screen density is selected and a subsampled process is then performed. The low pass characteristics of the dot size averaging and the coarse sampling in the conversion ensure that the contone comprises only low spatial frequency component. It essentially repeats the average grey level over a relative large area. This low channel signal can then be encoded by any contone coding scheme. Due to the small information content compared to the fine detail part, the signal is simply sent in direct PCM.

The next step is to construct the highs channel. A low passed version of the halftone picture is reconstructed using exactly the same screen structure and phase.

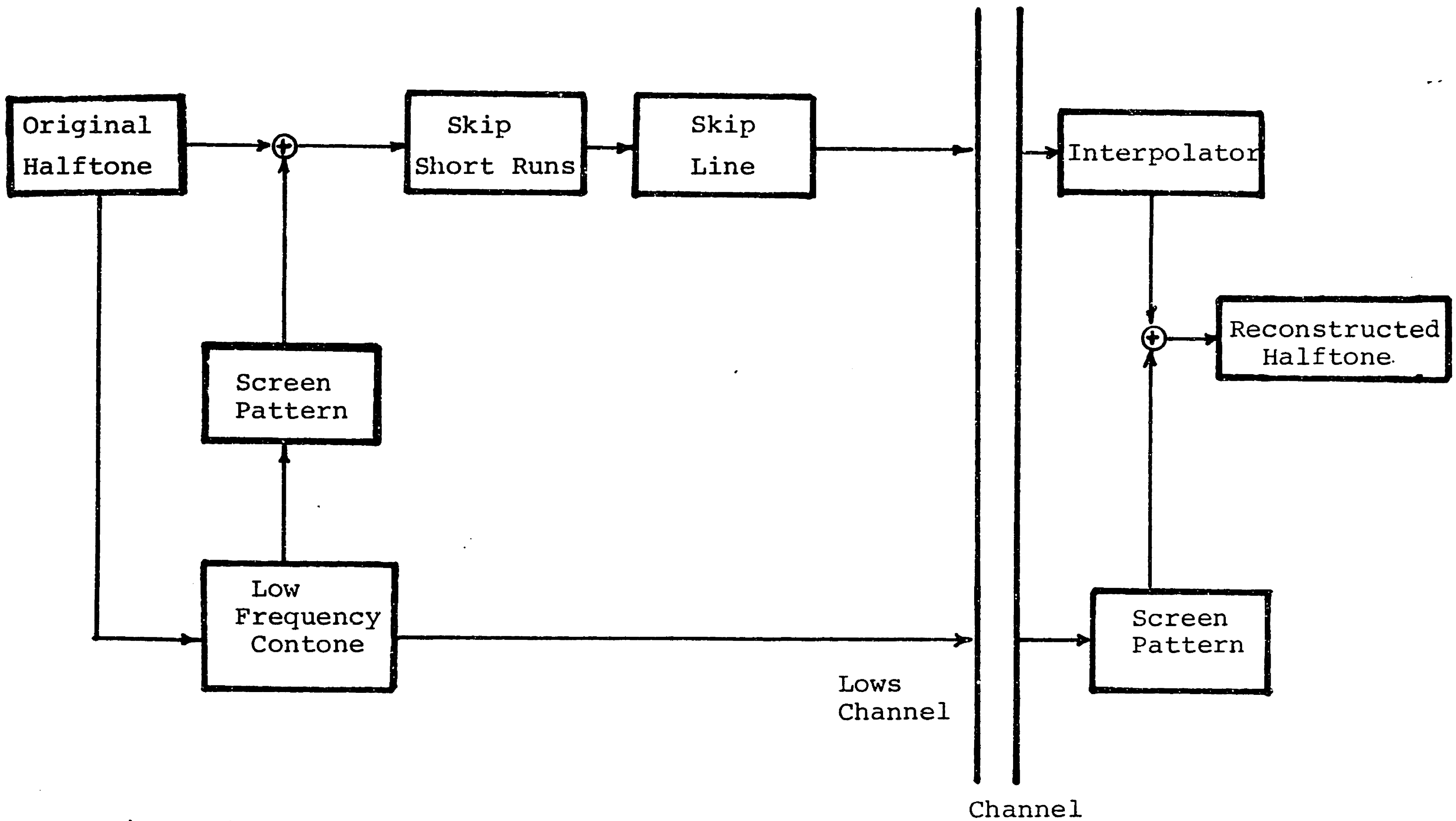


Figure 6.4 Block diagram of two channel halftone encoding.

After this conversion, all dots are of regular shape, due to the low frequency characteristics. This picture will be found sufficient for some printing purposes that do not need full quality but certainly is not suitable for high quality printing, since much detail would be lost.

To recover the detail portion of a halftone, a difference picture can be obtained by comparing the original with the reconstructed version using an "exclusive or" operation. This high detailed picture can then be stored or transmitted separately. A high quality image can now be obtained by superposition of the two separate pieces of information.

Decomposing a halftone picture into two parts has some advantages. First, it removes the redundancy between adjacent pels. The low pass version contains substantially all the information related to the high correlation for grey levels of adjacent pels. Second, it is easier to extract and eliminate the pictorial information that is not sensitive to the human visual system from the difference picture.

6.2.2. Selection of Subsampling Factor

In decomposing halftone pictures, the subsampling rate in the lows channel must be selected. There is some efficiency trade-off between the subsampling density and coding of the difference picture. For a dense sampling, more bits are allocated to render the apparent grey values. Due to the finer sampling grid, more detail information about the texture is preserved. On the other hand, coarse sampling decreases the data in the low spatial components. But more bits are then needed to reproduce the lost detail because more information is now in the difference picture. For pictures with a small amount of detail and mostly flat areas, a better choice is a slightly larger subsampling factor. Conversely, for pictures rich in fine detail, a smaller subsampling factor gives a better result.

The total number of bits in the original halftone has been plotted versus average area (i.e., subsampling distance) in Figure 6.5. It is easily seen that as the subsampling grid becomes finer, the average grey level conveys more information about the texture. This process is of course limited by the line density of the halftone. When the subsampling grid is comparable to the dot size, there is no further improvement available by shrinking the subsampling grid because the areas of averaging will overlap with each other so that no detail with higher frequency can be reproduced in the averaging process.

6.2.3. Skipping of Short Runs

Characteristics of the difference picture have been checked carefully. By calculation as well as theoretical analysis, it is found that if an invertible encoding is to be implemented, there is only moderate compression by decomposing the halftone into two parts. The total number of runs in a typical halftone picture is shown in Figure 6.0. The theoretical limit is obtained by calculating the entropy using

$$\text{Entropy} = \sum_{i=1}^{\text{all length}} \left(\frac{N_i}{S}\right)\log_2\left(\frac{N_i}{S}\right)$$

where

$$S = \sum_{i=1}^{\text{all length}} N_i$$

is the total number of runs. and N_i is the number of runs with length equal to i pels. The entropy calculation gives a good lower bound for the average number of bits required to encode each run. A lower bound for the total number of bits required can be easily obtained by

$$\text{Total bits} = S \cdot \text{Entropy}$$

When there are only low frequency signals in the picture, the difference picture will consist almost entirely of zeroes because the lows channel essentially contains all the information.

For a complex picture, there are more runs with shorter length. The total number of runs versus different subsampling area is plotted in Figure 6.5 when short runs are skipped. As the texture becomes more complex, the information content in the difference picture increases quickly. That means more runs with shorter length show up in the highs channel. The coding efficiency decreases very fast. Eventually, we cannot do any better by this means than by directly encoding the original halftone. Then what is the point of using this two channel coding scheme? The answer is that by using this scheme, the redundancy due to the visual system can be detected and removed easily. By removing the redundancy in the difference picture, a higher coding efficiency can be achieved.

It can be seen that a good percentage of the runs are 1 or 2 pels long (46.53% and 18.76% respectively for IEEE test picture). If these short runs can be neglected, then the total number of runs will be substantially reduced and the average length increased. The next question is whether those short runs are important in determining the quality of a halftone. If they affect the quality appreciably, then we have no other choice but to transmit them and the scheme will not be efficient. On the other hand, if they are not important to the visual system, then, we may proceed to omit them and improve the efficiency at the cost of minor picture

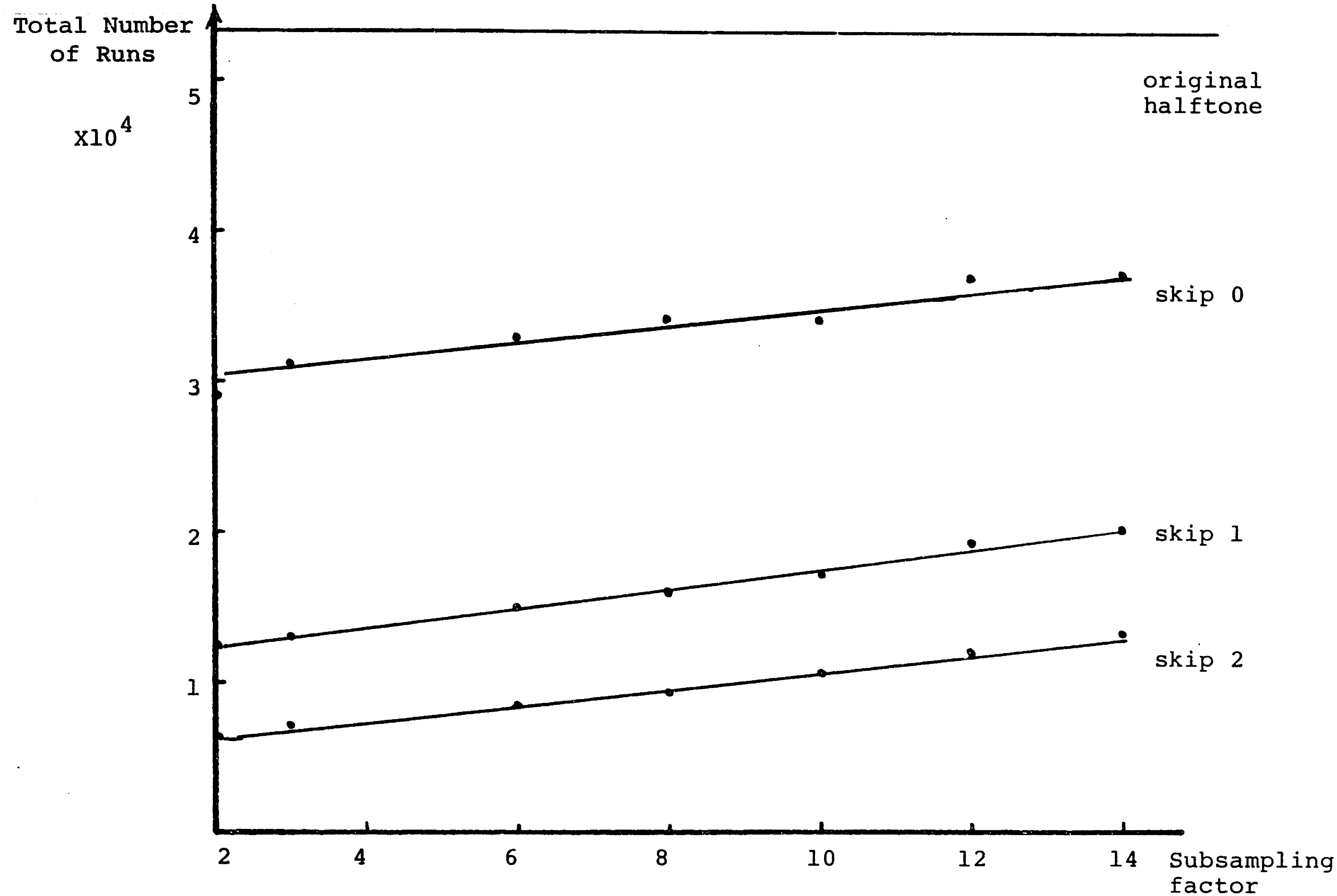


Figure 6.5 Distribution of runs in highs channel.

degradation.

The Autokon uses 722 pels per inch as the resolution for binary pels. The major reason is the requirement for rendition of enough grey values in a dot size area. As mentioned before, too few grey values can create a contouring effect. When the resolution is increased to above 300 lpi at a normal viewing distance of 12 inches, the response of the visual system has tapered off so much that a test pattern is perceived as a uniform grey rather than an alternating pattern unless the contrast is extremely high. At the Autokon resolution of 722 lpi, a one pel run is beyond visual resolution. By skipping these short runs, the total number and the distribution of runs will be greatly improved for run-length encoding. This scheme is especially useful when there is not much detail in a picture. In that case, skipping errors with short length creates almost no discernible difference.

Skipping runs with length of one and two has been tried. The compression ratio is improved, however, the quality of the picture is degraded and the degradation depending on the original halftone. When only runs of length 1 are omitted, results on the the IEEE facsimile test chart shows that only a slight difference can be detected in the high contrast resolution chart area. In the CROWD picture, if examined carefully, some loss of detail in the bark of the tree can be seen. As for the CMAN picture, even when runs with length of 2 are omitted, it shows little discernible difference. Overall, the omission of short runs with 1 pel length is shown to have only a very small effect on the visual quality but greatly improves the coding efficiency. For less complex pictures, the omission of runs with length of 2 in the difference picture can further improve the coding efficiency without further impairing the picture quality. All pictures show appreciable degradation when runs of length three are also skipped.

Other than the scanning direction redundancy, there is also redundancy in the line direction. The close correlations of pels in the vertical direction create the similarity between adjacent lines in the difference picture. Due to this similarity, skipping some lines before transmission and then recover them by repeating or interpolation will not cause too much degradation either, if appropriate processing has been carried out. This processing is discussed in detail in the next section.

6.2.4. Vertical Skipping of Lines

Due to the performance of the visual system, short horizontal differences can be neglected without causing noticeable degradation. The same principle can be applied in the vertical direction. One simple way is to skip every other line while transmitting the difference picture and then use an interpolator to recover the missing lines. If an appropriate interpolator is used, the quality of the resulting picture can be very close to that of the original. The interpolator used in this thesis is not very complex. In a fine detail difference picture, the number of white pels is substantially smaller than the black ones because of the correlation in the reconstructed halftone and the originals. The interpolation should therefore emphasize the white pels. Basically, it interpolates only when the two lines have some white pels (a white pel indicates there is need to correct the low channel pel) overlapped with each other as shown in Figure 6.6. Figure 6.6a and Figure 6.6b indicate the effect of interpolation while Figure 6.6c shows that no interpolation will be performed.

Two steps are followed in interpolating the current line from its two neighboring lines. First, an AND operation is performed on the two lines to create the skeleton of the interpolated line. For each white run, chances are that interpolation is required at the two ends. Next, a filling routine is called to actually perform the interpolation. End points of the current run are obtained from runs of the previous as well as next line. All pels between the left and right end point are then assigned to be white. Experiments

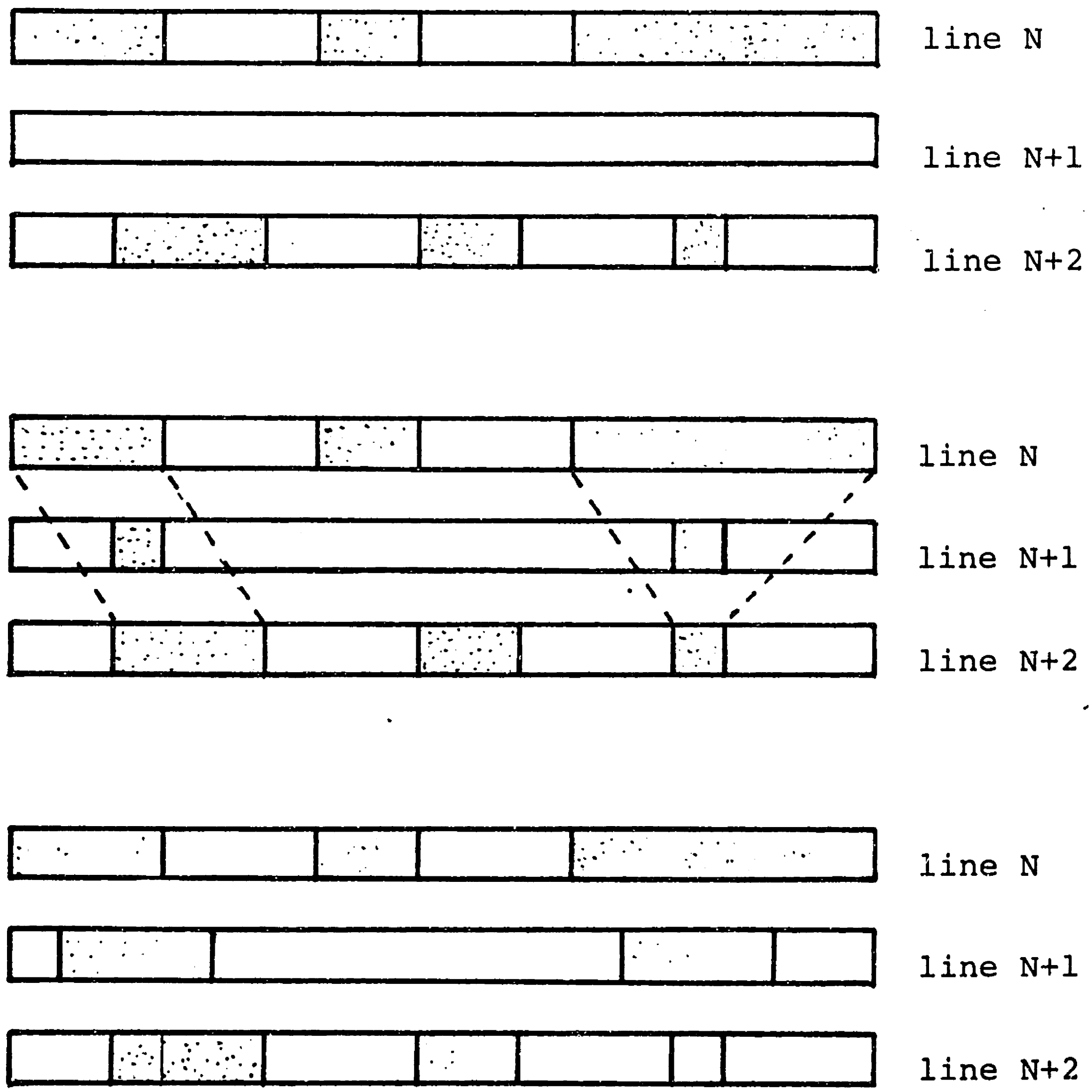


Figure 6.6 Interpolation of binary lines.

prove that this method works extremely well with halftone pictures. By using this line interpolation, a compression of 2 is achieved in addition to any other compression scheme. Theoretically, the same technique can be applied when more lines are skipped.

6.2.5. Result of 2 Channel Halftone Encoding

The advantages of this system are:

(1) We can trade off picture quality against transmission speed and storage space etc.

(2) Existing coding schemes can be used to encode the low detailed picture, such as DPCM, transform coding, dither coding etc..

The above mentioned technique has been applied to three pictures: IEEE test chart (IEEE), crowd picture (CROWD) and camera man picture (CMAN). In both IEEE and CMAN, subsamplings are accomplished over a 14X14 area while in CROWD 10X10 is used. The encoded highs channel uses CCITT run-length code and skipping single pel runs as well as every other line. Output picture is kept at the same line width as the picture to be encoded. Total number of output picture lines are then used to calculate the compression ration. The IEEE is compressed from 2912 lines to 289 lines. For CMAN, it is 188 lines from 4096 lines original and for CROWD it is 221 from 4096. Since no compression is done on the lows channel, it takes 7 bits per converted contone pel for a 65 line halftone. The overall compression ratios with respect to the original halftone scanned from the Autokon and store on disk are calculated as below:

for IEEE

$$\frac{2912}{288} = 10.1 \qquad \text{for highs channel}$$

$$\frac{14x14}{7} = 28 \qquad \text{for lows channel}$$

$$CR = \frac{1}{(1/10.1)+(1/28)} = 7.42$$

for CROWD

$$\frac{4096}{221} = 18.5 \qquad \text{for highs channel}$$

$$\frac{10x10}{7} = 14.3 \qquad \text{for lows channel}$$

$$CR = \frac{1}{(1/14.3)+(1/18.5)} = 8.07$$

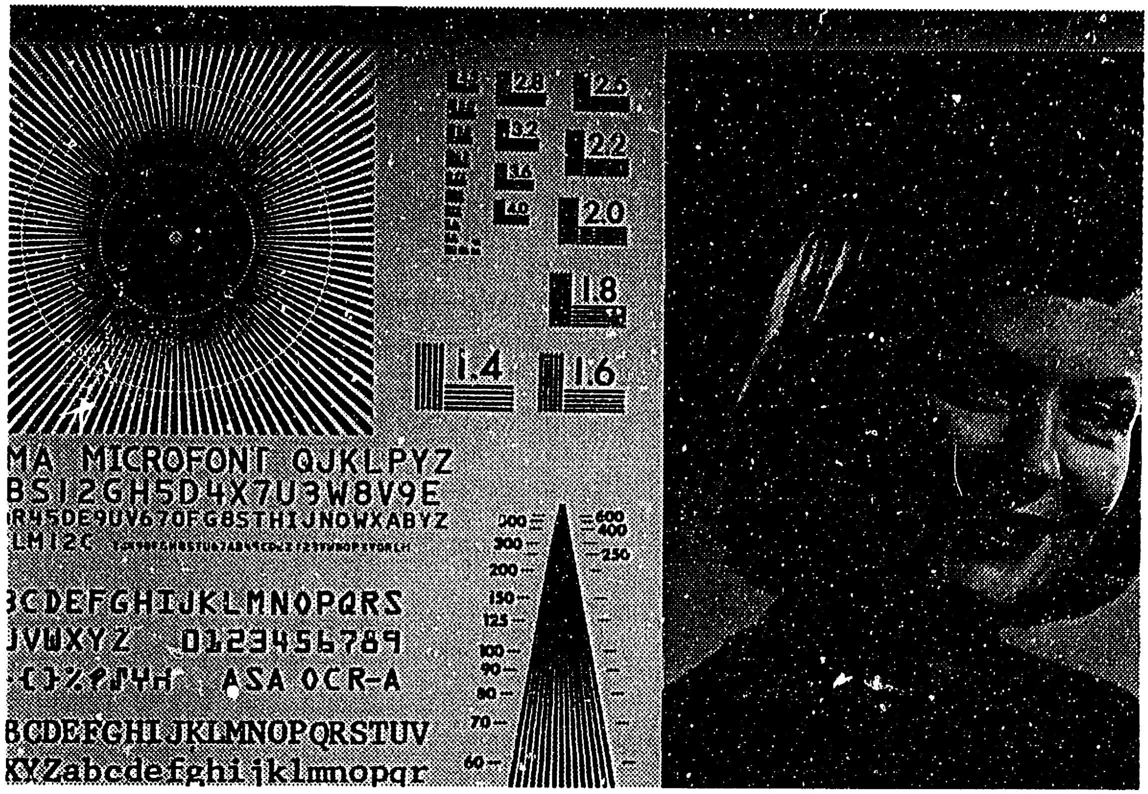
for CMAN

$$\frac{4096}{188} = 21.8 \qquad \text{for highs channel}$$

$$\frac{14x14}{7} = 28 \qquad \text{for lows channel}$$

$$CR = \frac{1}{(1/21.8)+(1/28)} = 12.3$$

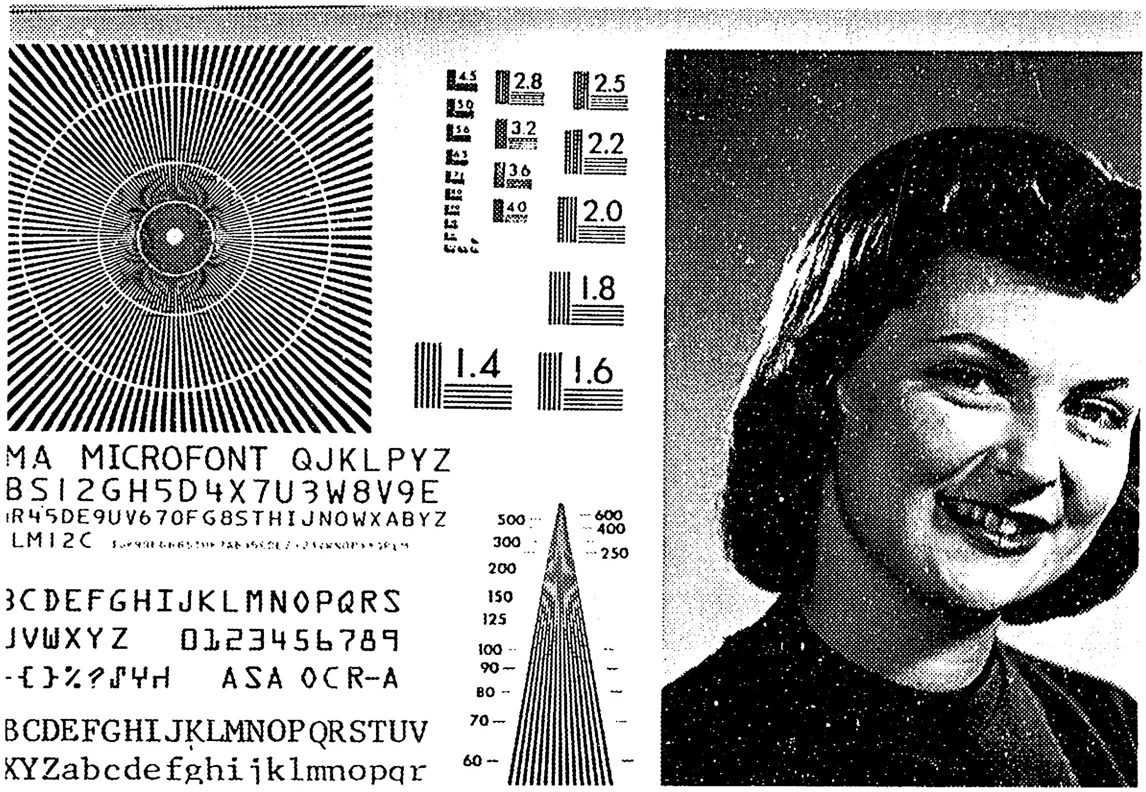
The original and decoded halftones are shown in Figure 6.7 through 6.12. When we have complete knowledge of the screen structure, this coding method is shown to be very efficient.



(a)

(b)

Figure 6.7 IEEE test picture. (a) original, (b) Reconstructed



(a)

(b)

Figure 6.7 IEEE test picture. (a) original, (b) Reconstructed

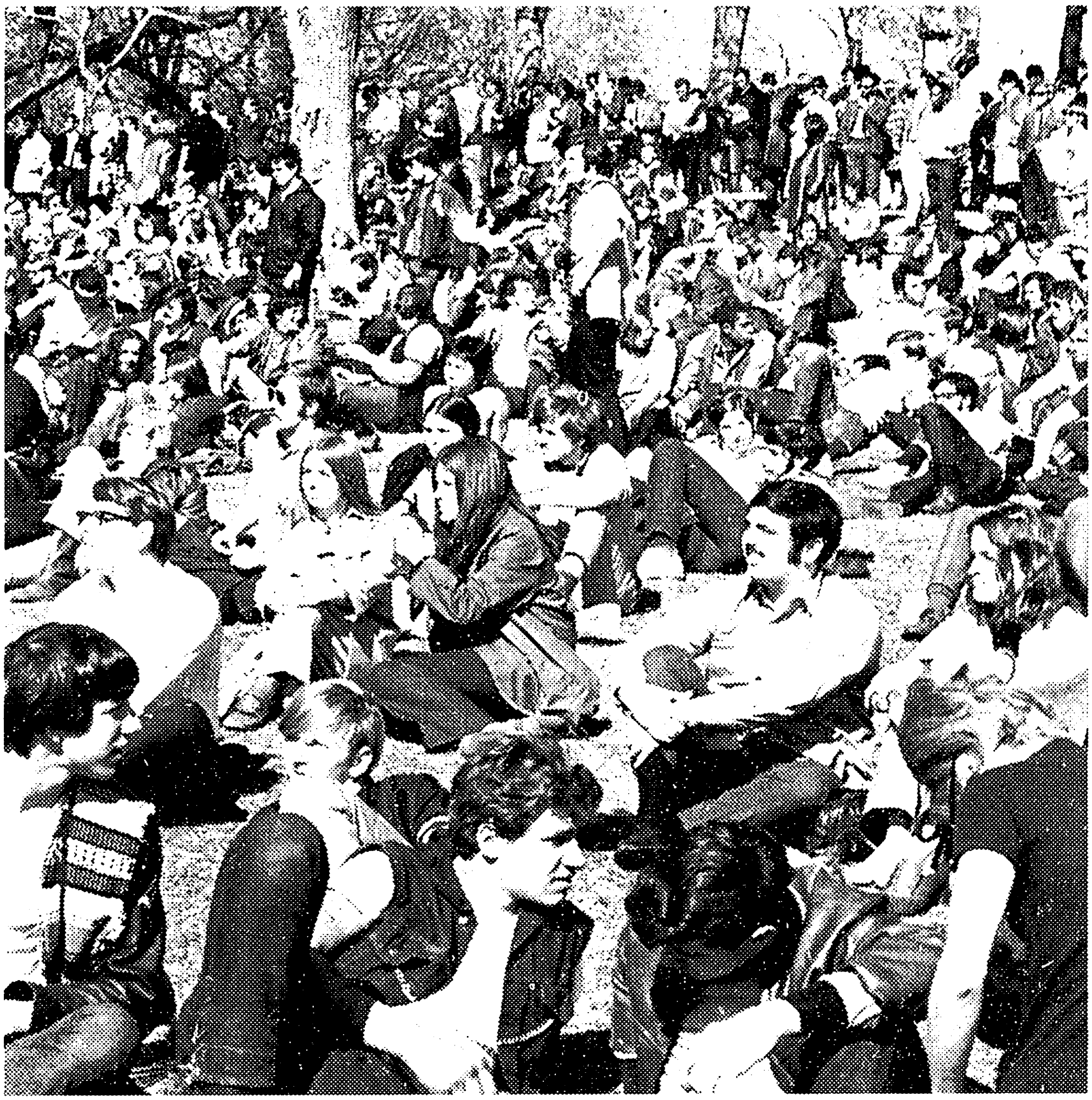


Figure 6.8 Original CROWD test picture.

Figure 6.8 Original CROWD test picture.

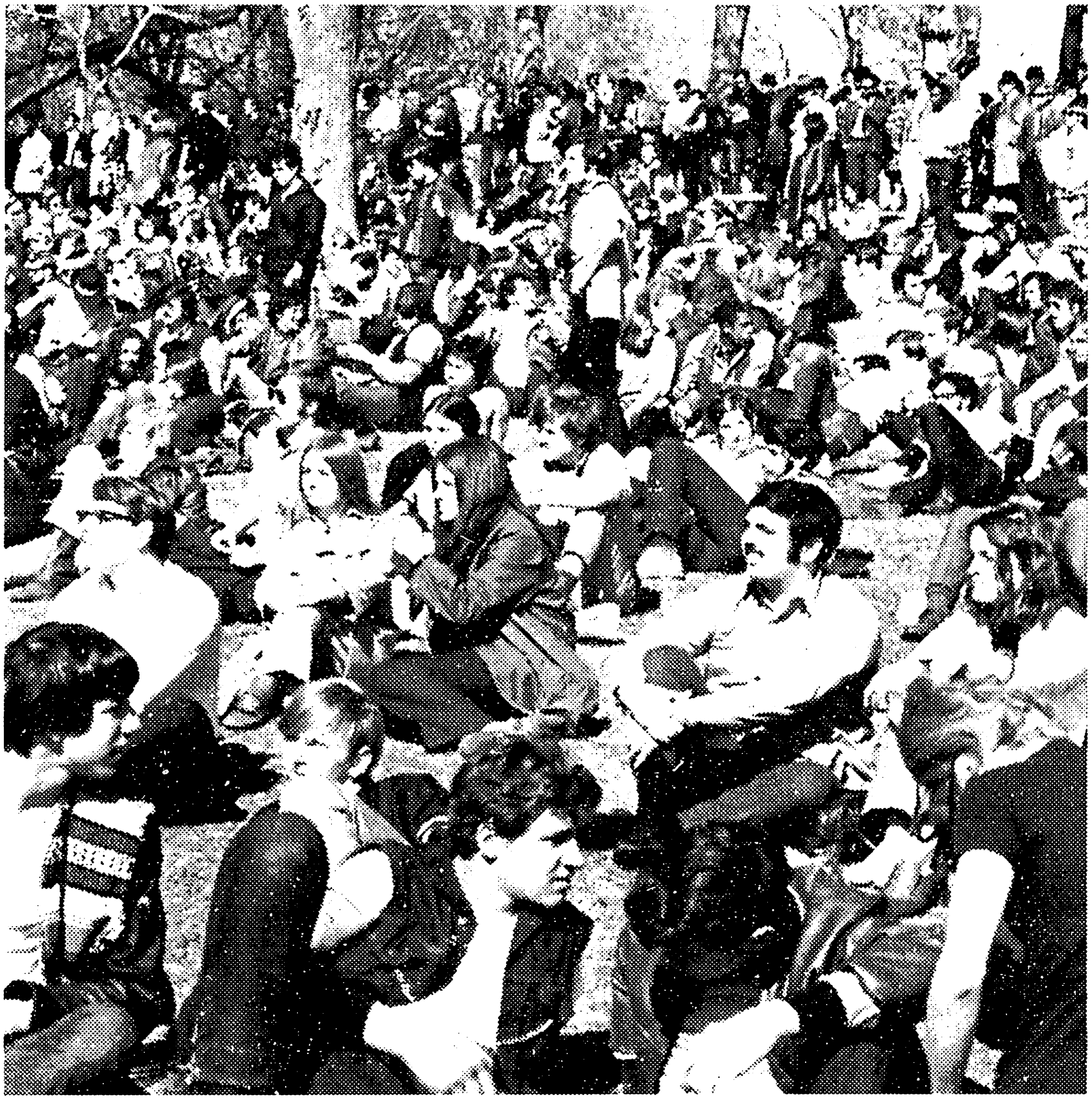


Figure 6.9 Reconstructed CROWD test picture.

Figure 6.9 Reconstructed CROWD test picture.

Figure 6.10 Original CMAN test picture.

Figure 6.10 Original CMAN test picture.

Figure 6.11 Reconstructed CMAN test picture.

Figure 6.11 Reconstructed CMAN test picture.

CHAPTER 7

Summary and Conclusions

7.1. Summary

At the beginning of the project, a number of goals were set for the work to be carried out. Briefly, they were:

(1) To study the nature of halftone pictures especially from the spatial frequency domain point of view and using a generalized dot structure.

(2) To evaluate screen patterns other than the traditional rectangular type.

(3) To convert halftones to contones for purposes of rescreening, enlarging, coding etc.

(4) To find limitations of a halftone process in rendition of details.

(5) To find a good coding scheme for storage and transmission of halftone pictures.

All of these objectives have been met satisfactorily.

On the screen structure, three hexagonal screens were implemented. These hexagonal halftones produced very good visual quality. The mathematical equations for rectangular halftone pictures were generalized to screens with general structure, spacing and orientation.

In order to study the conversion from half tone images to contone images, a number of algorithms were developed. The desired goal for these algorithms was a Moire-free picture. Based on this criteria, the results of the processing of halftone images using these algorithms were compared. It was found that among all of them the "dot size averaging" algorithm produces the best results. This algorithm carries out its work in a number of steps, the first one being the determination of the dot structure of the image. Algorithms were also implemented to do that for halftone images which dot structure was completely unknown. These were quite successful. Based on the determined dot structure an averaging mask is created which is used to process the given image to get rid of the dot structure and retrieve a contone image. The dot size averaging algorithm was also augmented with an adaptive detail extraction method to deal with pictures which originally had lot of high contrast detail but had screened by means of a coarse screen. The detail extractor works by determining the texture of a small area in the halftone picture. In areas where the texture is complex, a direct scan with a small scanning aperture is simulated. In all other areas the ordinary dot size averaging process is used.

On the limitation of halftone process, experiments were conducted to find the sampling density required to reproduce halftones with quality comparable to that obtainable from directly screening the originals. The required density is relatively low compared to what is required to scan a halftone using one bit per pel format. Depending on the characteristics of pictures and the halftone screen density used, a sampling density of 180 samples per inch is found to be adequate for almost all pictures using halftone screen up to 100 dots per inch.

On the coding of halftone pictures, two approaches were investigated. The two-channel coding scheme is used for coding-before-screening approach and a 180 lpi sampling density that is found adequate for full quality halftone reproduction. Compression ratio between 4 and 7 can be achieved as compared to the one bit per pel halftone format for pictures directly scanned from the Autokon. For coding-after-screening approach, two methods are implemented. One uses a multi-pel predictor to improve the statistics of the resulting binary pictures. The other decomposes a halftone into lows and highs channel and skips short runs to get rid of the signals that have little effect on the quality of the pictures. Compression ratios as high as 12 to 1 were achieved, again, with respect to the halftones directly scanned from the Autokon.

7.2. Conclusions

Up till now, not much work has been done on image processing with halftone pictures. The main reason is that the two properties associated with a halftone picture, the bi-level characteristic and the semi-periodic structure, make it extremely difficult to process without generating Moire patterns. In this thesis, conversion methods were implemented to convert a halftone into a contone while keeping as much information as possible. Pictures so obtained can be processed by the ordinary image processing methods such as scaling, tone scale transformation, or even for the screen density conversion. This should open the door for image processing with halftone pictures.

About the coding of halftone pictures, advantages and disadvantages for these two approaches are summarized in the following. Approach (a) is the coding-before-screening approach which was not using screen information while a two-channel coding scheme is used to code the contone pictures. Approach (b) is the coding-after-screening approach. In this case information about the screen patterns has been used as a kind of predictor to reconstruct the halftones. One dimensional runlength code by the CCITT group 3 method is used for detail transmission.

(1) Flexibility: Approach (a) has the advantage in this respect. Since the contone picture is coded and transmitted, conversion to halftones of different line density or screen type is very easy. It can be done by simply changing a screen function.

(2) Picture quality: For pictures with a small amount of fine detail, both approaches produce the same quality using the comparable number of bits. When the picture is rich in high contrast detail such as the IEEE test pattern, approach (a) can not reproduce the fine detail unless the contone is sampled at a relatively dense grid. But by doing so, more data is needed and the compression ratio will be degraded.

(3) Efficiency: In approach (a), the efficiency depends on the sampling density. If there are high contrast details in the original contone, then it needs high sampling density to digitize, resulting in lower compression ratio. In this case, approach (b) is a better choice. On the other hand, when the original consists mainly of plain texture, or only a low quality halftone is required, then using approach (a) with relatively low sampling density is better than using (a) and skipping longer runs. When the picture consists mainly of low spatial frequency, there is less information in the

texture. Use of more bits to represent the grey levels is a better trade off. When very high contrast details are present, then texture becomes more important and the code after screen approach starts showing an advantage either in the picture quality at the same coding efficiency or better efficiency with the same picture quality.

(4) Complexity of implementation: Approach (a) needs a low pass filter, a nonlinear quantizer, a pseudorandom noise generator etc. Approach (b) needs a screen structure for each halftone screen, one line of buffer if skipping of short runs and skipping of lines are both used, and a runlength encoder. Both approaches are quite simple compared with coding schemes such as transform coding. No complex computations are required especially in the case of approach (b). Both should be suitable for real time operation due to the simplicity of implementation in hardware.

(5) Effects of errors. Approach (a) is less prone to transmission errors because error are restricted to a local area. When an error occurs in the lows channel, it affects the average grey level over the subsampled area. If the error is in the highs channel, it affects only one pel. Overall, the error is localized to either a single

pel or the interpolated area for reconstructed lows signal. On the other hand, approach (b) uses runlength encoding. If there is an error in a line, the error propagates to the pels in the same line and renders the line unsuitable for image reconstruction. However, error correction codes can be used to decrease the net error rate. If an error does occur, error concealing algorithms can be applied to minimize the effect.

As for the selection of which approach to be used in transmitting and storing halftone pictures, it depends on the quality of halftones needed and the nature of the pictures to be processed. For pictures of simple texture or when the full quality is not required, a low sampling density coupled with an effective contone coding scheme is a better choice. On the other hand, when the picture has large amounts of high contrast detail and the full quality is to be reproduced in the halftone process, then coding-after-screening approach is clearly better.

7.3. Suggestions for Further Research

On the binary representation of contone pictures, it is found that the error diffusion method produces the best quality so far as the fine detail rendition and Moire pattern are concerned. There is one shortcoming in that algorithm, namely, the artifacts in the plain area when grey level is near the highlight or the shadow. One potential solution to get rid of this problem is by randomizing the coefficients for the error to be diffused to the neighboring pels. Another method is to combine the two schemes: contone screen and error diffusion. A contone screen produces less visible patterns under the same circumstances. It is logical to suggest that a combination of the two methods should produce a better quality.

So far, the contone to halftone conversion is accomplished by fixed screen. An adaptive scheme for this conversion is an interesting topic. A good screen pattern should produce noiseless dots in plain areas while generating a sharp texture for detailed areas. A fixed screen pattern usually is optimized only to one of the above objectives. Thus a screen pattern that produces fine details enhances the noise as well [75]. If the screen pattern can be adaptive to the local texture of the picture, then both noise suppression and detail enhencement can be achieved in the screening process. The result would be a substantially better halftone than any currently available screen. A fixed screen can be used to do the first pass conversion. The algorithm used in chapter three for detail extractor can then be applied. A map indicating details in the picture can be obtained. Using this detail map, screens with detail enhancement can then be applied in those areas rich in detail.

The same adaptive principle can be applied for encoding halftone pictures. The coding scheme in chapter six that skips short runs creates some loss of detail even when only the very short runs are omitted. This is especially true when large number of short runs are clustered and all of those short runs are skipped. A better way to improve the picture quality is to divide the picture into small areas and with the length of the short runs that are to be skipped determined by the complexity of the picture at each local area. If fine detail prevails, then more bits are transmitted for the shape of the dots. That means skipping only very short runs. In the extreme case, maybe no skipping is allowed in order to render the details. On the other hand, if almost no detail in the area can be detected, then runs with much longer length can be skipped without degrading the picture quality. This adaptation of skipping short runs could improve either the quality or coding efficiency.

The contrast versus resolution test has been experimented yet the result is not utilized in the coding-before-screening approach. Picture sampled at the appropriate density can be sent to a preprocessor where the low contrast detail is filtered. This picture now has less information content. Without the unreproducible detail signals in the image, a better coding efficiency can be achieved.

In the process of picture transmission, one may not want to obtain the full quality from the transmitting side. For example, if we know the newspaper halftone process will be substantially degraded in the printing, then we really only need to get pictures that presumably will produce halftones equivalent to what can be obtained by the best original under the same constraint. We know that very often a large amount of information must be transmitted for a slight improvement in picture quality. If quality can be compromised, then a better algorithm can be applied to achieve a higher compression ratio. On the other hand, some receiving sites may require higher quality halftones. To satisfy both situations, a progressive scheme is quite desirable. In this scheme, the picture is sent in a progressive fashion. A "skeleton" of the pictorial information is sent first. Extra information is then transmitted to further improve the quality such as adding more detail, more exact grey levels etc. The amount of information needed can be determined by the receiving site. Transmission can be stopped at the time when further information does not improve the final reproduction of the halftones being transmitted. If high quality is not needed in the reproduction, both time and storage space can be saved by simply sending as little information as necessary for the rendition of the desired quality. When better or full quality is needed, extra information can then be sent at the extra cost of transmission and computation time and storage space.

The mask designed for this work had very sharp transition in space domain. Namely, a value of one is assigned inside the mask, and zero outside. From the information about the screen pattern, it is possible to design a filter with the similar zero-crossing at the harmonics of the screen pattern which is desirable for the Moire pattern elimination. For example, the alternating phase inversion in between the harmonics of the screen pattern can be corrected to be in phase for the entire spatial frequencies. The characteristics of the filter can also be designed to preserve more signal components by having sharper notches at the harmonics of the screen pattern. Theoretically, this can be done by appropriately assigning the coefficients to the mask instead of uniformly assigning it to a constant. The implementation will be more complex. How to design such filters, and how much improvement from such implementations is possible, are very interesting topics for future research.

APPENDIX

A. Fourier Transform in Generalized Coordinates

The application of reciprocal basis and multi-dimensional Fourier transform has been developed in crystallography for a long time. Multi-dimensional sampled space in the frequency domain was investigated by Petersen et al [69], where the authors proved that the most efficient sampling grid is the hexagonal grid so far as the highest spatial frequency components are concerned. A further investigation of the two-dimension hexagonal raster has been published by Mersereau [58][59], where the representation of the hexagonal raster in terms of the ordinary orthogonal basis was used. The result is somewhat complex and not so straightforward as in the orthogonal raster. The fast Fourier transform (FFT) exists for the above representations [29][73]. It is slightly different from the FFT in orthogonal coordinates. A hexagonal sampling process is also described by Kelly [48], where the Fourier transform of impulses in a hexagonal raster is obtained by superposition and scaling of axes. The method presented below is to use the reciprocal basis to obtain the same result but in a more familiar and compact form.

B. Reciprocal Basis

An n-dimensional vector space needs n vectors as a basis to span the entire space. These n vectors can be arbitrarily chosen as long as they are linearly independent. If we choose $(\bar{r}_1, \bar{r}_2, \ldots . \bar{r}_n)$ as our coordinates, then for any vector in the space we have a unique representation.

Out of $(\bar{r}_1, \ldots . \bar{r}_n)$ there exists another set of vectors $(\bar{R}_1, \bar{R}_2, \ldots . \bar{R}_n)$ called the reciprocal basis that has the following property:

$$\bar{r}_i \cdot \bar{R}_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

where $\cdot$ indicate the inner product and

δ_{ij} is the Kronecker delta function.

C. Fourier Transform

The Fourier transform of a function in (r_i) coordinates can be written as

$$F(\bar{G}) = \int_{\text{entire space}} dv f(\bar{r}) \exp[-2\pi \bar{G} \cdot \bar{r}] \qquad (1)$$

Let us limit ourselves to the 2-dimensional case of interest. If we choose $\bar{r}_1$, $\bar{r}_2$ to be an orthonormal basis, then the reciprocal basis is nothing more than itself. We write $f(\bar{r})$ as $f(x,y)$, where $r = x\bar{r}_1 + y\bar{r}_2$

If we write $F(\bar{G}) = F(u,v)$ where $\bar{G} = u\bar{r}_1 + v\bar{r}_2$

we then have

$$F(u,v) = \iint_{-\infty}^{\infty} dxdyf(x,y)\exp[-2\pi j(ux\overline{r}_1\cdot\overline{r}_2+uy\overline{r}_1\cdot\overline{r}_2+vx\overline{r}_1\cdot\overline{r}_2+vy\overline{r}_1\cdot\overline{r}_2)]$$

$$= \iint_{-\infty}^{\infty} dxdyf(x,y)\exp[-2\pi j(ux+vy)] \quad (2)$$

It is easily seen that Equation (2) is the familiar form. Now if $\overline{r}_1$, $\overline{r}_2$ are arbitrary (but linearly independent), then we can find the reciprocal basis $\overline{R}_1$, $\overline{R}_2$ in the frequency domain. The spectrum can be written as

$$F(\overline{G}) = F(g_1,g_2)$$

$$= |\overline{r}_1 x \overline{r}_2| \iint_{-\infty}^{\infty} dx_1 dx_2 f(x_1,x2)\exp[-2\pi j(g_1x_1+g_2x_2)] \quad (3)$$

In the two-dimensional case, the area element is

$$ds = |\overline{r}_1 x \overline{r}_2| dx_1 dx_2 = |(\overline{r}_1 dx_1) x (\overline{r}_2 dx_2)|$$

Note that in Equation (3), except for the magnitude factor $|\overline{r}_1 x \overline{r}_2|$, all other computations are the same as Equation (2). Equation (3) has the new meaning that the Fourier transform is expressed in the reciprocal basis.

D. Examples

(a) Spectrum of a slanted bar as shown in Figure I. where the angle between the two basis vector, is θ. The function can be

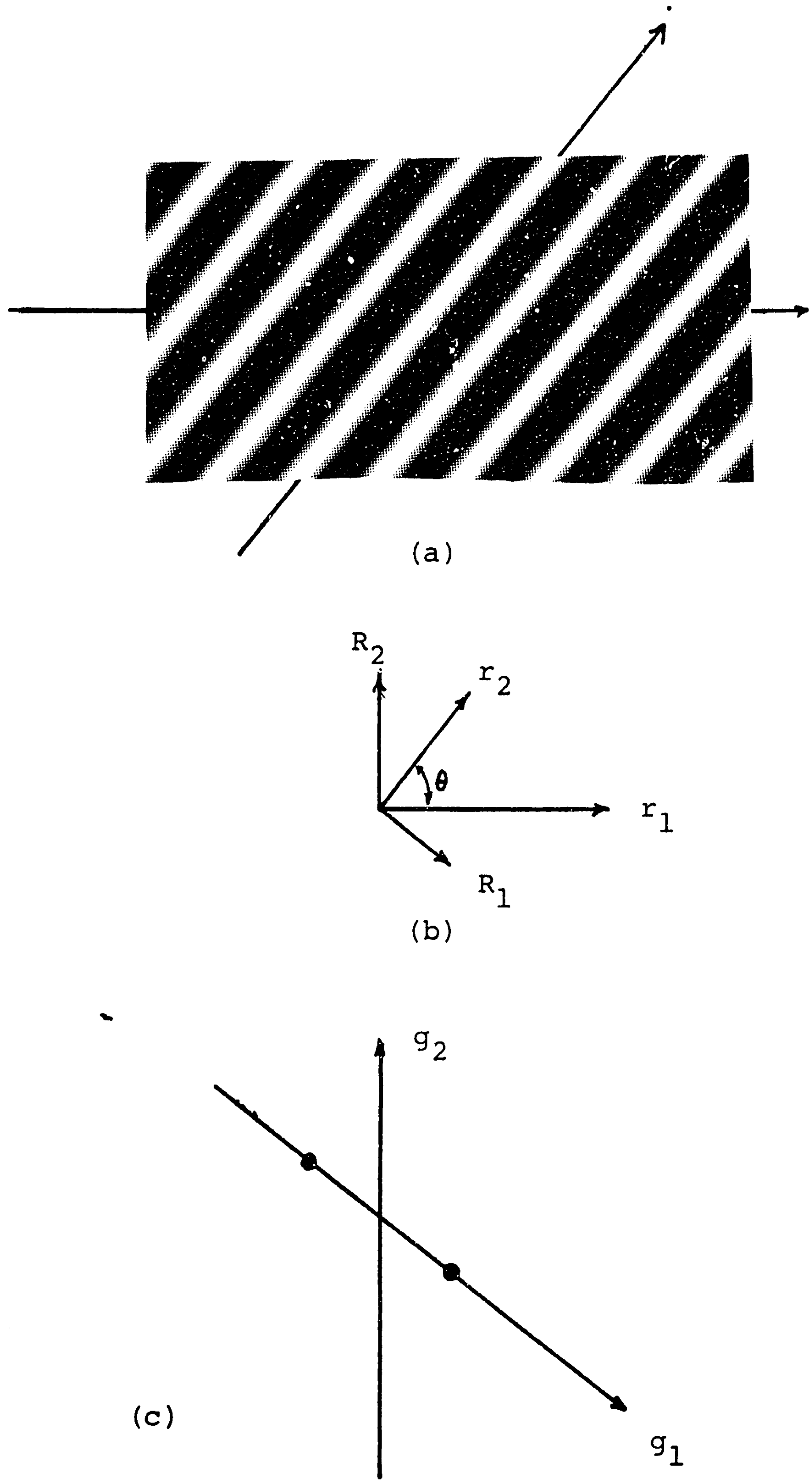


Figure I. Slant bar pattern. (a) is the function, (b) is the reciprocal coordinates, and (c) is its Fourier spectrum.

written as

$$f(x_1,x_2) = \cos 2\pi x_1$$

Now $|\bar{r}_1 \times \bar{r}_2| = |\bar{r}_1||\bar{r}_2|\sin\theta = \sin\theta$

The reciprocal basis is shown in Figure II.b as $\bar{R}_1$ and $\bar{R}_2$. Substituting into Equation (3), we have

$$F(g_1,g_2) = \sin\theta\iint_{-\infty}^{\infty} dx_1 dx_2 \cos 2\pi x_1 \exp[-2\pi j(g_1x_1+g_2x_2)]$$

Note that $|\bar{R}_1| = |\bar{R}_2| = \frac{1}{\sin\theta} = \csc\theta$

so

$$F(g_1,g_2) = \sin\theta\int_{-\infty}^{\infty} dx_2 \exp[-2\pi j g_2 x_2]\int_{-\infty}^{\infty} dx_1 \cos 2\pi x_1 \exp[-2\pi j g_1 x_1]$$

$$= \sin\theta\ \delta(g_2)\{\ \delta(g_1+1)+\ \delta(g_1-1)\}/2 \qquad (4)$$

Equation (4) can be shown to represent an array of two dimensional impulses as in Figure I.c.

We would get the same result if we first obtained the Fourier transform of $\cos(2\pi\ \csc\theta)x$ in the ordinary orthonormal basis and then rotated the result by an angle of $\theta - \frac{\pi}{2}$.

Note also that in Figure II, distances from the two impulses to the origin are $\frac{1}{\sin\theta}$ instead of 1.

(b) Two dimensional hexagonal sampling pattern.

A step-by-step approach to find the Fourier transform pair of the hexagonal sampling pattern has been discussed by Kelly [48]. Now armed with the technique of reciprocal basis, we can attack the same problem in a single step.

The 2-D infinite hexagonal sampling function is shown in Figure II.a, where each point represents a 2-D sampling impulse. Written as Dirac delta function, each sampling impulse can be represented by

$$\delta(x_1-m)\delta(x_2-n)$$

Now the hexagonal sampling function can be written as

$$\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\delta(x_1-m)\delta(x_2-n) \tag{5}$$

It is easily seen that the reciprocal basis is as shown in Figure II.b. The angle spanned between the basis and the reciprocal basis of the hexagonal sampling grid is $\pi/3$ and $2\pi/3$ respectively.

The Fourier transform represented in the reciprocal basis can be obtained from Equation (3).

$$F(g_1,g_2) = \sin\frac{\pi}{3}\iint_{-\infty}^{\infty} dx_1 dx_2 \sum_m\sum_n \delta(x_1-m)(x_2-n)\exp[-2\pi j(x_1g_1+x_2g_2)]$$

$$= \sin\frac{\pi}{3}\int_{-\infty}^{\infty} dx_2 \exp[-2\pi j x_2 g_2] \sum_n \delta(x_2-n)$$

$$\cdot\int_{-\infty}^{\infty} dx_1 \exp[-2\pi j x_1 g_1] \sum_m \delta(x_1-m)$$

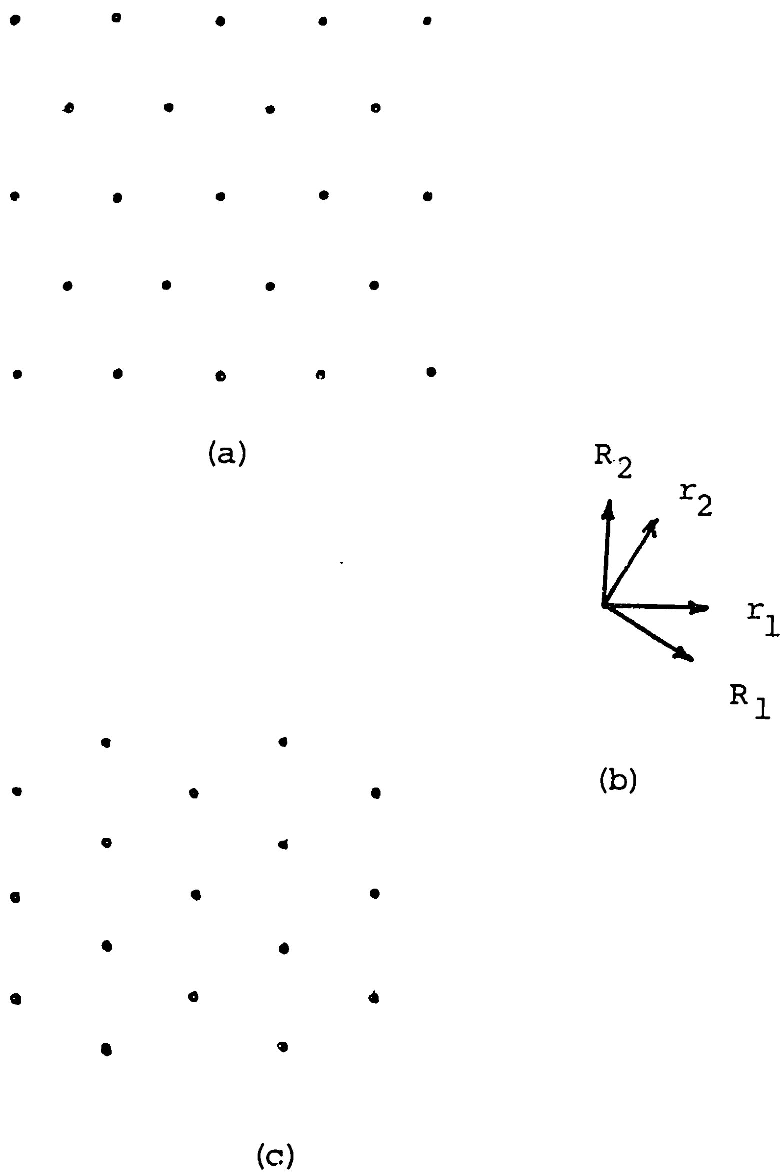


Figure II. Hexagonal sampling grid.

$$= \sin\frac{\pi}{3} \sum_k \delta(g_2-k) \sum_l \delta(g_1-l)$$

$$= \frac{\sqrt{3}}{2} \sum_k \sum_l \delta(g_1-l)\ \delta(g_2-k) \tag{6}$$

Equation (5) and (6) are quite similar. The hexagonal sampling function has hexagonally spaced impulses but the direction has been rotated by $\frac{\pi}{3}$.

If we expand the axis of Figure II.a such that the magnitude

$$|\bar{r}_1'| = |\bar{r}_2'| = \left(\frac{2}{\sqrt{3}}\right)^{1/2}$$

and directions are the same as $\bar{r}_1$, $\bar{r}_2$ respectively, then their reciprocal basis $\bar{R}_1'$, $\bar{R}_2'$ has the properties

$$\bar{r}_1' \cdot \bar{R}_1' = 1, \qquad \text{and} \qquad \bar{r}_2' \cdot \bar{R}_2' = 1.$$

We have

$$|\bar{R}_1'|\,|\bar{r}_2'|\ \frac{\sqrt{3}}{2} = 1$$

$$|\bar{R}_1'| = \left(\frac{2}{\sqrt{3}}\right) / \left(\frac{2}{\sqrt{3}}\right)^{1/2} = \left(\frac{2}{\sqrt{3}}\right)^{1/2} = |\bar{r}_1'|$$

Similarly

$$|\bar{R}_2'| = |\bar{r}_2'|$$

The sampling function $\sum_m\sum_n \delta(x_1'-l)\ \delta(x_2'-n)$ then transforms to

$$F(g_1',g_2') = \left(\frac{2}{\sqrt{3}}\right)^{1/2}\left(\frac{2}{\sqrt{3}}\right)^{1/2}\sin\frac{\pi}{3}$$

$$\cdot\iint_{-\infty}^{\infty} dx_1'dx_2' \sum_{mn}\sum \delta(x_1'-m)\ \delta(x_2'-n)\exp[-2\pi j(x_1'g_1'+x_2'g_2')]$$

$$= \sum_k \sum_l \delta(g_1'-k)\ \delta(g_2'-l)$$

which is the same as developed by Kelly.

(c) The Fourier transform of a function that is unity inside a hexagon and zero elsewhere. The function is shown in Figure III.

We could accomplish this in the usual way, using the orthonormal basis. If instead, we represent the function in the hexagonal basis $\bar{r}_1,\bar{r}_2$, with

$$|\bar{r}_1| = |\bar{r}_2| = a, \quad \text{and} \quad \bar{r}_1 \cdot \bar{r}_2 = \cos\frac{\pi}{3} = 1/2$$

then we can represent its Fourier transform in the reciprocal basis ($\bar{R}_1$, $\bar{R}_2$) using Equation (3)

$F(g_1,g_2)$

$$= \frac{\sqrt{3}}{2}a^2\int_{-1}^{0} dx_2 \int_{-(1+x_2)}^{1} dx_1 \exp[-2\pi j(x_1g_1+x_2g_2)]$$

$$+ \frac{\sqrt{3}}{2}a^2\int_{0}^{1} dx_2 \int_{-1}^{1-x_2} dx_1 \exp[-2\pi j(x_1g_1+x_2g_2)]$$

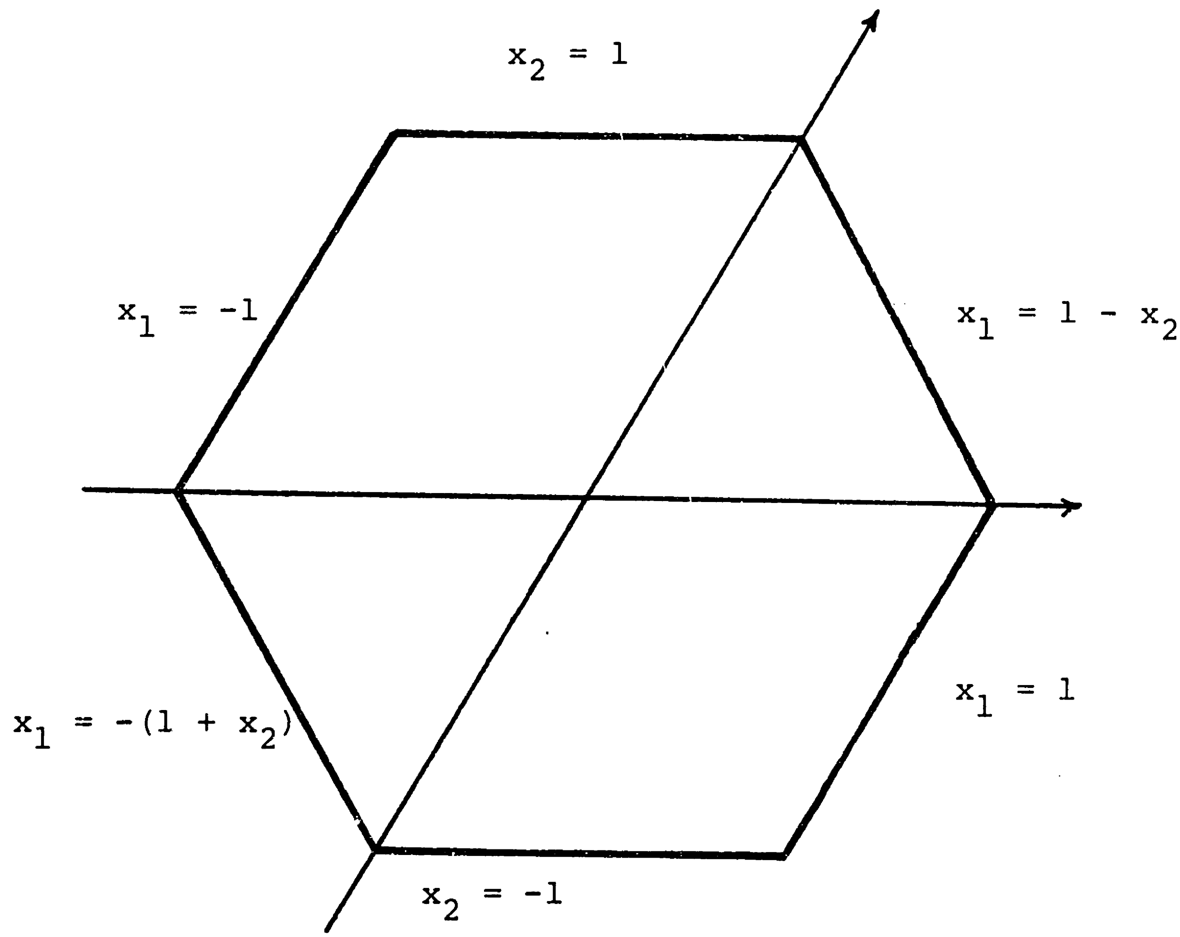


Figure III. A hexagon which Fourier spectrum is to be found.

$$= \frac{\sqrt{3}}{2}a^2\{\int_{-1}^{0} dx_2 \exp[-2\pi j x_2 g_2] \frac{\exp[-2\pi j x_1 g_1]}{-2\pi j g_1}\Bigg|_{-(1+x_2)}^{1}$$

$$+ \int_{0}^{1} dx_2 \exp[-2\pi j x_2 g_2] \frac{\exp[-2\pi j x_1 g_1]}{-2\pi j g_1}\Bigg|_{-1}^{1-x_2} \}$$

$$= \frac{\sqrt{3}}{2}a^2\{\int_{-1}^{0} dx_2 \exp[-2\pi j x_2 g_2] \frac{\exp[-2\pi j g_2]-\exp[2\pi j g_1(1+x_2)]}{-2\pi j g_1}$$

$$+ \int_{0}^{1} dx_2 \exp[-2\pi j x_2 g_2] \frac{\exp[-2\pi j g_1(1-x_2)]-\exp[2\pi j g_2]}{-2\pi j g_1} \}$$

$$= \frac{\sqrt{3}}{2}a^2\{\frac{\exp[-2\pi j g_1]}{-2\pi j g_1}\cdot\frac{\exp[-2\pi j x_2 g_2]}{-2\pi j g_2}\Bigg|_{-1}^{0}$$

$$- \frac{\exp[2\pi j g_1]}{-2\pi j g_1}\cdot\frac{\exp[2\pi j(g_1-g2)x_2]}{-2\pi j(g_2-g_1)}\Bigg|_{-1}^{0}$$

$$+ \frac{\exp[-2\pi j g_1]}{-2\pi j g_1}\cdot\frac{\exp[-2\pi j(g_2-g_1)x_2]}{-2\pi j(g_2-g_1)}\Bigg|_{0}^{1}$$

$$- \frac{\exp[2\pi g_1]}{-2\pi j g_1}\cdot\frac{\exp[-2\pi j x_2 g_2]}{-2\pi j g_2}\Bigg|_{0}^{1} \}$$

$$= \frac{\sqrt{3}}{2}a^2\{ \frac{\exp[-2\pi j g_1](1-\exp[2\pi j g_2])}{-4\pi^2 g_1 g_2} - \frac{\exp[2\pi j g_1](1-\exp[2\pi j(g_2-g_1)])}{-4\pi^2 g_1(g_2-g_1)}$$

$$+ \frac{\exp[-2\pi j g_1](\exp[-2\pi j(g_2-g_1)]-1)}{-4\pi^2 g_1(g_2-g_1)} - \frac{\exp[2\pi j g_1](\exp[-2\pi j g_2]-1)}{-4\pi^2 g_1 g_2} \}$$

$$= \frac{\sqrt{3}}{2}a^2\{\frac{(\exp[-2\pi jg_1]+\exp[2\pi jg_1])-\exp[2\pi j(g2-g_1)]-\exp[-2\pi j(g_2-g_1)]}{-4\pi^2 g_1 g_2}$$

$$+ \frac{-(\exp[-2\pi jg_1]+\exp[2\pi jg_1])+\exp[-2\pi jg_2]+\exp[2\pi jg_2]}{-4\pi^2 g_1(g_2-g_1)}\}$$

$$= \frac{\sqrt{3}}{2}a^2\{\frac{2\cos 2\pi g_1-2\cos 2\pi(g_2-g_1)}{-4\pi^2 g_1 g_2} + \frac{2\cos 2\pi g_2-2\cos 2\pi g_1}{-4\pi^2 g_1(g_2-g_1)}\}$$

$$= \frac{\sqrt{3}a^2\{\cos 2\pi g_1[(g_2-g_1)-g_2]-(g_2-g_1)\cos 2\pi(g_2-g_1)+g_2\cos 2\pi g_2\}}{-4\pi^2 g_1 g_2(g_2-g_1)}$$

$$= \frac{\sqrt{3}a^2}{4\pi^2 g_1 g_2(g_2-g_1)}\{g_2\cos 2\pi g_2-g_1\cos 2\pi g_1-(g_2-g_1)\cos 2\pi(g_2-g_1)\}$$

(d) Fourier spectrum of hexagonal halftone images.

The ordinary rectangular halftone screen has been discussed by Kermisch and Roetling [50]. Their method is to find the spectrum of a halftone picture with uniform grey level that reveals the structure of the screen. The same technique can be applied to hexagonal halftones with the difference that hexagonal coordinates are used instead of the familiar orthonormal coordinates. Assume a periodic pattern in the hexagonal system consisting of identical parallelograms with each side having length of T as shown in Figure IV. We then define the halftone function as

$$g(x,y;f) = \sum_{m}\sum_{n} g_0(s-mT,y-nT;f)$$

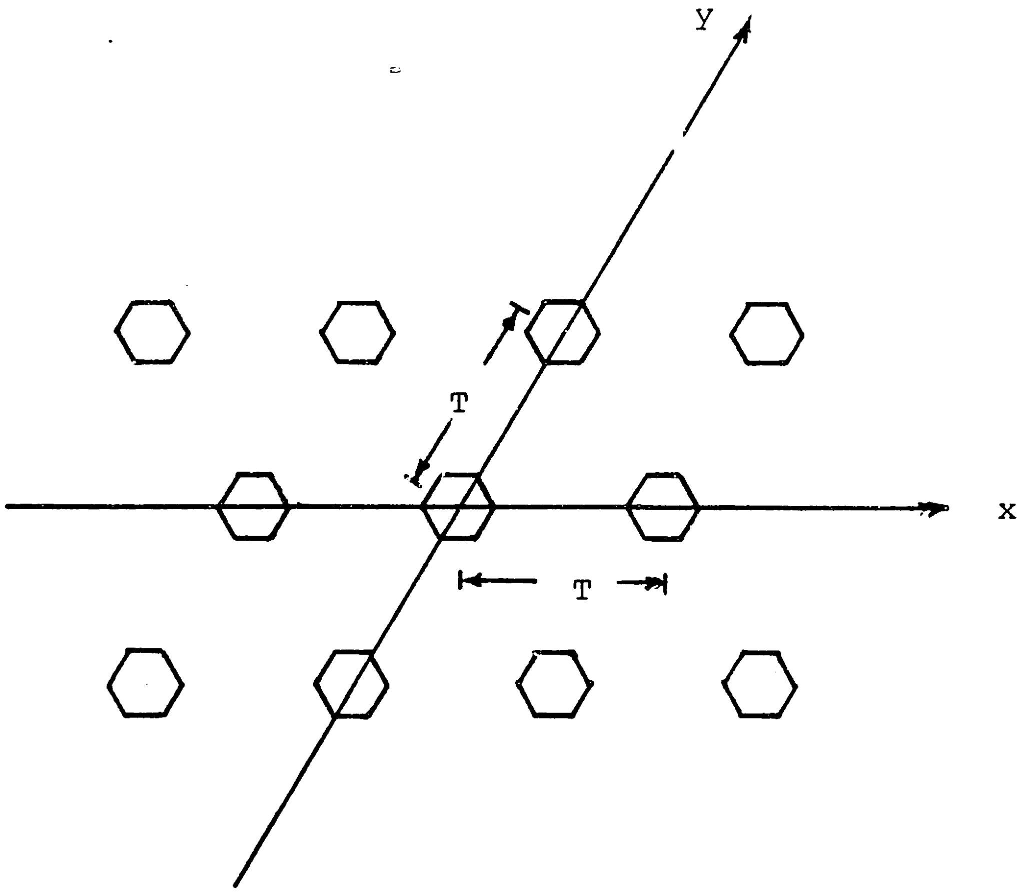


Figure IV. Shape of dot in the hexagonal screen discussed.

$$= \frac{2}{\sqrt{3}} g_0(x,y:f) ** \sum_{mn}\sum \delta(s-mT)\ \delta(y-nT)$$

$$\text{where } g_0(x,y;f) = \begin{cases} 0 \text{ or } 1 \text{ for } |x| < \frac{T}{2},\ \ |y| < \frac{T}{2}, \text{ depends on f} \\ 0 \text{ otherwise.} \end{cases}$$

Notation ** indicates the two-dimensional convolution.

Also

$$(\frac{1}{T_2}) \iint_{-\infty}^{\infty} g_0(x,y;f)dxdy = f$$

where f is a fixed grey level.

Note that directions of x,y are not orthogonal, rather they are separated by an angle of $\pi/3$. Now we denote $G(u,v;f)$ and $G_0(u,v;f)$ the 2-D Fourier transform of $g(x,y;f)$ and $g_0(x,y;f)$ respectively. It is easily seen that

$$G(u,v;f) = \frac{\sqrt{3}}{2} \iint_{-\infty}^{\infty} dxdyg(x,y;f)\exp[-2\pi j(ux+vy)]$$

$$= \frac{\sqrt{3}}{2} \iint_{-\infty}^{\infty} dxdy \iint_{-\infty}^{\infty} dx'dy'g(x-x',y-y';f)$$

$$\sum_{mn}\sum \delta(x'-mT,y'-nT)\exp[-2\pi j(ux+vy)]$$

$$= \frac{\sqrt{3}}{2} \iint_{-\infty}^{\infty} dxdy\ g(x-x',y-y';f)\exp[-2\pi j(x-x')u]\ \exp[-2\pi j(y-y')v]$$

$$\cdot \iint_{-\infty}^{\infty} dx'dy'\ \sum_{mn}\sum \delta(x'-mT,y'-nT)\exp[-2\pi j(x'u+y'v)]$$

$$= G(u,v;f)\left(\frac{1}{T^2}\right) \sum_{k}\sum_{l} \delta\left(u-\frac{k}{T}\right) \delta\left(v-\frac{l}{T}\right)$$

$$= u_0^2 \sum_{m}\sum_{n} G_0(mu_0,nu_0;f) \, \delta(u-mu_0) \, \delta(v-nu_0)$$

where $u_0 = 1/T$.

Now a 2-D halftone function $h(x,y)$ can be written as

$$h(x,y)=\iint_{-\infty}^{\infty} g[x,y;f(s,t)] \, \delta(x-s) \, \delta(y-t)\,ds\,dt$$

$$= \frac{2}{\sqrt{3}} \, g[x,y;f(s,t)] ** \sum_{m}\sum_{n} \delta(x-mT) \, \delta(y-nT)$$

We then have

$H(u,v)$

$$= \left(\frac{1}{T^2}\right) \sum_{m}\sum_{n} \iint_{-\infty}^{\infty} ds\,dt \; G_0[mu_0,nu_0;f(s,t)]\exp[-2\pi j[(mu_0-u)s+(nu_0-v)t]]$$

This is quite similar to the rectangular raster as developed by Kermisch, only the frequency is represented in the hexagonal reciprocal basis.

(e) Further applications

Mersereau has shown that hexagonal rasters are more efficient in the sense of frequency resolution for a fixed number of sampling points. He also developed the hexagonal DFT. Yet he still works in the rectangular coordinates. The representation is thus complex and some new techniques are required. For the hexagonal coordinates system as in the examples of the previous section, a hexagonal DFT can be easily obtained from the sample points if the reciprocal basis is used instead of the rectangular coordinates. All formulas for the ordinary rectangular DFT are then applied. All the mathematics is thus easier to understand and the form is also simpler. Existing computer programs are thus usable. The only differences are that (1) instead of a rectangular area, we now have to work with the parallelogram, (2) directions for the spatial frequency are different from the spatial sampling direction, rather it is in the directions of the reciprocal basis. It may be necessary to design new techniques of digital filtering because the frequency response is expressed in the reciprocal basis. It is believed that a simple translation between axes can solve this problem. All we have to do is to express the frequency response of a filter in the reciprocal basis, and then sample the frequency response in the reciprocal raster.

BIBLOGRAPHY

[1] Allebach, J. P. and Liu, B. (1976), "Random Quasiperiodic Halftone Process," Journal Opt. Soc. Am. vol. 66, Sept. pp. 909-917.

[2] Allebach, J. P. and Liu, B. (1977), "Analysis of Halftone dot Profile and Aliasing in the Discrete Binary Representation of Images," Jour. Opt. Soc. Am. vol. 67. no.9, Sept. pp. 1147-1154.

[3] Amidon, E. L. (1971), "Algorithm Selection of the Best Method for Compression Map Data String," Commun. ACM vol. 14, no. 2 pp. 769-774.

[4] Arnemann, S. and Tasto, M. (1973), "Generating Halftone Pictures on Graphic Computer Terminals Using Run-Length Coding," Computer Graphics and Image Processing, vol. 2, pp. 1-11.

[5] Autokon User's Manual, ECRM Burlington

[6] Bayer, B. E. (1973), "An Optimal Method for Two-Level Rendition of Continuous-Tone Pictures," IEEE ICC, vol. 1, pp. 26-11 to 26-16.

[7] Behane, D. (1971), "Method and System for Reconstruction of Halftone Images," US Patent 3604846 issued Sept. 1971.

[8] Buckley, R. (1981), "Digital Color Image Coding and the Geometryof Color Space," Ph. D. Dissertation, Dept. Elec. Eng. and Comp. Sci. MIT.

[9] Capon, J. (1959), "A Probabilistic Model for Run-Length Coding of Pictures," IRE Trans. on Info. Theory, vol. IT-5, pp. 157-163.

[10] Chow, L.K., Deekshatulu, B. L. and Loh, L. S. (1974), "Some Computer Experiments in Picture Processing for Data Compaction," Computer Graphics and Image Processing, vol. 3, pp. 203-214

[11] Cornsweet, T. N. (1970), "Visual Perception," Academic Press

[12] Davidson, M. L. (1966), "A Perturbation Analysis of Spatial Brightness Interaction in Flashed Visual Fields," Ph.D. Thesis, Univ. of California, Berkeley.

[13] Davidson, M. L. (1968), "Perturbation Approach to Spatial Brightness Interaction in Human Vision," Journal Opt. Soc. Am. vol. 58, Sept. pp. 1300-1309.

[14] Decoulon, F. and Johnson, O.,(1976) "Adaptive Block Scheme for Source Coding of Black and White Fascimile," Electronics Letters, vol. 12, No, 3 Feb. pp. 61-62.

[15] Dooley, R. P. (1975) "Predicting Brightness Appearance at Edges Using Linear and Non-Linear Visual Describing Functions," Presented at SPSE Annual Meeting, May 14, 1975, Denver Colorado.

[16] Ellis, J. R. (1977), "An Investigation of Approximate Bit-Plane Encoding of Pictorial Information," Ph. D. Dissertation, Dept. Elec. Eng. and Comp. Sci. MIT.

[17] Engeldrem, D. G. (1974), "A Mathematical Model of A Contact Screen System," TAGA pp. 1-31.

[18] Floyd, R. and Steinberg, L. (1975), "An Adaptive Algorithm for Spatial Grey Scale," SID Digest 1975, pp 30-31.

[19] Frank, A. J. (1973), "High Fidelity Encoding of Two Level , High Resolution Images," IEEE ICC 26-5 to 26-10.

[20] Frank, A. J. and Groff, R. H. (1976), "On Statistical Coding of Two-Tone Image Ensembles," Proc. SID, vol. 17, second quarter, pp. 102-110.

[21] Gallager, R. G. (1968), "Information Theory and Reliable Communication," John Wiley and Son.

[22] Golay, M. J. E. (1969), "Hexagonal Parallel Pattern Transformations," IEEE Trans. Comput. vol. C-18, pp. 733-740.

[23] Goren, R. N. and Marks, L. M. (1974), "Document Screen for Extended Range Electrophotography," Journal SPSE, vol. 18, no. 6, Nov. pp. 605-609.

[24] Grad, R. (1976), "Digital Picture Processing Techniques for Publishing Industry," Computer Graphics and Image Processing, vol. 5, pp. 151-171.

[25] Graham, D. N. (1966), "Image Transmission by Two-Dimensional Contour Coding," Ph.D. Dissertation, Elec. Eng. Dept. MIT.

[26] Grass, R (1978), "An Integreatd Image Compression/Enhancement System," M.S. Thesis, M.I.T., E.E.C.S. Dept. November.

[27] Habibi, A. (1971), "Comparison of nth-Order DPCM Encoder with Linear Transformations and Block Quantization Techniques," IEEE trans. on Commun. Tech. vol. COM-19, no. 6, Dec. pp. 948-956.

[28] Hallows, R. L. and Klensh R. J. Jr., (1968), "Electronic Halftones," IEEE Spectrum, Oct. pp. 64-72.

[29] Harris, D. B. McClellan, J. H. Chan, K. S. K. and Schuessler, H. W. (1977) "Vector Radix Fast Fourier Transform," 1977 IEEE Int. Conf. on ASSP Record, pp. 548-551.

[30] Huang, T. S. (1972), "Runlength Coding and its Extentions in Picture Bandwidth Compression," edited by T. S. Huang and O. J. Tretiak, Gordon and Breach.

[31] Huang, T. S. (1974), "Digital Transmission of Halftone Pictures," Computer Graphics and Image Processing, pp. 195-202

[32] Huang, T. S. (1975), "Bounds on the Bit Rate of Linear Runlength Codes," IEEE Trans. Info. Theo. vol. IT-21, Nov. pp. 707-708.

[33] Huang, T. S. (1975), "Easily Implementable Suboptimum Runlength Codes," Proc. ICC 1975 pp. 7-8 to 7-11.

[34] Huang, T. S. and Koller, H. U. (1975), "Coding of Multilevel Graphics," IEEE Trans. Commun. vol. COM-23, June, pp.598-606.

[35] Huang, T. S. and Hussian, A. B. S. (1975), "Facsimile Coding by Skipping White," IEEE Trans. Commun. vol. COM-23, no. 12, pp. 1452-1466.

[36] Huang, T. S. (1977) "Coding of Two-Tone Images," IEEE Trans. Commun. vol. COM-25, Nov. pp. 1406-1424.

[37] Jaffe, E., Durbin, H. C. and Prince, R. W. (1964), "A Photographic Technique for determining the number of discrete Tones Required for Digital Representation of Pictures," Proc. Seventh Int. Conf. Printing Research Inst. London, pp. 283-290.

[38] Jarvis, J. F. (1964), "A New Technique for Displaying Continuous Tone Image in Black and White Facsimile Systems," IBM Tech. Disclosure Bulletine, vol. 9, pp. 636-637.

[39] Jarvis, J. F. (1976), "A New Technique for Display Continuous Tone Images on a Bi-Level Display," IEEE Trans. Commun. vol. COM-25, pp. 891-898.

[40] Jarvis, J. F. (1976), "A Survey of Techniques for the Display of Continuous Tone Pictures by Bi-Level Displays," Computer Graphics and Image Processing, vol. 5, pp. 13-40.

[41] Johnsen, O. and Netravali, A. (1981), "Progressive transmission of Two-Tone Images," IEEE Trans. communcation vol. COM-29, nol 12, Dec. pp 1934-1941.

[42] Jayant, N. S. (1970), "Adaptive Delta Modulation with One Bit Memory," BSTJ vol. 49, March, pp. 321-342.

[43] Judice, C. N., Jarvis, J. F. and Ninke, W. H. (1974), "Using Ordered Dither to Display Contunuous Tone Pictures on an AC Plasma Pannel," Proc. of SID, vol. 14/15, 4th quarter pp. 161-169.

[44] Judice, C. N. (1975), "Dithervision - A Collection of Techniques for Displaying Continuous Tone Still and Animated Picture on Bilevel Display," SID Digest, pp. 32-33.

[45] Judice, C. N. (1976), "Data Reduction of Dither Coded Images by Bit Interleaving," Proc. SID vol. 17, pp. 92-101.

[46] Judice. C. J., White, A. B. and Johnson, R. L. (1976), "Transmission and Storage of Dither Coded Images Using 2-D Pattern Matching," Proc. SID vol. 17, pp. 85-91.

[47] Kato, H. and Goodman, J. W. (1975), "Nonlinear Filtering in Coherent Optical Systems through Halftone Screen Process," Applied Optics, vol. 14, Aug. pp. 1813-1824.

[48] Kelly, D. H. (1979), "Methods & Designs Manipulation of Two-Dimensionally Periodic Stimulus Patterns," Behavior research Methods & Instrumentation, vol. 11, pp. 26-30.

[49] Kennedy, J. D. (1971), "Halftone Images Using Computer Graphics," IEEE Trans. Computer, vol. C-20, pp. 584-585.

[50] Kermisch, D. and Roetling, P. G. (1973), "Fourier Spectrum of Halftone Images," Journal Opt. Soc. Amer. vol. 65, June, pp. 716-723.

[51] Klensch, R. J. Meyerhofer, D. and Walsh, J. J. (1970), "Electronically Generated Halftone Pictures," TAGA Proc. Vol. 22, pp. 302-320.

[52] Knowlton, K. and Harmon, L. (1977), "Computer Produced Grey Scales," Computer Graphics and Image Processing, vol. 1, pp. 1-20.

[53] Kunt, M. (1975), "A Statistical Model for Correlation Functions of Two-Level Digital facsimiles," IEEE Proc. (letters) Feb. pp. 327-328.

[54] Kunt, M. (1977), "Statistical Model and Information Measurements for Two-Level Digital Fascimile," Information and Control, vol. 33, April pp. 333-351.

[55] Limb, J. O. (1969), "Design of Dither Waveforms for Quantized Visual Signals," BSTJ vol. 48, pp. 2555-2582.

[56] Lippel, B. and Kurland, M. (1971), "The Effect of Dither on Luminance Quantization of Pictures," IEEE Trans. Commun. Tech. vol. 6, pp. 879-888.

[57] Luczak, E. and Rosenfeld, A. (1976), "Distance on a Hexagonal Grid," IEEE Trans. Comput. vol. C-25, pp. 532-533.

[58] Mersereau, R. (1978), "Two-Dimensional Signal Processing from Hexagonal Raster," Proc. ICASSP-78 Record pp. 739-742.

[59] Mersereau, R. (1979), "Hexagonally Sampled Two-Dimensional Signals," Proc. IEEE, vol. 67, no. 6, pp. 930-949.

[60] Meyr, H., Rosdolsky, G. and Huang, T. S. (1974), "Optimum Runlength Codes," Proc. Trans Commun. vol. COM-22, no. 6, June, pp. 826-835.

[61] Morrin, T. H. (1974), "A Black-White Representation of a Gray Scale Picture," IEEE Trans. Computers, vol. C-23, pp/ 184-186.

[62] Morin, T. H. (1976), "Chain-Link Compression of Arbitrary Black-White Images," Computer Graphics and Image Processing, pp. 172-189.

[63] Musmann, H. G. and Preuss, D. (1977), "Comparison of Redundancy Reducing Codes for Facsimile Transmission of Documentations," IEEE Trans. Commun vol. Com-25 Nov. pp. 1426-1433.

[64] Netravali, A. N., Mounts, F. W. and Bowen, E. G. (1976), "Ordering Techniques for Coding of Two-Tone Facsimile Pictures," BSTJ vol. 55, no. 10, pp. 1539-1552.

[65] Netravali, A. N. and Prasada, B. (1977), "Adaptive Quantization of Picture Signal Using Spatial Masking," Proc. IEEE, vol. 65, no. 4, pp. 536-548.

[66] Netravali, A. N.,Mounts, F. W. and Beyer, J. D. (1977), "Techniques for Coding Dithered Two-Level Pictures," BSTJ vol. 56, no. 5 pp. 809-819

[67] O'Neal J. B. Jr. (1966), "Predictive Quantizing System (Differential Pulse Code Modulation) for the Transmission of Television Signals," BSTJ vol. 45, may pp. 689-721.

[68] Parilidi, T. (1974), "Technique for Optimal Compaction of Picture and Maps," Computer Graphics and Image Processing vol. 3, pp. 215-224

[69] Petersen, D. P. and Middleton, D. (1962), "Sampling and Resconstruction of Wave-number-Limited Functions in N-Dimensional Euclidean Spaces," Information and Control. vol. 5, pp. 279-323.

[70] Phillips, J. W. Ransom, P. L. and Singleton, R. M. (1975), "On the Construction of Holograms and Halftone Picture with Ink Plotter," Computer Graphics and Image Processing, pp. 200-208.

[71] Pocket Pal (1974) International Paper Company.

[72] Pratt, W. K. (1978), "Digital Image Processing," John Wiley and Son.

[73] Rivard, G. E. (1975), "Direct Fast Fourier Transform of Bivariate Functions," IEEE Trans. ASSP, vol.ASSP-25, pp. 250-252.

[74] Roberts, L. G. (1962), "Picture Coding Using Psuedo-Random Noise," IRE Trans. Info. Theory vol. IT-8, pp. 145-154.

[75] Robinson, A. H. (1973), "Multidimensional Fourier Transforms and Image Processing with Finite Scanning Apertures," Applied Optics, vol. 12, Oct. pp. 2344-2352.

[76] Roetling, P. G. (1976), "Halftone Method with Edge Enhancement and Moire Supression," Journal. Opt. Soc. Amer. vol. 66, no. 10, pp.985-989.

[77] Roetling, P. G. (1976), "Visual Performance and Image Coding," Proc. SID vol. 17, pp. 111-114.

[78] Roetling, P. G. (1977), "Binary Approximation of Continuous Tone Images," Proc. SPSE vol. 21, no. 2, pp. 60-65.

[79] Rosenfeld, A. and Kak, A. (1976), "Digital Image Processing," Academic Press.

[80] Rossol, L. (1971), "Generation of Halftone by Computer Controlled Microfilm Recoder," IEEE Trans. Computers vol. C-20, pp. 642-646.

[81] Rothgorat, U. (1975),"Run-Length Coding Method for Black-and-White Facsimile with a Ternary Code as an Intermediate Step," Electronic Letter, Feb. pp. 101-102.

[82] Schindler, H. R. (1970), "Delta Modulation," IEEE Spectrum, vol. 7 Oct. pp. 69-78.

[83] Schreiber W. F. (1975) "Coding of Halftone Pictures," unpublished notes.

[84] Schreiber, W. F. (1976), "Combining Text and Graphics in the Computer/Communciation Environment," Paper presented at Profit View Nov.

[85] Schreiber, W. F. and Buckley R. (1981), "Two-Channel picture Coding system:II - Adaptive Companding and Color," IEEE trans. on Commun. vol. COM-29, no. 12, Dec. pp. 1849-1858

[86] Schreoder, M. R. (1969), "Image from Computers," IEEE Spectrum, vol. 6, Mar. pp. 66-78.

[87] Schroeder, M. R. (1969), "Image from Computers and Microfilm Plotters," Commun. ACM, vol. 12, feb. pp. 95-101.

[88] Sidhu, G. S. and Boute, R. T. (1972), "Property Encoding: Application in Binary Picture Encoding and Boundary Following," IEEE Trans. Computers vol. C-21, pp. 1206-1215.

[89] Spencer, D. R. and Huang, T. S. (1969), "Bit-Plane Encoding of Continuous Tone Pictures," Proc. Symp. Computer Processing in Commun. Brooklyn, N.Y. Plolytechnic Press.

[90] Stoffel, J. C. and Moreland, J. F. (1981), "A Survey of Electronic Techniques for Pictorial Image Reproduction," IEEE Trans. communcation. vol. COM-29, no. 12. Dec. pp 1898-1925.

[91] Streifer, W, Goren R. N. and Mark, L. W. (1974), "Analysis and Experimental Study of Ruled Halftone Screens," Applied Optics, vol. 13, June pp. 1299-1317.

[92] Stucki, P. (1975), "Comparison and Optimization of Computer-Generated Digital Halftone Pictures," SID Digest, pp. 34-35.

[93] Tollenaar, D. and Ernst, P. A. H. (1964), "Conditions for Halftone Prington," Proc. Seventh Int. Conf. Printing Research Inst. London, pp. 1-17.

[94] Usubuchi, T., Omachi, T. and Iinuma, K. (1980) "Adaptive Predictive Coding for Newspaper Facsimile," Proc. IEEE vol. 68., no. 7. July pp. 807-813.

[95] Van Nes, F. L. and Bouman, M. A. (1965), "The Effects of Wavelength and Luminance on Visual Modulation Transfer," Excerpta, Miedical, Intern. Congr. Ser. No. 125, "Performance of the Eye at Low Luminance," Proc. of the Symposium, Delft.

[96] Wesner, J. W. (1974), "Screen Pattern used in reproduction of Continuous tone from Graphics," Applied Optics, vol. 13, July, pp. 1703-1710.

[97] Wintz, P. A. (1972), "Transform Picture Coding," Proc. IEEE vol. 60 pp. 43-53.

[98] Yan, J. K. and Sakrison, D. (1977), "Encoding of Images Based on a Two-Component Source Model," IEEE Trans. Commun. vol. COM-25, Nov. pp. 1315-1322.

[99] Yule, J. A. C. (1964), "The Optimum Tone Rendering in Halftone Reproduction Process," Proc. Seventh Int. Conf. Pringing Research Inst. London, pp. 17-46.