University of Alberta

IMPORTANCE DRIVEN HALFTONING

by

Lisa M. Streit ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**.

Department of Computing Science

Edmonton, Alberta
Fall 1998

0-612-34422-3

University of Alberta

Library Release Form

Name of Author: **Lisa M. Streit**

Title of Thesis: Importance Driven Halftoning

Degree: Master of Science

Year this Degree Granted: 1998

Lisa M. Streit

Lisa M. Streit
11207 22ave
Edmonton, Alberta
Canada, T6J-4V4

Date: June 9th 1998

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Importance Driven Halftoning** submitted by **Lisa M. Streit** in partial fulfillment of the requirements for the degree of **Master of Science.**

. . . .
Dr. John Buchanan

. . . .
Dr. Ben Watson

. . . .
Dr. Desmond Rochfort

Date: June 9th 1998

Abstract

Many halftoning algorithms have been proposed to simulate the original continuous tone images as closely as possible. Most of these techniques are designed to preserve particular image attributes. This thesis proposes a new generalized halftoning technique based on preserving artifacts in the image that have been predetermined by the user to be *important*. This new technique allows for both diversity in specifying the important attributes and control over local placement of ink. The flexibility of this technique allows for the creation of many different results. This thesis does not address what image attributes are important or how to display them, rather it introduces a new technique called *importance driven halftoning*. Results obtained by varying the parameters of this new technique will be illustrated and compared with traditional halftoning results, limited resource renderings and non-photorealistic rendering results.

We've all seen another wrong and right,
We know everything's not black and white,
There's always somethin' in between.

by Clint Black

Acknowledgements

This thesis is a result of interactions with many people both technically and socially.

Thanks to Mom and Dad for reminding me of what I *should* be doing and for teaching me the value of hard work. Thank you Wes, my friend and companion for understanding my need to loose myself and all track of time. Thank you for being there with me through the last two degrees. Thanks to Michael and Valerie for reminding me to laugh at the chaos.

A source of tremendous support at the University of Alberta throughout my Master's degree was the graphics and render groups. Thank you for being my second family. I have learned a tremendous amount about both graphics and life. I have cherished the experience to learn and grow with all of you. Thank you Oleg for the guidance in halftoning. Thank you Paul for the thorough proof-reading. Thanks to Mario, for reminding me Edmonton is not the beaches of Brazil. Thank you Jean Paul for listening during my late night bitching sessions. Last, but definitely not least, thank you John for being a supervisor and a friend and including us in your adventures.

Supervisory Committee

Dr. John Buchanan, Computing Science
Dr. Ben Watson, Computing Science
Dr. Desmond Rochfort, Art and Design

Contents

List of Tables

List of Figures

Chapter 1

Introduction

As graphical user interfaces become an integral part of everyday interaction with computers, the ability to display images on a variety of devices with varying limitations is essential. Images are becoming a popular medium for the efficient communication of information. The typical user expects the display of images to be transparent to the hardware on which the image is displayed.

While advances in technology allow monitors to display an increasing number of colors at increasing resolutions, the process of reducing or compressing image information or reducing the demand for band width is continually useful in multi-media applications. A significant amount of computer graphics and image processing research addresses the problem of displaying images on devices with limited capabilities. This means that often some of the image information is lost or unable to be displayed. These hardware limitations are perhaps more prominent in the printing industry. Often in color printing, an image with a continuous color spectrum must be printed using four colors of ink on white paper. Continuous grey-scale images often must be printed using one tone, black, on white paper.

In order to create the illusion of more tones than are actually available using the given display hardware the image must first be converted from a continuous tone image into an image composed of a discrete number of tones. This is typically done by digitizing the image. A digitized image is constructed from a regular grid of dots or pixels that are assigned numerical values. An 8 bit image containing 256 tones of grey can contain pixels with values ranging between 0 and 255[1]. Halftoning is the process of converting a grey-scale image into an image composed of two tones; white and black. There are two questions that must be addressed in order to design a halftoning technique. In a 2D image:

- How many pixels should have the value 0 (white) and how many should have the value 1 (black)?
- Where should these binary pixels be placed or how should they be arranged within the image?

[1] In this thesis 0 represents white and 255 or 1 represents black

1.1 Overview of Halftoning

The problem of portraying the illusion of continuous tone images using a discrete number of tones has been around since the early 1600's. This has been a challenge among many artists involved in an art-form called *mezzotint* [Ulic87]. The origin of the word **halftone** comes from the Italian word *mezzotint* describing a process that can be closely paralleled with the modern halftoning technique of random ordered dither. Artists randomly scratched an area of copper plating to create wells to hold ink. By varying the density of these scratches the artist is able to vary the amount of ink deposited on the copper plate and thus control the tone in this region. This technique was used to create the illusion of continuous tone images using copper plates and a single tone of ink during the 17th and 18th centuries.

Modern halftoning to date, has been primarily concerned with creation of two toned images that look as much like the original grey-scale image as possible (photorealistic halftoning[2]). The halftoning process takes advantage of the human eye's ability to spatially integrate or the eye's inability to view high frequency individual tones. This means that an image with half black and half white pixels evenly interspersed can create the illusion of mid-tone grey (50% black) image. Halftoning research has loosely identified goals that must be achieved in order to produce a two-tone photorealistic halftoned image. These goals include:

1. Creating the illusion of the same number of grey-scale intensities in the original image as in the halftoned image.

2. Ensuring the halftoned image does not contain *artifacts*[3] that are not found in the original image.

3. Ensuring the halftoned image does contain the details that are found in the original image.

The goal during the development of initial halftoning techniques was to achieve accurate grey-scale reproduction. This requires creating a halftoned image that induces the illusion of containing the same number of absolute tones as the number of grey-scale intensities found in the original image. There are two classes of halftoning methods with this goal: threshold matrix methods and error diffusion methods.

Threshold matrix methods [Ulic87] approximate a grey-scale intensity in the original image by some predefined arrangement of pixels. This arrangement, namely the number of pixels to be set and the position of the pixels, is defined by the matrix known as a dither or threshold matrix. There are two types of dither matrices used in threshold matrix halftoning: clustered and dispersed. Clustered dither matrices (see figure 1.1(b)) grow the arrangement of black pixels from the center of the matrix outwards as the intensity becomes darker. Dispersed dither matrices (see figure 1.1(c)) set pixels

[2]For the purposes of this thesis, photorealistic halftone is defined as a halftoned image that appears as visually similar to the original continuous tone grey-scale image as possible

[3]For the purposes of this thesis, artifacts are defined as unwanted high frequency image information

(a)

(b) $$\begin{vmatrix} 14 & 12 & 13 & 16 \\ 5 & 4 & 3 & 10 \\ 6 & 1 & 2 & 11 \\ 9 & 7 & 8 & 15 \end{vmatrix}$$

(c) $$\begin{vmatrix} 10 & 6 & 11 & 7 \\ 4 & 14 & 1 & 15 \\ 12 & 8 & 9 & 5 \\ 2 & 16 & 3 & 13 \end{vmatrix}$$

(d) $$\begin{vmatrix} 0.0 & 0.0 & 0.4375 \\ 0.1875 & 0.3125 & 0.0625 \end{vmatrix}$$

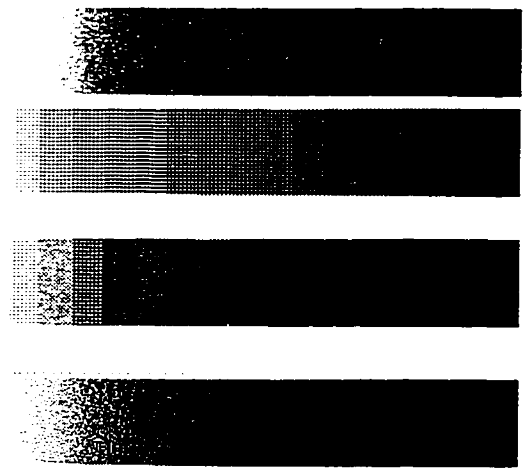


Figure 1.1: (a) Original Image; (b) Clustered Threshold Matrix with the 4x4 threshold matrix on the left; (c) Dispersed Threshold Matrix with the 4x4 threshold matrix on the left; (d) Floyd-Steinberg's Error Diffusion using the error matrix on the left.

apart from each other as the intensity becomes darker, so as to not form clusters. Threshold matrix halftoning methods are simple and efficient. They do however, typically cause *banding*[4] (see figure 1.1(b) and (c)) to occur in the halftoned image since the threshold matrix can only approximate a limited and discrete number of grey-scale intensities. This can be corrected by increasing the size of the threshold matrix. However, increasing the size of the matrix causes the regular artifacts introduced by the tiling of the dither matrix to become larger and hence more obvious.

Error diffusion methods [Floy76, Ulic87] threshold a single pixel and set it to black or white in the halftoned image. The error between the pixel in the original grey-scale image and the pixel in the halftoned image is calculated and distributed to neighboring pixels along some predetermined path through the image. This error distribution can also be done over groups of pixels or regions instead of single pixels. The next pixel or region along the path is then thresholded based not only on the intensity of the region in the original image but also on the error that has been passed to it from neighboring regions. Error diffusion methods tend to introduce artifacts into the halftoned image (see figure 1.1(d)) due to the sequential, predetermined path of diffusion. However, this technique also eliminates the banding effect seen when using threshold matrix halftoning methods.

Hybrid halftoning methods incorporate the efficiency of threshold matrix methods with the advantages of error diffusion methods. One example of this is varying the threshold used in error diffusion according to some threshold matrix [Bill83] or over some range (*i.e.* white noise) [Ulic87]. A second example is locally thresholding the pixels according to a threshold matrix, but distributing error accumulated over the threshold matrix from one matrix region to another [Bill83]. These hybrid methods tend to reduce the introduction of artifacts that are common to both threshold matrix

[4]The appearance of regular, usually false contours

and error diffusion halftoning methods.

All the above mentioned methods achieve reasonable results in approximating grey-scale intensities, however these methods sometimes introduce artifacts or destroy image details, such as edges.

More recent work in halftoning has focused on the preservation of edges in addition to grey-scale intensity reproduction. Edges are important image attributes that should be preserved throughout the halftoning process. There have been three classes of techniques developed that attempt to satisfy these goals. The first class pre-processes the original image to enhance edges in order to help their preservation through the halftoning process. The second class alters the halftoning technique to avoid destroying edges during halftoning. The third class post-processes the halftoned image to enhance edges after halftoning.

As more is learned about how humans perceive images and what image attributes are important there may be a desire to ensure the preservation of other important image attributes throughout the halftoning process. There may also be a need to re-order their importance. Current halftoning techniques are designed and tailored to achieve very specific goals and are thus hard to adapt to ensure the preservation of other important image features.

1.2 Motivation

A virtually instantaneous perception of the image is formed upon viewing the image, however the longer the image is viewed, the more the perception of the image develops. The human eye is known [Snyd85a] to work as a multi-resolution filter. Thus, once this initial impression of the image is formed the eye then focuses at varying levels of detail on various attributes of the image. A complete set of rules that define the order of image attribute importance have not yet been identified. However, since the human perceptual system works at multiple resolutions it is productive to design or construct images with this notion in mind. Thus, not only must important image attributes be identified, but it must also be known at what resolution these attributes become important and the preservation of these attributes at the necessary resolutions must be ensured. For example, grey-scale intensity changes seem to be noticeable even at low resolutions (on a global scale), where as edges and high frequency information alterations seem only noticeable at some of the higher resolutions (on a local scale). The author is not aware of a halftoning technique that allows the user to specify a general importance function in order to identify important attributes of the image and use this information at multiple resolutions for the halftoning of local and global image attributes.

Limited resource rendering is another well-studied problem in computer graphics. Limiting the quantity of information in the image and hence reducing or "compressing" the size of the image is useful for transporting image information faster, while using less bandwidth. This efficiency is necessary for use in multi-media applications. Halftoning already reduces the number of resources needed for the display of an image by reducing the number of intensities in the image. The required resources can be further reduced by decreasing the number of display primitives (black pixels) in the image. The ability to easily adjust the quantity of display primitives used in the halftoned image can be very beneficial to the printing industry. Not only can it reduce costs of printed images by reducing the amount of toner used, but it can perhaps assist the quality of the images during various moments in the toner cartridge's life span. For example, with laser printers there is often bleeding of ink between pixels commonly known as *dot gain*. This is more prominent when a new toner cartridge is used. Due to dot gain, dark regions become much darker than desired, at the expense of clarity and preservation of detail. Reducing the amount of ink while retaining relative intensity can correct this problem. This technique can also be useful for the construction of draft copies of documents, when high quality is not needed and resources such as ink and printing time are limited. A halftoning technique that easily adjusts the placement of ink dependent on the number of display primitives may be very useful for creating images with less resources. Currently, traditional halftoning techniques do not provide an easy method to limit the number of required display primitives.

Finally, it may be desirable to preserve intensity approximation, yet use various local effects to create these intensities. This altering of the local arrangement of ink in line segments or textures is quite commonly used in the creation of non-photorealistic renderings. The balance between texture and tone is achieved by retaining the global distribution of display primitives, but altering the local distribution of ink within the display primitive. For instance, individual pixels can be clustered to form circular regions, or clustered to form line segments. Thus, being able to control the structure of display primitives throughout the halftoning process can lead to control over image texture or style. Variations in style can lead to the creation of results that range from photorealistic to non-photorealistic.

This thesis presents a generalized halftoning technique that can distribute a user-defined number of display primitives at multiple-resolutions according to a predefined importance function. Once the specified number of display primitives have been distributed the local position of ink is defined by the structure of the chosen type of display primitive. This technique will be shown to create effects such as traditional halftoning and non-photorealistic rendering.

1.3 Scope of Thesis

This thesis formulates a new generalized halftoning technique to achieve the following goals:

1. Allow the user to specify any importance function or combination of importance functions that identify important image attributes at varying resolutions.

2. Allow the user to have control over the number of display primitives and ultimately the amount of ink that is used in the halftoned image.

3. Allow the user to control the local position or structure of ink by altering the type of display primitive used in the halftoned image.

This technique assumes that the user has defined the important image attributes as a function F(x,y). The importance function will be used as a black box defined by the user. This thesis will not be concerned with how to best render a given image attribute and thus, will not attempt to identify which type of display primitives should be used for the rendering of different image attributes. Rather it will illustrate how the type of display primitives and their position can be controlled to highlight different image attributes. This thesis is primarily concerned with the design of a generalized halftoning technique that will allow the user to specify various types and quantities of display primitives as well as various importance functions. This generalized technique is loosely defined by addressing the following three questions:

1. How many display primitives are to be used in the halftoned representation of the image?

2. How does the importance function affect the distribution of the display primitives?

3. What affect does different display primitives have on fine local placement of ink?

The three user controlled parameters of this halftoning technique form the answers to these questions which in turn form the solution to the two halftoning sub-problems, as outlined at the beginning of this chapter. The number of display primitives used and where they are placed are indirectly controlled by the user, but ultimately defined through the generalized halftoning technique.

In order to allow the importance function to be defined and evaluated at multiple resolutions, a pyramidal representation [Will83] of the image is used. This representation will allow for the approximation of both global and local ink distribution. In this manner, relative approximations can be retained in order to satisfy our visual system sensitivities and yet the local position of ink can be adjusted in order to satisfy both the visual system and display medium requirements. Controlling the local placement of ink enables the preservation of global image information while still allowing for the creation of non-photorealistic and limited resource renderings effects.

This thesis will present the necessary background needed to provide the motivation and understanding for this new importance driven halftoning technique. Specifically, a comprehensive overview of halftoning techniques, an overview of the basics of the human visual system and an overview of non-photorealistic rendering (NPR) techniques are presented. The details of the importance driven halftoning method and its associated parameters are presented and various results are shown. Finally, these results will be compared with other traditional halftoning and NPR results.

1.4 Analysis of Results

The results of importance driven halftoning will be compared with both traditional halftoning and NPR results. The comparison will be done in two ways: visual comparison and quantitative metrics. The visual comparison will be done by presenting similar results and analyzing specific components of the images. Ulichney's [Ulic87] metrics of power spectrum and anisotropy will be used as the first quantitative measure for assessing frequency distributions and symmetry throughout the halftoned image. A multi-scale edge metric proposed by Veryovka *et al* [Very98] will also be used to assess the preservation of edges or introduction of false edges. The evaluation of importance driven halftoning non-photorealistic results will be done by visually comparing NPR images with importance driven halftoning rendered images, as there does not currently exist a quantitative measure for such analysis.

1.5 Summary and Direction of Thesis

This chapter presented a brief overview of halftoning leading to a description of the motivation for importance driven halftoning. The scope of the thesis was outlined and the method for evaluating importance driven halftoning is given. Chapter two reviews the background literature on halftoning, introduces the basics of non-photorealistic rendering techniques, and overviews the basics of the constraints and capabilities of the human visual system. Chapter three describes a detailed formulation of what must be achieved to provide a solution for the halftoning problem and describes the details of the importance driven halftoning technique. Chapter four discusses the results obtained by using importance driven halftoning with varying parameter values. Results similar to traditional halftoning techniques, NPR, and limited resource rendering results are created using this new technique. Chapter five presents the evaluation of the importance driven halftoning technique as well as the visual and quantitative evaluation of this technique's results by comparing these results with results of other rendering techniques (both halftoning and pen-and-ink illustrations). Chapter six summarizes the objectives, details, and results of importance driven halftoning as well as discusses future extensions to this technique.

Chapter 2

Background and Related Information

Halftoning is a necessary step for the display of continuous tone images on binary devices. The goal of this process is to achieve, as stated by Ulichney, "the illusion of continuous-tone images from a judicious arrangement of binary picture elements". This process relies on the properties of the human visual system for the creation of photorealistic halftoned and non-photorealistic halftoned images.[1] The human eye works as a filter that spatially integrates over a small area and thus has the inability to distinguish fine details at increasing distances. This means that an area of interspersed black and white pixels will be interpreted as a grey tone. There are other properties of the human visual system that can be utilized in conjunction with attributes of continuous tone grey-scale images to help guide or induce a certain illusion or interpretation of the image. The first section of this chapter will outline the basics of the human visual system and how it can be used in conjunction with halftoning.

Halftoning, as with any process that projects a continuous scale onto a binary scale, has an obvious loss of information. The original objective of halftoning was to absorb this loss of information by using the properties of the human visual system to create the illusion of continuous tone and hence create a photorealistic halftoned image. A second and more recent objective is to use halftoning to purposely introduce artifacts into the image. This involves altering or adding information to the halftoned image for the production of special effects. These special effects include simple highlights of image attributes or artistic, texture effects. The second section of this chapter will give a comprehensive overview of traditional halftoning techniques and some methods for assessing the results of these techniques. The third section of this chapter will outline the necessary background information on non-photorealistic rendering techniques applicable to halftoning. Finally, background on multi-resolution image representations and an outline of importance driven halftoning are presented

[1] For the purposes of this thesis a **photorealistic halftone** is defined as a halftoned image that appears as visually similar to the original continuous tone grey-scale image as possible and a **non-photorealistic halftone** is defined as a halftoned image is similar to the original continuous grey-scale image, but will have added artifacts and texture effects.

in this chapter.

2.1 Perceptual Basics

The overall objective of halftoning is to construct images that induce certain interpretations or visual impressions. Once the image is constructed, the human visual system is used to view the image and form this interpretation. In order to successfully achieve the illusion of identical appearances or alternatively the appearance of special effects there is a need to both understand and utilize some of the basic components of the human visual system in order to influence the perception of the image.

The human visual system is a vast, complex, and not completely understood component of the human body. Some halftoning methods have been based on visual or perceptual models ([Mull92, Kolp92, Papp92, Chu92, Hilg94]). However, there is no currently accepted, complete and simple visual or perceptual model. Thus, these halftoning methods use various perceptual models and are quite complex. The human visual system has been loosely divided by Snyder [Snyd85a]) into three segments: the physical eye, the neural pathways that connect the eye to the brain, and the portion of the brain responsible for vision coordination (*i.e.* the visual cortex). Current research has revealed a lot of information about the physical eye, less about the activities of the neural pathways, and very little about the activities in the visual center of the brain.

Halftoning has used attributes of the human visual system in order to achieve its objectives. These attributes are known responses of the physical eye and the neural pathways. A few of the commonly accepted characteristics of the human visual system that are used in both image processing and graphics are:

1. The human visual system's ability to spatially integrate at certain spatial frequencies.
2. The human visual system's ability to act as a multi-resolution passband filter.
3. The human visual system's dependence on background illumination for the perception of intensity.
4. The human visual system's preference toward certain visual organizations.

In this section details about these three characteristics will be presented. As well an outline will be presented of how the human visual system perceives some common attributes of images, particularly attributes common in continuous grey-scale images.

2.1.1 Characteristics of the Human Visual System

Many attributes of the human visual system can be used in halftoning techniques. Attributes that traditional halftoning techniques utilize involve both how the human eye operates and how the human visual system forms interpretations of images and image attributes. This section will discuss how the human visual system spatially integrates at certain spatial frequencies, how the visual system is

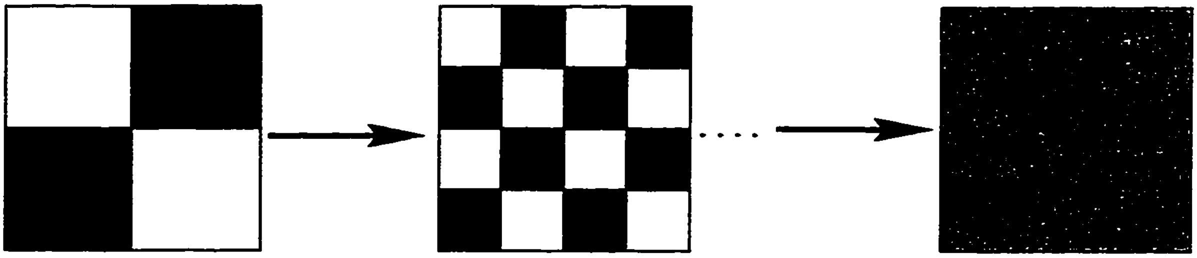


Figure 2.1: Viewed up close the image appears as an arrangement similar to the image on the left. Viewed at a distance and thus, higher spatial frequency the eye performs spatial integration over the region to have the image interpreted as the image on the right (a mid tone of grey - 50 % black)

able to focus at multiple resolutions for the perception of different levels of detail, how contrast is interpreted and how visual organization affects the interpretation of visual clarity.

Spatial Integration

The human eye does not have the ability to distinguish small features with high spatial frequencies at increasing distances. Thus, as viewing distance increases the spatial frequency increases. At a certain threshold the eye ceases to see the individual features and approximates the region by averaging the features of that region together[Fole90]. If an area of equally interspersed black and white pixels is viewed at a distance, the eye will spatially integrate over this region to perceive a region of mid-tone grey (see figure 2.1). The viewing distance only has to be large enough so that spatial frequencies of the features reaches 3 cycles/degree [Snyd85a]. Below this frequency the visual system is less efficient in integrating over the area and individual features are detected.

Multi-Resolution Filter

The human visual system [Snyd85a] behaves as a Fourier analyzer in the spatial domain. This means it has the ability to perform multi-resolution filtering in order to focus on varying levels of detail. For example, the visual system may form an initial interpretation of an image, but upon further examination finer details of the image are noticed. This is simply because the initial interpretation was formed from a filtered, low resolution version of the image and the finer details were later observed from higher resolution versions of a particular image area.

Relative Contrast or Perceived Brightness

Gonzalez and Woods[Gonz92] states that the perceived brightness of a region is not simply a function of intensity. The perceived foreground brightness also relies heavily on the background illumination or intensity. This is illustrated in figure 2.3. Thus, the perceived intensity really depends on relative contrast. If this contrast is small enough there will be no perceived difference. The threshold where the contrast is not discernible is defined by the weber ratio (see figure 2.2). This was devised using a human subject with the initial background illumination set to I and the difference in intensity

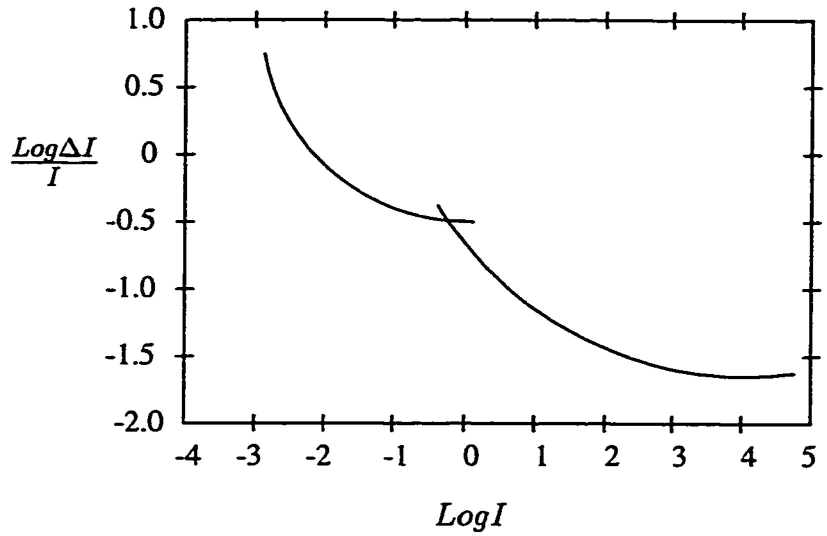


Figure 2.2: Weber Ratio as a function of intensity. The weber ratio is the minimum detectable (50% of the time) change in intensity (ΔI) given the background intensity I.

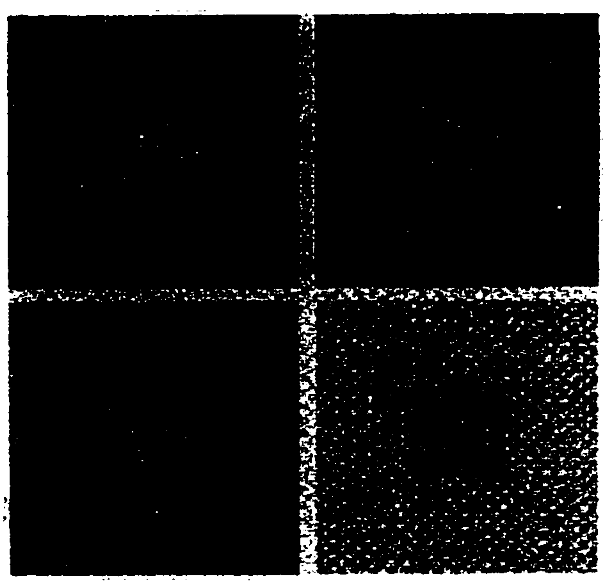


Figure 2.3: Perceived intensity is a function of both foreground and background intensity. The square in the center is the same intensity in all four regions, but is perceived as darker as the background becomes lighter.

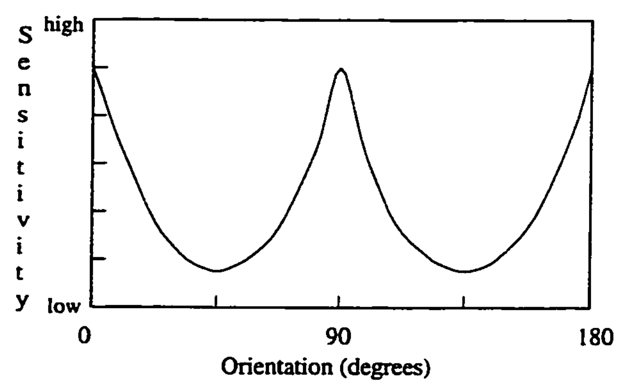


Figure 2.4: Comparing spatial frequency with sensitivity shows that the eye is more sensitive to horizontal and vertical orientations than any others. Ulichney states that this disparity in sensitivity increases with spatial frequency, see Ulichney [Ulic87] page 84.

between the object and its background, ΔI, gradually increased until the human subject can perceive a difference in brightness. The quantity ΔI_c/I where I_c is discriminable 50% of the time is called the weber ratio. Figure 2.3 shows that as the background intensity or illumination is decreased the contrast needs to be increased before a change in intensity is detected. This means that in regions of low background luminance the eye is less sensitive to contrast than in regions of high background luminance.

Visual Organization and Clarity

The human visual system is sensitive to different types of phenomena in an image or scene. Two examples of this phenomena are the tendency to recognize attributes aligned at particular orientations and the tendency to form visual clarity.

The human visual system is more sensitive to image artifacts aligned [Snyd85b] horizontally (0^o) and vertically (90^o) than to features aligned at other orientations [Ulic87, Knut87] (see figure 2.4). This means that any attributes aligned in this fashion will be noticed first and hence be more obvious.

Foley *et al* [Fole90] claims that if the interpretation of an image is readily apparent this is due to visual clarity, which can be created by using the visual organization of information to reinforce or emphasize image attributes or logical organization. Rules of visual organization have been proposed by the Gestalt psychologist Wertheimer [Wert39] as early as 1930. These rules are primarily concerned with similarity, proximity and good continuation. In Foley *et al* [Fole90] p.418 the following is stated:

> The rule of *similarity* states that two visual stimuli that have a common property are seen as belonging together. Likewise, the rule of *proximity* states that two visual stimuli that are close to each other are seen as belonging together. The rule of *closure* says that, if a set of stimuli almost encloses an area or could be interpreted as enclosing an area then the viewer sees that area. The *good-continuation* rule states that, given a juncture of lines, the viewer sees as continuous those lines are smoothly connected.

This is illustrated in figures 2.5 and 2.6 from Foley *et al*

Since the human visual system is sensitive to certain orientations and favors visual clarity, the placement of image attributes should favor the correct interpretations. This means placing image attributes and features so as to not aid in the interpretation of artifacts or false edges.

2.1.2 Application to Continuous Tone Grey-Scale Images

Using our knowledge of how the human visual system forms interpretations the placement of image attributes can be tailored to favor particular interpretations of the final image. It may be desirable to try to favor the perception of a continuous grey-scale image similar to the original image or the perception of the original image drawn artistically or altered in some way.

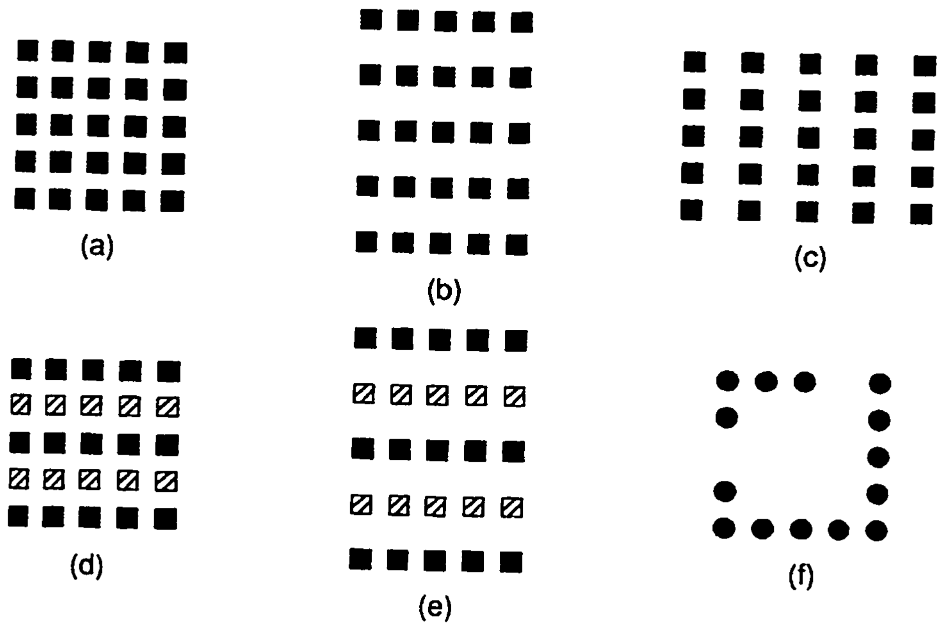


Figure 2.5: From Foley *et al* [Fole90] Gestalt rules: In (a), the squares are undifferentiated. In (b), proximity induces a horizontal grouping; in (c), it induces a vertical grouping. In (d), similarity induces a horizontal grouping, which is further reinforced in (e) by a combination of proximity and similarity. In (f), closure induces a square of dots, even though two dots are missing.

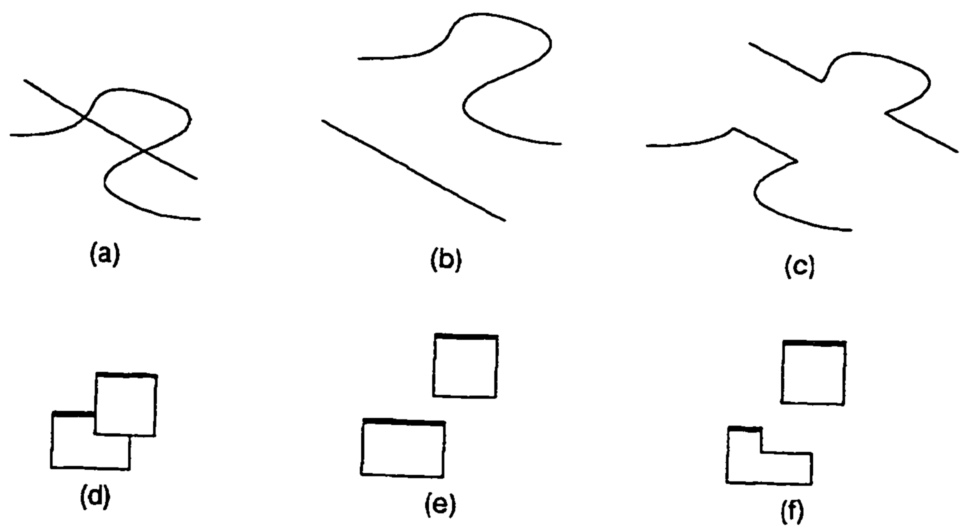


Figure 2.6: From Foley *et al* [Fole90] More Gestalt rules. The two intersecting lines in (a) could be interpreted as show in either (b) or (c). Good continuation favors (b). In a more applied context, the two overlapping windows of (d) could be interpreted as shown in either (e) or (f). Good continuation favors (e).

All continuous grey-scale images have common properties that must be displayed in a particular fashion in order for the required interpretation to be formulated. There are some properties that seem to be of particular importance in influencing the interpretation of the image and have thus, been of interest throughout the halftoning process and in the evolution of the results of this process. Some of these properties are:

- The accurate representation of the same number of perceived grey-scale intensities as found in the original continuous grey-scale image.
- The approximation of uniform regions of intensity (constant, increasing and decreasing).
- The representation of edges or high contrast regions.
- The preservation of important image attributes.

Approximation of Grey Tones

Creating a black and white image that is perceived as much like the original grey-scale image as possible involves closely approximating the grey-scale intensities of the original image. Three problems involved in accomplishing this objective are all associated with the perception of intensity.

First, the number of grey tones in the image can be approximated by utilizing the visual system's ability to spatially integrate. Using black pixels that are sufficiently small and viewing the image from a minimum required distance, the spatial frequency needed to induce the perception of grey tone can be achieved.

Second, studies [Gonz92] show that there exists a minimum difference in intensity before this change is detectable by the human visual system. This limits the number of intensities that need to be approximated to the number detectable by the human visual system. Also, each intensity only needs to be approximated by a sufficiently close intensity to be perceived as the same intensity.

Third, considering that the perceived brightness of an intensity change relies on the background illumination or intensity, perhaps only relative intensity or contrast needs to be preserved. This can lead to the creation of perceptually similar images where one has less grey-tones than the other.

Thus, by utilizing the eye's ability to spatially integrate and keeping in mind the visual system's limitation to decipher minimum intensity changes and perceived brightness a binary image can be constructed using these attributes to absorb some of the loss of information inevitable in the halftoning process.

Regions of Uniform Intensity

Another attribute of many continuous grey-scale images are regions of uniform intensity. Regions of uniform grey-scale intensity are typically halftoned using some arrangement of black and white pixels. In the original images these regions are continuous and lack obvious components. Due to the human visual system's desire for visual clarity (closure and good continuation), care must be taken not to induce the perception of obvious connected components in the arrangements of black and white pixels. These obvious, unwanted, connected components are commonly referred to in halftoning research as *artifacts*.

In regions of uniform increasing or decreasing intensity not only must the introduction of unwanted artifacts be limited, but the continuity in intensity change must be preserved. Perceptually, obvious transitions between arrangements of black and white pixels in order to approximate increasing and decreasing grey-scale intensities, must be eliminated or smoothed out.

Regions of High Contrast or Detail

Another property common to many continuous tone grey-scale images are regions of high contrast or edges. Since the visual system favors closure and good continuation of connected components (particularly edges) in the original image it is essential that these perceptions are also induced with the photorealistic halftoned image. As well, the perception of connected components in the halftoned image that are non-existent in the original image must be limited. This addition of components is called the introduction of false edges or false features.

Perceptually Relevant Image Attributes

Occasionally, certain image attributes may be of more importance than others. In these cases there may be a desire to ensure preservation of these image attributes at the expense of others. For example, the attributes in the foreground of an image may be much more important than those in the background and hence it may be satisfactory to blur detail in the background, while ensuring the preservation of detail in the foreground. However, in a typical halftoned image, image component information is not available and defining image components is a large, complex problem currently being investigated by the image processing community. In the future, guidelines for knowing which components of an image are perceptually relevant may be developed. Once these are defined and these components can be identified, the preservation of these image attributes can be ensured by encapsulating this information into an importance function. Until then, halftoning algorithms are generally limited to use 2D intensity information only.

2.2 Traditional Halftoning

Throughout the evolution of the halftoning process researchers have identified image attributes that must be preserved in order to create a photorealistic halftoned image. Initial halftoning attempts focused on re-creating the grey-scale intensities as closely as possible without the introduction of noticeable artifacts to the image. Artifacts are usually defined as the presence of regular or irregular high frequency information in the halftoned image that was not present in the original image. Other more recent objectives include the preservation of edges and detailed or high contrast regions of the image.

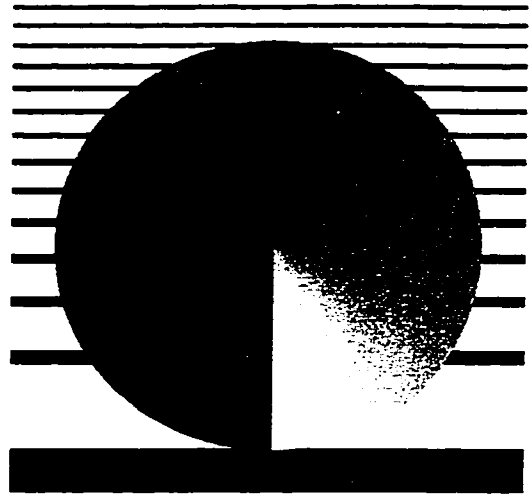


Figure 2.7: Original continuous tone gray-scale image. A computer generated continuous tone grey-scale circular ramp with black and white bands in the background.

2.2.1 Grey-Scale Intensity Approximation/Preservation

Many different techniques were developed primarily to approximate grey-scale intensities without the introduction of artifacts. These techniques can be loosely classified as: Threshold Matrix and Error Diffusion techniques.

Threshold Matrix halftoning techniques create the binary image by approximating each grey-scale intensity by an $n \times m$ arrangement (matrix)[2] of black and white pixels or by using the matrix to threshold a region of $n \times m$ region of pixels in the original image. This $n \times m$ matrix is a set of thresholds which define how to order or arrange the black pixels and is commonly known as a *dither matrix* [Baye73, Jarv76a]. A summary of these *ordered dither* techniques before 1976 is described by Jarvis and Roberts in [Jarv76b]. There are two methods for arranging the thresholds within the dither matrix: clustered and dispersed [Ulic87]. Clustered ordered dither methods cluster the thresholds within the matrix such that the black pixels are arranged to form clusters that grow from the center of the threshold matrix outwards over the range of grey-scale intensities. Dispersed ordered dither methods arrange the black pixels such that clusters are not formed within the matrix. This is illustrated in figures 2.8 and 2.9.

Threshold matrix techniques are fast and particularly useful for some display devices, but they trade off artifact size for the number of approximated grey-scale intensities. Threshold matrix techniques are fast because only a simple comparison is needed, in order to set the pixel. They are particularly useful for display devices that cannot set individual pixels. For example, on laser printers the ink bleeds causing black pixels to be larger than intended. This effect is generally termed *dot gain*. For these devices a clustered threshold matrix can be used to set pixels in clusters rather than individually. An analysis of halftoning techniques for displays at different arrangements and

[2]generally the matrix is square, so $m = n$

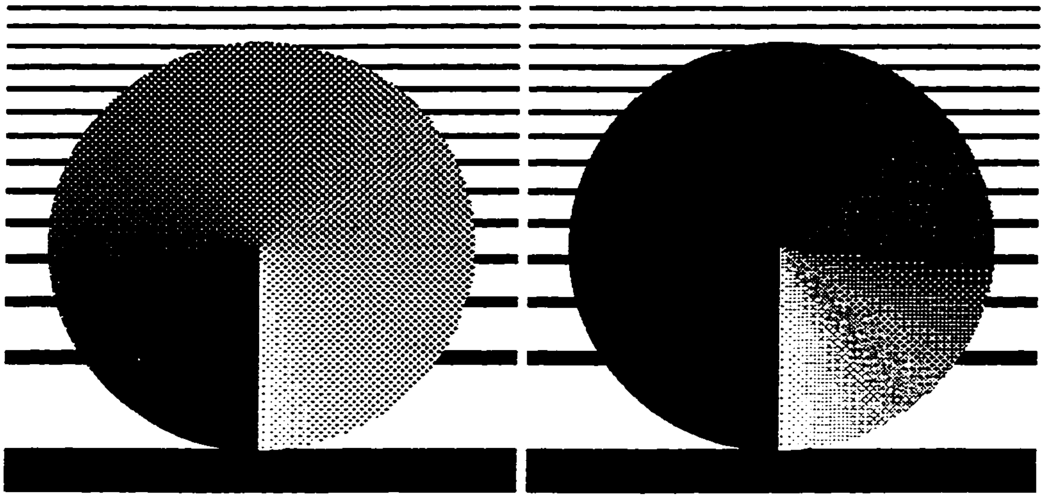


Figure 2.8: 8x8 Threshold Matrix: left - clustered; right - dispersed. Both are 200dpi images.

angles was presented by Ulichney [Ulic87]. Holladay [Holl80] presented a method for calculating the threshold matrix for any angled screen. Threshold matrix techniques form obvious ***banding*** due to the discrete number of intensities that can be approximated ($n^2 + 1$) and the lack of distribution of the *quantization error*. The number of intensities that can be approximated can be increased, hence decreasing the banding effect by increasing the size of the threshold matrix. However, as the size of the matrix increases so does the size of the clusters, thus creating larger, more obvious artifacts. It has been noted that using larger threshold matrices not only allows for the approximation of more tones, but eliminates some of the artifacts created from tiling the matrix over the image. Thus, much of the current research on these techniques has concentrated on creating large matrices that do not introduce artifacts into the image. Purgathofer *et al* [Purg94] briefly overviews these attempts. Chu and Watunyuta [Chu92] have tried creating an optimal dither matrix based on the characteristics of a particular model of the human visual system. Ulichney [Ulic93] constructed a dither matrix by evaluating the measure of gaps or voids and clusters with each additional pixel setting. Ostromoukhov *et al* [Ostr94, Ostr95a] attempted to reduce the obviousness of the regular artifacts by rotating the dither matrix. Mitsa and Brathwaite[Mits95] used wavelets to construct the optimal dither matrix by controlling the spectral energy. Dalton [Dalt95] described various perceptions of binary texture and outlined possible models used in the creation of optimal dither screens. Currently, research has determined that the human visual system is less sensitive to patterns of ***blue noise*** [Ulic87] and creating a dither matrix using blue noise, called a ***blue noise mask***, results in the introduction of less obvious artifacts [Ulic88, Mits92, Schu94]. Yao and Parker [Yao95] recently addressed the problem of compensating for ***dot gain*** when using blue noise masks. Finally, Buchanan and Streit [Buch98a] suggested using a segmented shape space filling curve as the threshold region. They stated that this reduces the artifacts associated with regular threshold matrices since the threshold matrix is not tiled regularly.

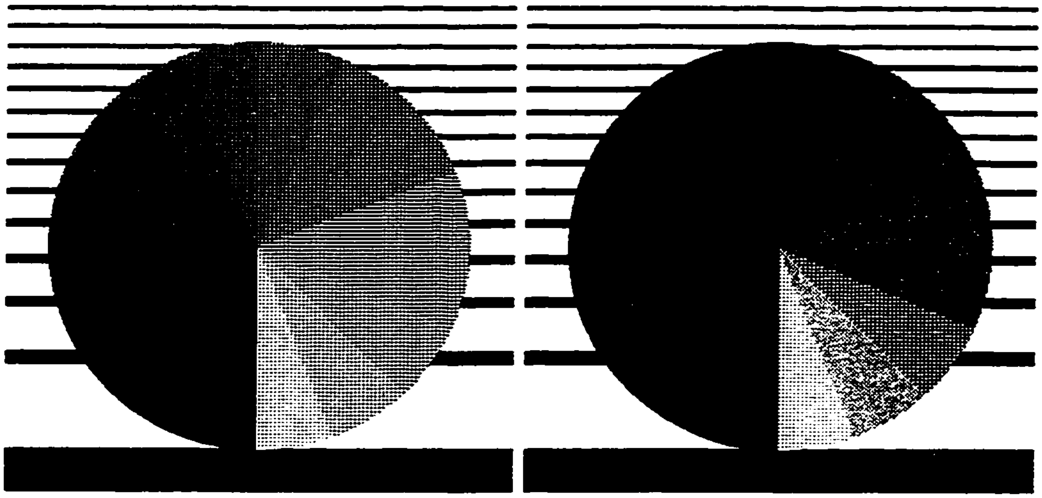


Figure 2.9: 4x4 Threshold Matrix: left - clustered; right - dispersed. Both are 200dpi images.

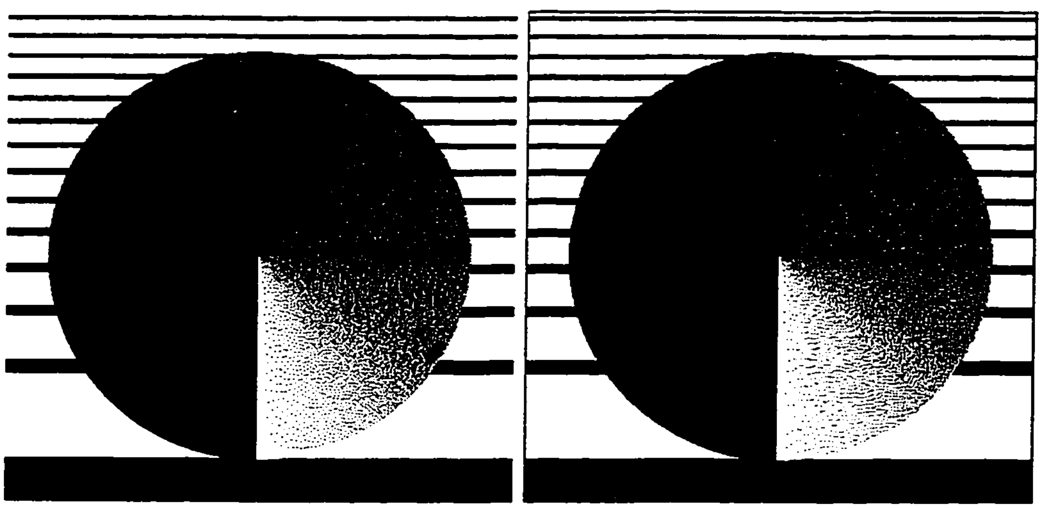


Figure 2.10: Left - Floyd-Steinberg Error Diffusion; Right - Naiman and Lam's error diffusion. Both are 200dpi images

Error diffusion techniques [Floy76] create the binary image by thresholding a region of pixels and then calculating the error over this region of pixels by calculating the difference between the binary approximation and the original image. Once the *quantization error* has been calculated it is diffused to surrounding regions. These regions are then thresholded by analyzing both the error that has been diffused to them and the original grey-scale intensities. Error diffusion techniques more smoothly approximate grey-scale intensities as seen by the lack of banding effects in figure 2.10. However depending on the path of diffusion through the image this technique may create obvious artifacts in the halftoned image. Research in this area has concentrated on altering both the error diffusion filter and the path through the image such that artifacts are not introduced. Ulichney [Ulic87] experiments with various error diffusion filters. Kolpatzik and Bouman [Kolp92] tried to create an optimal error diffusion filter by combining the traditional filter with the lowpass characteristics of the contrast sensitivity function for the human visual system. Wong [Wong95] tried to "look-ahead" to predict the error and set the pixel accordingly in order to minimize the propagation of error.

Traditionally, the path over the image has been left to right and top to bottom [Floy76] (see figure 2.10). Knuth [Knut87] suggested specifying the order of processing within the matrix and distributing error only to unprocessed neighbors. Witten and Neal [Witt82] suggested using a one-dimensional Peano curve (based on the Hilbert polygon) divided into segments for thresholding regions. Cole [Cole91] later suggested using Murray polygons rather than Hilbert polygons, since many picture shapes are rectangular rather than square. Velho and Gomes [Velh91] suggested using a one-dimensional space-filling curve and then tried to tailor this technique to adaptively place the cluster within the region according to the darkest part of the underlying image. Rodriguez [Rodr94] attempted to reduce the "rivers" of error introduced by traditional Floyd-Steinberg error diffusion by processing each scanline in a forward and a backward pass. He found this created more symmetric artifacts that were less noticeable to the human visual system. Naiman and Lam [Naim96] suggested taking a wavefront path through the image. Distribution of the quantization error starts at the center of the image and is diffused outwards towards the edges of the image.

Some halftoning techniques have been devised which attempt to combine both error diffusion and threshold matrix techniques. Billotet-Hoffmann and Bryngdahl [Bill83] altered the threshold used during error diffusion according to a threshold matrix. They also diffused error from one region to the next using threshold matrix techniques within that region. Ulichney[Ulic87] perturbed the constant threshold typically used in error diffusion techniques so that it varied within a given range (white noise). Velho and Gomes [Velh91] diffused error from one region to the next along their one-dimensional space filling curve. Knox [Knox93] analyzed the theory of the error diffusion technique and attempted to explain why these modifications yield slight improvements to the results of traditional error diffusion. Zhang and Webber [Zhan93] combined Knuth's [Knut87] dot diffusion with Velho and Gomes [Velh91] space filling curve diffusion. Knox [Knox94] reduced the distribution of quantization error by weighting the image intensity more heavily than the error. Buchanan and Streit [Buch97] suggested combining error diffusion techniques in manners similar to both Ulichney [Ulic87] and Billotet-Hoffman and Bryngdahl[Bill83].

2.2.2 Edge Preservation

A more recent objective in halftoning has been the preservation of edges in the image in addition to approximating the grey-scale intensities without the introduction of artifacts. This notion comes from the fact that edges are high contrast areas and the human eye is sensitive to ([Knut87, Fole90]) this type of information.

There have been three approaches taken in order to achieve this objective. The first approach involved enhancing the edges in the original image before the image is halftoned in order to help preserve the edges through the halftoning process [Jarv76b, Knut87]. The second approach adjusts the halftoning technique to locally compensate for edges when they have been detected. Thurnhofer [Thur94] suggested using a non-linear error diffusion filter to adapt to image characteristics so as to reduce the destruction of fine detail, at the same time not overemphasizing edges where it is not appropriate according to Weber's law. Knox [Knox94] suggested weighting the error less than the actual intensity, reducing the amount of error distributed, thus lessening the distortion of edges. Velho and Gomes [Velh95] used space filling curve halftoning and altered the size of the cluster along the curve inversely to the magnitude of the gradient over the cluster. Thus, in regions of high contrast or rapidly changing intensity the cluster was smaller. Buchanan and Verevka [Buch95] used space-filling curve halftoning and altered both the placement of black pixels within the clusters and the size of the clusters dependent on whether an edge was present. The third approach involves re-enhancing the edges after the image has been halftoned in order to compensate for any destruction of the edge throughout the halftoning process. Buchanan *et al* [Buch98b] suggested re-drawing the edges in the halftoned image in regions where edges have been detected. The edge is drawn in one of three possible ways: a row of white pixels, a row of black pixels or a row of white pixels and a row of black pixels.

In addition to grey-scale intensity approximation and edge preservation other image attributes may need to be preserved. None of the above techniques can be easily adjusted to accommodate for this change in focus. All of the techniques have been specifically designed for the preservation of some specific image attribute. This thesis presents a method for halftoning based on an importance function. These results will be presented in [Stre98].

2.2.3 Computational Trade-off

Most of these early techniques were designed around very harsh restrictions on memory and complexity. The most efficient in both these areas are ordered dither methods and with a slight increase in memory, simple error diffusion methods such as Floyd-Steinberg error diffusion. With the increase in computational powers and available memory other techniques have been developed to produce more optimal results. These techniques halftone the image and then evaluate the halftoned image according to some quality metric which can be based on anything from a model of the human visual system to a printer model. Once the halftoned image has been evaluated, it is altered based on the metric to improve the evaluation of the image. This alteration and re-evaluation is repeated until the maximum number of iterations or the optimal result is reached. Analoui and Allebach [Anal92] utilized binary search to find the optimally halftoned image with a difference metric incorporating the contrast sensitivity function as the measure for evaluation. Mulligan and Ahumada [Mull92] used sequential adjustment with a measure of the vector length of error filtered by the contrast sensitivity function to find an optimized solution with the option of using simulated annealing for the optimal solution. Pappas and Neuoff [Papp92] used a measure of least squared error on a model incorporating characteristics of printers and the human visual system. Finding the minimal least squared error is a difficult problem to solve in two dimensions, so he suggests using an approximation. Zakhor *et al* [Zakh92] used distortion between blocks of the image as a measure and used quadratic programming to find an optimal solution. Since these techniques generally make multiple passes over the image they require much more memory and computational time and are thus, only useful for situations where these extra resources are available.

2.2.4 Image Quality Measures for Traditional Halftoning Results

To aid in the assessment of these various halftoning methods and results, different measures of quality have been developed for both qualitative and quantitative assessment. Many of these quality metrics that have been specifically designed to assess the quality of photorealistic halftoned images. Some of these metrics are based on printer or human visual system models [Grog92, Lin93, Mits93]. Other metrics used to assess the quality of halftoned images are quantitative metrics. Ulichney [Ulic87] analyzes the presence of various frequency information and symmetry of signal and noise information in the image using plots of the power spectrum, anisotropy, and exposure plots. Veryovka *et al* [Very98] measures the preservation of edges and the introduction of artifacts at multiple scales of the image. In this thesis the latter two quantitative measures will be used to evaluate importance driven halftoning results.

2.3 Special Effect Halftoning

The theme throughout traditional halftoning techniques is to hide or mask any artifacts that may inadvertently be introduced into the image. Currently, much of the work in halftoning is done on the generation of blue noise masks [Ulic88] or constructing the perfect dither arrays [Chu92, Ulic93, Mits95, Dalt95].

There is however, an area of computer graphics that focuses on introducing effects into images. In these type of images it is important, not only to ensure that these effects are preserved, but that other artifacts are not introduced. One way to make this process more efficient, is to develop a set of halftoning techniques that introduce *wanted artifacts*. There has been some initial research in this area [Sloa92, Sloa93, Ostr95b, Buch96].

2.4 Non-Photorealistic Rendering

In this section a brief overview will be given of an area of non-photorealistic rendering that is closely related to halftoning. Results from this area will be used throughout this thesis as a basis of construction and evaluation of results which resemble non-photorealistic rendering (NPR) results. This area is pen-and-ink illustration.

Pen-and-ink illustration can be closely paralleled with the process of halftoning. This type of illustration results in the creation of a binary image that has placed ink in such a manner as to create the perception of more tones. Landsdown and Schofield [Land95] reviewed NPR techniques. Three of the common effects used in NPR, particularly in pen-and-ink illustration are:

- Tone - The reproduction of various tones from the original scene or image. As Winkenbach and Salesin [Wink94] states tone refers to "the amount of visible light reflected toward the observer from a point on a surface".
- Texture - the structure implied by the placement of ink, color or "strokes". NPR often uses *stroke textures* for the simultaneous creation of both tone and texture.
- Outline - As Winkenbach and Salesin state [Wink94] realistic scenes generally do not contain definite outline, however, it is a natural means for portraying objects especially in children's drawings.

Pen-and-ink renderings create these three effects in a variety of ways. Tone is typically created by defining different *stroke textures* for varying intensities [Wink94] or by overlaying multiple strokes or textures until the desired tone is achieved [Sali94]. Tone and texture has also been achieved using multiple strokes or parametric curves [Wink96] [Elbe95] and contour lines [Pnue94]. Texture is defined by the type, thickness, orientation, placement, and quantity of strokes. In this way, texture and tone may not be independent. Outline can be created using edge detection [Sait90] and the strategic placement of skeletal strokes [Hsu94].

This emphasis on effects can be compared with the desired results of halftoning. Halftoning, as mentioned previously, is concerned with the reproduction of grey tones as is pen-and-ink illustration. Another important issue is the preservation of detail and the reduction of unwanted artifacts or texture in the halftoned image. Thus, it can be concluded that halftoning has a similar concern with the creation and placement of texture. Finally, halftoning has more recently put emphasis on the preservation of edges, which can be closely compared to NPR's emphasis on outlines.

Both rendering techniques, pen-and-ink illustration and halftoning can be generalized to place ink or display primitives both globally and locally. In pen-and-ink illustration the stroke textures and placement of strokes determine the global placement of ink, where the type of stroke and the size, width or curvature will determine the local placement of ink. Schlechtweg and Strothotte [Schl96] alter the global placement of display primitives depending on which areas of the image are more important and hence need more detail. They alter the local placement of ink by adjusting the shape or structure of the display primitive. This global and local manner of ink placement is not common among halftoning techniques, although it seems to have potential. Halftoning techniques have typically worked on a local scale, being concerned with the local placement of ink to control artifacts, but also with the global placement of ink to preserve relative tone.

2.5 Multi-resolution Image Representation and Analysis

There are many different ways to store and analyse images. Two common types or classes are the following:

- Pyramidal or MIP map representations as described by Williams [Will83]
- Wavelet Representations [Stol96]

Much research has focused on using these different image representations for various purposes. Wavelets [Mall89, Stol95a, Stol95b] can be used to hierarchically decompose the image into a coarse representation and a set of coefficients. This representation allows for the expansion of this coarse representation into details by using the wavelet coefficients. It is an efficient method of image storage at the small expense of computation time needed to expand the image details by evaluating a function over the wavelet coefficients. This representation is not used for importance driven halftoning simply because computation time is more important than storage space. Instead, a MIP map type of approach has been used. MIP maps [Will83] represent the image at multiple resolutions, by evaluating a 2×2 box filter over a 2×2 area and storing the result as a new representation of the image. For an $n \times m$ image, the amount of storage room needed is $3/2(n \times m)$. The typical MIP Map as used by Williams [Will83] uses intensity values for the values in the pyramid. In importance driven halftoning a similar structure is used, except the values are the result of a user chosen importance function evaluated over the image.

2.6 Importance Driven Halftoning

In this thesis a halftoning technique is proposed that allows the user to have control over both the global and local placement of ink. With this control the user will be able to indirectly control the tone and texture in the halftoned image for the purposes of creating both photorealistic halftoned results and non-photorealistic halftoned results. This technique has four main characteristics:

- Distribution of ink is driven by the user-defined importance function.

- The number of display primitives is user-defined.
- The user has control over the type of display primitives.
- The Multi-resolution/pyramidal representation allows for control of both global and local distribution.

Halftoning research has concentrated on preserving grey-scale reproduction and more recently preserving edges. Importance driven halftoning allows the user to specify an *importance function* which identifies image attributes that must be preserved or highlighted throughout the halftoning process. This technique adjusts the placement of ink according to the given importance function, resulting in placing ink in areas where important attributes are more prominent.

Second, this new halftoning technique allows the user to specify the number of display primitives. Restricting the number of display primitives permits creation of limited resource renderings. When using intensity for the importance function the alteration of the quantity of display primitives allows the user to alter the absolute intensity of the image (darken or lighten the entire image).

Third, the user can specify the type of display primitives to be used. Altering the type of display primitives allows the user to control the local placement of ink. For example, the local clustering of ink is very different if a line segment is used as the display primitive as opposed to a single pixel.

Finally, the pyramidal representation of the image helps in the identification of the important attributes at varying resolutions. In this manner, both a global and local assessment can be made of the number of display primitives needed in a certain area. If intensity is the importance function the low resolution representations will help to retain relative contrast across the image and the higher resolutions will help to adjust local placement of ink thus, retaining control over texture and artifacts.

Importance driven halftoning is a general technique that allows the user to directly control the number of display primitives and indirectly control the placement of them through the definition of an importance function and choice of display primitives. It attempts to preserve user-defined image attributes throughout the halftoning process by using the visual system's ability to spatially integrate and act as a multi-resolution filter.

This technique allows the user to either directly or indirectly control parameters to create both photorealistic and non-photorealistic effects. The human visual system's ability to spatially integrate will be used to create the illusion of grey tones. Also, the ink will be distributed from a global representation to a local representation in an attempt to preserve important attributes (possibly edges or fine detail) while avoiding the introduction of regular artifacts. Since this is a generalized technique controlled mostly by user-defined parameters, there are many effects that can be created. In the following chapters, details of the technique are discussed and a wide range of results are presented.

2.7 Chapter Summary

In this chapter, an overview of the basic attributes of the human visual perceptual system, current halftoning techniques and non-photorealistic rendering techniques were presented. First, the overview of the human visual system outlined basic characteristics such as ability of the visual system to spatially integrate and perform as a passband filter, the visual system's dependency on background illumination for the perception of intensity and it's preference for visual organization and clarity. This basic knowledge of the human visual system was then applied to characteristics of continuous tone, grey-scale images that are needed for the formation of a particular perception. Second, a comprehensive discussion of current halftoning techniques was given including methods designed to preserve intensity, to preserve edges and to create optimally halftoned images according to some measure. A brief overview of some of the current measures used to evaluate these halftoned results was also presented. Third, an overview of one method of non-photorealistic rendering was presented. A comparison of some of the similarities between this method (pen-and-ink illustration) and halftoning was discussed. Finally, an overview of multi-resolution image representations and an outline of importance driven halftoning was presented. In the following chapters, the technical details of a new technique called *Importance Driven Halftoning* will be introduced. This technique can produce various results according to the user's desires. A collection of results, the parameters used to obtain them and how these results compare with other current results will be presented in chapters 4-6. Some of the results in chapters 4 and 5 will be compared to the results of pen-and-ink illustration techniques.

Chapter 3

Importance Driven Halftoning

In this chapter, the details of and the motivation for a new halftoning technique based on the preservation of important image attributes are outlined. This technique allows the user to specify an *importance function* that defines the important image attributes, the quantity of display primitives, and the type of display primitives used to halftone the image. Importance driven halftoning allows for both global approximation of the importance values as well as local approximation and control. Having control over the local placement of display primitives allows for the creation of many different effects or textures, while retaining the important attributes in the image.

This chapter will outline the problem and motivation for the solution, as well as discuss the details of the components of the system, user control over parameters, and use of this information in the halftoning process. This chapter will first outline the problem and motivation for importance driven halftoning. Second, it will give an overview of the technique describing the ability to control both global and local arrangements of ink. Third, it discusses the technical details of constructing the multi-resolution representation of the image, distributing the weights and display primitives, and adjusting the local structure of the display primitives. Finally, the user controlled parameters are presented.

3.1 Problem Formulation/Motivation

This technique is proposed mainly as a generalized halftoning technique. Previous halftoning techniques are formed on the premise that a particular image attribute is important and thus needs to be preserved throughout the halftoning process. Standard rules for identifying what is important in our interpretation of the image have not yet been developed. Some basic image components seem to be more important than others. However, once these components are combined it is undetermined which are most important and which can be neglected. As research in this area progresses attributes other than the commonly accepted grey tone reproduction and edge preservation may be identified. Importance driven halftoning is created to use any important attribute or combination of important attributes the user determines necessary.

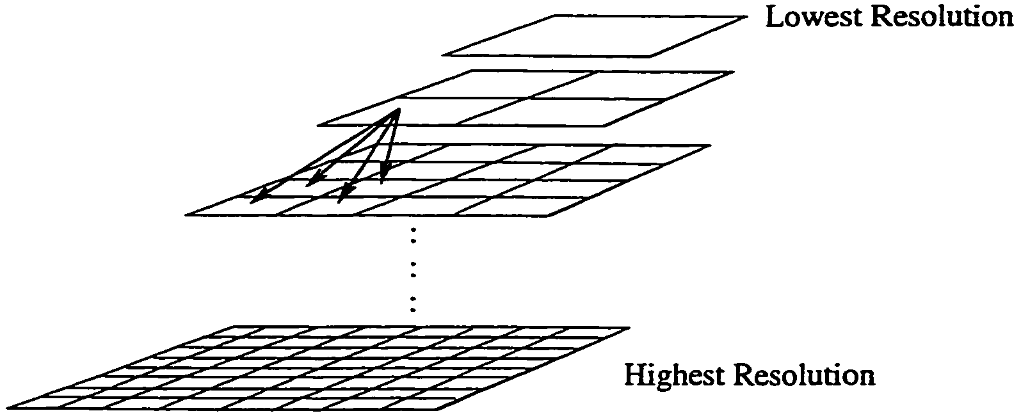


Figure 3.1: The lowpass pyramid is a filtered representation of values related to the image. The highest resolution of the pyramid are values obtained from evaluating the importance function over the original image. The rest of the pyramid is a filtered representation of this highest resolution level with the highest resolution at the bottom of the pyramid and the lowest resolution (single value) at the top of the pyramid. The display primitives are distributed in a top-down manner through the pyramid according to these values. The display primitives at each region in the level above are distributed to four corresponding regions in the level below.

As a secondary objective, this technique must be able to analyse the image on both a global and local scale in order to determine the placement of ink. The human visual system is known to form an initial global interpretation of the image before analyzing the image on a local scale. Due to this behavior, it must ensured that both the global and local portrayal of the image does not result in false interpretations. In order to ensure this, the proposed technique uses a multi-resolution representation of the image for determining the global and local placement of display primitives during the halftoning process. By being able to assess the image at a global level, relative importance of the image attributes can be preserved. This means two areas of the image that have the same importance must be rendered with the same number of display primitives. However, being able to assess the image at a local level allows the local adjustment of these display primitives for the creation of various effects while still preserving relative importance throughout the image.

Altering the number and type of display primitives can lead to differing effects. These effects include photorealistic halftoned, non-photorealistic halftoned and limited resource rendered images. Once again, a photorealistic halftoned image is an image halftoned in a manner to appear as much like the original continuous tone grey-scale image as possible. A non-photorealistic halftoned image is an image halftoned in a manner to appear similar to the original continuous tone grey-scale image, but with added texture or artistic effects perhaps creating the illusion of a sketched or drawn image. Finally, limited resource rendering uses a reduced amount of resources to reproduce the image. This may involve eliminating attributes of the image that are of lesser importance. In halftoning, reducing resources typically means the reduction of the number of display primitives or the amount of ink used. Thus, the type and number of display primitives used depends on the desired result. Examples of all three types of effects will be illustrated in chapter 4.

3.2 Overview of the Technique

Importance driven halftoning uses a pyramidal representation of the image, as shown in figure 3.1. The bottom level of the pyramid is the evaluation of the importance function over the original image. The other levels are filtered versions of the bottom level to form lower resolution representations. Once the pyramid is constructed the display primitives are distributed down through this *lowpass pyramid* [Perl95] according to the values in the pyramid, until a higher resolution representation of the image is reached. Once the desired resolution is reached the placement of the display primitives can be adjusted only within a local area. Importance driven halftoning uses the continuous tone grey-scale image in addition to the three user defined parameters as input and creates a binary image as output. The three user defined parameters are:

1. The importance function.
2. The number of display primitives.
3. The type of display primitive.

These three parameters are defined at the beginning of the halftoning process and the binary image is rendered automatically from them.

The chosen importance function must be defined to rate a region of the image for the presence of the desired component. Thus it must return a high value if the region is very important and a lower value if the region is of less importance. The values from evaluating the importance function over the original image become the lowest level (highest resolution - see figure 3.1) of the *lowpass pyramid*. This representation of the image is filtered using a box filter to create lower resolution representations of the image that form the upper levels of the pyramid. Once a single value representation of the image is created, the lowpass pyramid is complete and ready to be used for the distribution of display primitives.

Using the lowpass pyramid the user-defined quantity and type of display primitives are distributed down through the pyramid from the top to the bottom. The display primitives are divided into different regions in the pyramid according to the values of each region. Thus, the general placement of display primitives is defined by the importance function. Once the display primitives have been globally distributed, the local placement or arrangement of the ink is controlled by the type of display primitive or by re-orienting or scaling the display primitive.

3.2.1 Prioritized Global Distribution

Using the lowpass pyramid for the distribution of display primitives, places the display primitives in areas where the importance function evaluates to larger values. Thus, more display primitives are placed in areas where important image attributes are more prominent. This distribution of display primitives assumes that more display primitives generally result in a better rendering of the important attributes, particularly, tone, edges, detail, and areas of high contrast. Of course, excessive amounts of display primitives can lead to destruction of these image attributes. The justification for this assumption can be made by considering current halftoning techniques as a basis. Current halftoning techniques simply address the problem of local arrangement of ink or display primitives. The number of display primitives used by current techniques is bounded only by the number of pixels in the image. If the number of display primitives is limited, then both the problems of where to place the display primitives as well as how to locally arrange them must be addressed. This is the case with limited resource rendering. The placement of more display primitives in areas where important attributes are more prominent is common in both the graphics and arts worlds. A common example of placing more display primitives in areas of important attributes is a typical sketch. A sketch is obviously intended to use minimal amount of resources and yet still portray the objects or features of the image. Thus, display primitives or strokes are placed in areas of the image to delineate the features of the image.

3.2.2 Local Structure and Positioning

Once the position of the display primitive is determined within a small local region, the display primitive and its immediate neighbors can be adjusted to ensure the portrayal of the intended local effect. In creating photorealistic halftoned images it is important to ensure that these local adjustments preserve patterns (*i.e.* edges or high contrast regions) present in the original image and do not create regular patterns not present in the original image. There are many ways to adjust the local placement of display primitives without disrupting the global positioning and quantity of display primitives. Some efforts that result in local adjustment are the following:

1. Altering the type of display primitive.
2. Rotating the display primitive.
3. Scaling the display primitive.
4. Thresholding the display primitive.

Depending on the type of display primitive some or all of the above can have a profound effect on the halftoned result.

Varying the type of display primitive alters the local placement of ink. This can lead to creating varying texture effects in the image. For example using a line segment as opposed to a square or matrix of the same number of pixels introduces different textures into the image. This is due to the clustering of the ink in a different manner. Finally, by altering the orientation or the size of the display primitive regular or irregular patterns of textures can be created. In figure 3.2 the result of using a single pixel versus a line segment as the display primitive can be compared with the original image. Both halftoned versions use approximately the same number of black pixels, they are simply distributed differently.

Figure 3.2: From top to bottom: (a) Original grey-scale image - 600dpi; (b) Importance Driven Halftoning using single pixel as display primitive - 200dpi; (c) Importance Driven Halftoning using a line segment oriented and scaled as a function of intensity - 200dpi. Both images in (b) and (c) set approximately 70,000 black pixels

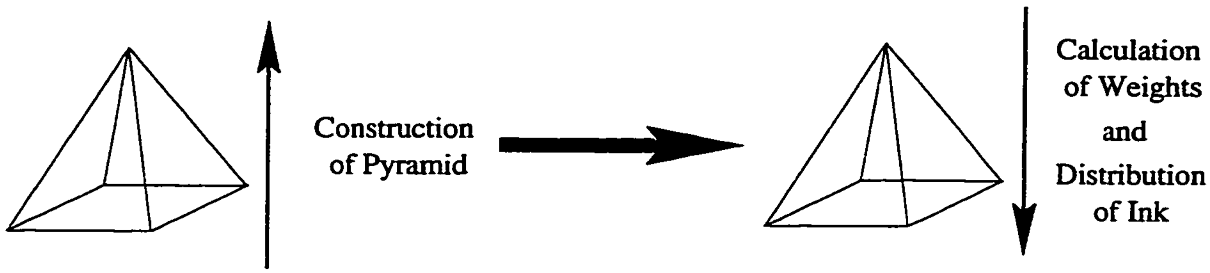


Figure 3.3: The process of importance driven halftoning involves constructing the pyramid from bottom to top to form the multi-resolution layers. The next step of the process involves calculating weights according to the importance function values in order to distribute the ink in a top-down manner.

Having the user control the importance function, the number of display primitives, and the type of display primitives allows for the creation of various effects. Using the lowpass pyramid or multi-resolution representation enables the user to indirectly control both the global and local distribution of display primitives. These coupled abilities allow the user to create a variety of effects on both a global and local scale by using different arrangements of ink.

3.3 Technical Details

Importance driven halftoning is essentially a three step process. The steps are:

1. Evaluation of the importance function and construction of the pyramid.

2. Calculation of weights and distribution of display primitives.

3. Local adjustment to placement of ink.

The pyramid is constructed from the highest resolution representation up to the lowest and the ink is distributed in the reverse order as shown in figure 3.3. Once the distribution of ink or display primitives reaches a high resolution, local adjustments can be made to the actual placement of ink. In this section, these three steps to importance driven halftoning will be discussed in detail.

3.3.1 Construction of the Lowpass Pyramid

As mentioned previously the premise for the creation of the lowpass pyramid is to have the ability to manipulate and analyse the image more as the human visual system perceives it - at multiple resolutions. Since the human visual system acts as a Fourier analyzer an image tends to be perceived at varying resolutions. In order to achieve the preservation of important image attributes, the human visual system must perceive the preservation of these attributes, at all resolutions.

The lowpass pyramid is constructed using two functions: The importance function and the filter. Both of these functions are static throughout the formation of the pyramid. The highest resolution (lowest level) of the lowpass pyramid is a matrix of values obtained after evaluating the importance function over the original image. This two dimensional matrix of values forms an "important attribute" representation of the image. Higher values in this matrix represent areas where important attributes are more prominent in the image. Details of the importance function will be discussed further in section 3.4.1.

The upper levels of the pyramid are formed by filtering this highest resolution representation of the image. Each level is formed by applying the filter to the level below. A standard non-overlapping box filter was used that simply averages four corresponding values in the level below to form a single value in the current level. This filtering approach is identical to that used in the formation of MIP maps by Williams [Will83]. For example, consider an image of resolution $2^p \times 2^p$ with relative or average intensity as the importance function. To construct the representation with resolution $2^{p-1} \times 2^{p-1}$ a box filter is passed over the $2^p \times 2^p$ representation that averages the intensities of four pixels to construct the value for one pixel at the level above (see Figure 3.1).

Compensating for box filter requirements

Using a box filter can present problems if the original image is not square or not a power of two. If the original $n \times m$ image does not meet these requirements, the image representation used is a $2^p \times 2^p$ image, where $p = \lceil log_2(max(m, n)) \rceil$. The edges of the image are padded with values that do not affect the distribution of display primitives in order to fill the excess area. This means the original $n \times m$ image is positioned in the center of the $2^p \times 2^p$ representation and the edges are filled in with the lower bound of the importance function as shown in figure 3.4. This buffer zone around the image can also be useful in situations where a display primitive may be rotated off the edge of the image.

An alternate method to applying these two functions for the creation of the lowpass pyramid is to filter the original image and then evaluate the importance function over these filtered values. This has the effect of actually changing the evaluation of the importance function. Applying the

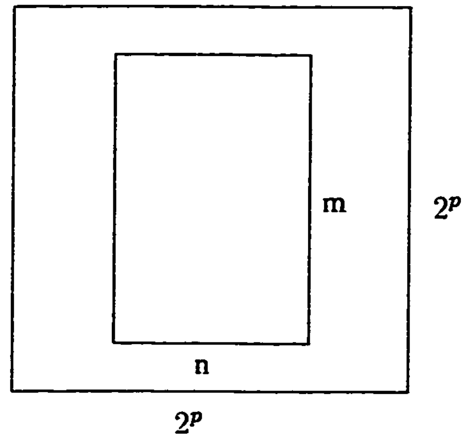


Figure 3.4: If the image size is not a power of two, an image representation that is a power of two is used. The difference in the sizes is padded with values that do not affect the distribution of display primitives.

functions in this order could actually result in a different rating of importance from what the user had intended. This effectively alters the importance function and thus, this approach was not used.

Finally, experiments were done using other non-uniform or overlapping filters such as Bartlett or Gaussian filters, however these approaches also lead to an alteration of the intended importance function. Using non-uniform filters alters the importance function to place more emphasis on different regions under the filter (*i.e.* center is more heavily weighted with the Gaussian filter). Similar alterations of importance occur with over-lapping filters. Realizing this distortion of importance function values, the non-overlapping box filter was chosen to be the best filter for the creation of the lowpass pyramid.

Once the lowpass pyramid has been constructed it is traversed in a top-down depth first search manner. As the pyramid is traversed, values in neighboring regions of each representation are compared and weights are calculated for each region.

3.3.2 Calculation of the Weights

The calculated weights determine how much of the total number of display primitives from the current level will be propagated to each of the four corresponding regions in the level below. Each region is weighted as a fraction of the sum of itself and it's three neighboring region values. Thus, the weights for the four approximate regions will always sum to one. Weighting each region in this manner makes the process of calculating weights independent of the importance function. This is needed for the creation of a generic halftoning technique. More specifically, if the importance function is relative intensity each region is weighted as a fraction of the summed intensity of the four corresponding intensity values. This identifies each region as having a percentage of the total display primitives passed from the representation above. Thus, if a region in the representation above

	R_1	R_2	R_3	R_4
Importance Func. Value	0.8	0.4	0.2	0.6
Corresponding Weight	0.4	0.2	0.1	0.3
	40%	20%	10%	30%

Table 3.1: Table of importance function values and corresponding weight values

corresponds to regions R_1, R_2, R_3, R_4 in the current representation with the importance function values as shown in table 3.1, these regions would have the corresponding weights shown in table 3.1.

By using relative intensity as the importance function and constructing the weights in this manner the ink will be concentrated in darker regions of the image. By representing the weights as a percentage of the total display primitives the intensities are kept relative to each other and independent of the number of display primitives used to halftone the image.

3.3.3 Distribution of the Display Primitives

The distribution of display primitives within the pyramid is determined by the weights. Distribution begins at the top of the lowpass pyramid with some initial amount of display primitives and proceeds downwards through the pyramid. This ensures that the display primitives are distributed evenly at a global level rather than replicating a regular pattern of local distribution. This process is less likely to introduce regular artifacts into the image. The quantity of display primitives for each location at the current level is divided into the four regions R_1, R_2, R_3, R_4 at the next level. The display primitives are distributed proportionally to the weights at each of these four locations. Thus the number of display primitives distributed to each location i is:

$$N'_i = trunc(w_i \times N)$$

where, N'_i is the number of display primitives at location i in the level below, w_i is the weighting or percentage at location i in the level below, N is the number of display primitives in the current level, and $0 \leq i < 4$. However, since the number of display primitives is a integer amount and the weight is a percentage, an error is introduced. Since the operation truncates the result of the weight factor, there may be extra display primitives that have not been distributed. The extra display primitives are calculated as follows:

$$extra = N - \sum_{i=0}^{i<4} N'_i$$

where *extra* is the number of extra display primitives for the level below. There are two methods for distributing these extra display primitives depending on the underlying values. The first method of distribution is used for regions of equal value and the second for the rest of the regions. The first method distributes the extra display primitives evenly among the regions. The second method calculates the error for each of the i regions as follows:

$$error_i = (w_i \times N) - (N'_i + t_i)$$

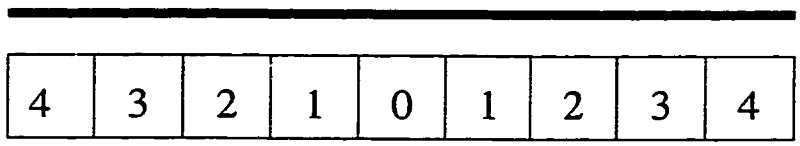


Figure 3.5: A line nine pixels long is thresholded according to a threshold matrix aligned with the line. The thresholds start from the center of the line and increase outwards. The line is thresholded from the center out because the center of the line is known to be dark, however other portions may not be.

where, $error_i$ is the absolute error between the approximation N'_i and the actual number of display primitives needed at location i at the level below, and t_i is the number of extra display primitives allocated to location i. The value t_i has an initial value of zero. The total number of extra display primitives are distributed according to the amount of error at each region, $error_i$. The region with the largest positive error is determined, an extra display primitive is allocated to that area and the value of t for that region is incremented. The number of extra display primitives is then decremented. Next, the errors at each of the four regions are recalculated. Again, one more display primitive is then allocated to the region with the largest positive error and the values of t and $extra$ are adjusted. This process is repeated until the number of extra display primitives is zero.

The display primitives are distributed in a top-down order. At each level the region with the highest value of the importance function is selected until the highest resolution is reached (depth-first traversal). In the case where the importance function is average intensity, the display primitives are distributed to the overall darkest regions of the image first.

Once the display primitives have been propagated through the pyramid to a smaller area of interest (a higher resolution) the ink distribution can be locally adjusted to the underlying image. This ensures that a valid local approximation can be obtained. Some methods of local ink adjustment will be discussed in the following subsection.

3.3.4 Local Adjustment of Ink

Local adjustment to the placement of ink has been investigated through three modifications to the display primitives.

1. Rotating the display primitive.

2. Scaling the display primitive.

3. Thresholding the display primitive.

With a single pixel as the display primitive the first two modifications do not affect the results. However, a small region of pixels can be thresholded. Using intensity as the importance function, the image can be thresholded using a threshold matrix. Using a matrix or larger area for the approximation compensates for the human visual system's inability to distinguish fine detail at increasing distances (ability to spatially integrate). This adjustment of display primitives would involve distributing the display primitives down to the highest resolution and thresholding this region of the image according to the corresponding threshold matrix values. In the case where intensity is the importance function, this yields results similar to traditional threshold matrix halftoning. The details and results of this modification will be discussed in chapter 4.

With a line segment as the display primitive all of the above modifications are possible. Rotating a set of line segments in an image can create very different results. Aligning the line segments uniformly or semi-uniformly can create the interpretation of distinct texture in the image. Aligning the line segments along the edges or regions of high gradient can help to enhance edges and connectedness in the image and create results similar to outlines in non-photorealistic rendering. The line segment can be aligned with various image properties such as intensity, gradient, variance etc. Scaling the line segment not only can reduce the amount of resources used, but can vary the texture throughout the image. Again the line segment can be scaled to intensity, gradient, variance or any other image property. Finally, thresholding the line segment has a similar effect to scaling it, but the length can be adjusted not only to a particular image property, but also to the image intensity directly under the placement of the line segment. The determined position of the display primitive is where the center of the line is placed. Thus, in order to threshold the line a threshold matrix is built from the center of the line outwards as shown in figure 3.5. The pixels of the line are thresholded according to this matrix, as is done with normal threshold matrix halftoning. This adjusts the length of the line segment directly to the underlying image. Thresholding the line segment will be shown to preserve detail in the image. By using these three simple modifications to the display primitives, importance driven halftoning can create different local effects, which can be used to create global effects depending on the amount of local adjustment.

3.4 Alteration of Parameters

In importance driven halftoning, the user has the ability to define three parameters for the creation of various effects. These three parameters are:

1. The importance function.

2. The number of display primitives.

3. The type of display primitives.

This section outlines some of the values used for these parameters and the general effect they have on the halftoned image.

3.4.1 Importance Function

The user defines the importance function which dictates the distribution of display primitives. As intensity and spatial proximity are the only information available in the original grey-scale image, the importance function must be based on this information. The importance function can be any function of intensity, variance, gradient, edge detection filter (sobel) or even a function identifying texture in the image. If average intensity is the importance function the highest resolution representation is simply the original image's intensity values. The importance function identifies the presence of important image attributes over a segment of the image. Thus, if the importance function is intensity g at location (x, y), then the importance function F, is simply:

$$F(x, y) = g(x, y)$$

The user may also choose to construct an importance function that combines several importance functions by weighting them as follows:

$$F(x, y) = w_1 \times f_1(x, y) + w_2 \times f_2(x, y) + ... + w_n \times f_n(x, y)$$

where w_x is a weight, $\sum_{x=0}^{x=n} w_x = 1$ and $f_1, ..., f_n$ are importance functions. For example, the following combines the importance functions of intensity and variance:

$$F(x, y) = 0.8 \times \text{Intensity Fcn(x,y)} + 0.2 \times \text{Variance Fcn(x,y)}$$

Altering the importance function or combination of importance functions, defines the placement of the display primitives based on different attributes in the image. The results of altering the importance function will be illustrated in the following chapter.

3.4.2 Number of Display Primitives

The user can specify the number of display primitives used in importance driven halftoning. Adjusting this parameter, especially in combination with the other parameters can create differing effects. By reducing the number of display primitives only the most important attributes of the image will be rendered by the placement of display primitives. This can be useful in limited resource rendered images. Using an average number of display primitives,

$$\text{Average No Of Display Primitives} = \frac{\text{Average Intensity} \times \text{No Of Pixels In Image}}{\text{No Of Pixels In Primitive} \times \text{Maximum Intensity}} \tag{3.1}$$

yields results similar to traditional halftoning, especially in the case where intensity is the importance function. Figure 3.6(a) is an image halftoned using Floyd-Steinberg's error diffusion with approximately the average number of display primitives as calculated in equation 3.1. Figures 3.6(b) to (d) are images halftoned using importance driven halftoning. Figure 3.6(b) is halftoned with half the number of display primitives as (a). Figure 3.6(c) has the same number of display primitives as (a) and figure 3.6(d) has twice the number of display primitives as (a).

3.4.3 Type of Display Primitives

The user can also specify the type of display primitive. Altering the type of display primitive adjusts the fine local placement of ink. The display primitive defines how the ink is clustered locally. By

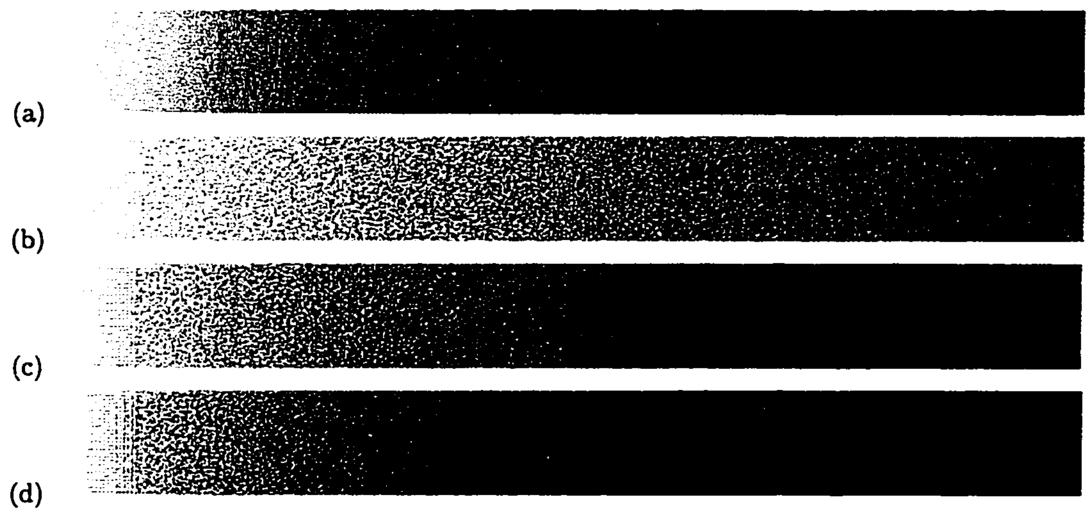


Figure 3.6: Top to Bottom: (a) Floyd-Steinberg's error diffusion using approximately an average number of display primitives. (b) Importance driven halftoning with half the display primitives as in (a). (c) Importance driven halftoning with the same number of display primitives as (a). (d) Importance driven halftoning with twice the number of display primitives as in (a).

using a single pixel for the display primitive results similar to traditional halftoning can be attained. Using a line segment as the display primitive can create sketched-like effects. By combining the alteration of display primitives with the importance function other non-photorealistic effects can be created. In the following chapters both a single pixel and a line segment will be used as the display primitive to illustrate various effects. As an example, the differences in textures between a single pixel and a line segment are shown in figures 3.2(b) and 3.2(c).

3.5 Chapter Summary

In this chapter the details and motivation of a new halftoning technique, called importance driven halftoning, have been outlined. Current halftoning techniques lack the ability to easily adjust which attributes are preserved in the image, how many display primitives are used, and the local structure or arrangement of ink. This new technique is based on the preservation of important image attributes, defined by the user. The user has control over the amount and type of display primitives used in this halftoning process. The user parameters ultimately give the user control over both global and local placement of display primitives and ink. Importance driven halftoning uses a lowpass pyramid constructed of values obtained from evaluating the importance function over the image and filtering it to multiple resolutions. From this pyramid of values, weights are determined that define the distribution of display primitives. Finally, the individual display primitives can be rotated, scaled and thresholded to create local texture or effects. Altering the importance function, the number of display primitives and the type of display primitives yields varying results that can be used to achieve traditional photorealistic halftoning objectives, limited resource rendering objectives or non-photorealistic rendering objectives. This chapter has outlined the components used in this technique, the process of importance driven halftoning, the details of the parameters, and a brief explanation of the results. The following chapters will illustrate the use of various parameters in importance driven halftoning, the details of how they were obtained, and a comparison with other current techniques for obtaining similar results. The comparison will be performed using Ulichney's [Ulic87] measures and measures proposed by Veryovka *et al* [Very98].

Chapter 4

Parameter Alteration in Importance Driven Halftoning

This chapter illustrates the results obtained by altering the parameters of importance driven halftoning. Importance driven halftoning is controlled by three parameters: the importance function, the quantity of display primitives, and the type of display primitives, that can be varied and altered in different combinations to obtain both photorealistic and non-photorealistic results. As mentioned in chapter 2, the main objectives of halftoning are to approximate grey-scale intensities and to preserve edges or detail present in the original image without introducing unwanted artifacts. During the construction of images different interpretations are induced by using various local and global effects. In importance driven halftoning, the user can create various local and global effects in order to suggest different interpretations by varying these three parameters. The first three sections of this chapter illustrate images created from altering each of the three parameters individually. The fourth section shows how the parameters can be altered simultaneously to create other results.

4.1 Original Images

The images shown in figure 4.1 will be used as the original continuous tone grey-scale images for halftoning examples in this chapter. The top two images are computer generated grey-scale ramps. The bottom left image is a grey-scale photograph of the ubiquitous Lenna and the bottom right image is a lithograph of a sheep herder by Peter Hurd [Hurd68].

4.2 The Importance Function

One of the parameters the user has control over is the importance function. As mentioned in chapter 3 the importance function identifies attributes or regions of the image the user has designated as important. Thus, an increase in the value of the importance function must correlate directly with a proportional increase in the importance of the underlying region of the image. Importance driven halftoning assumes that important image attributes are better rendered with more display primitives

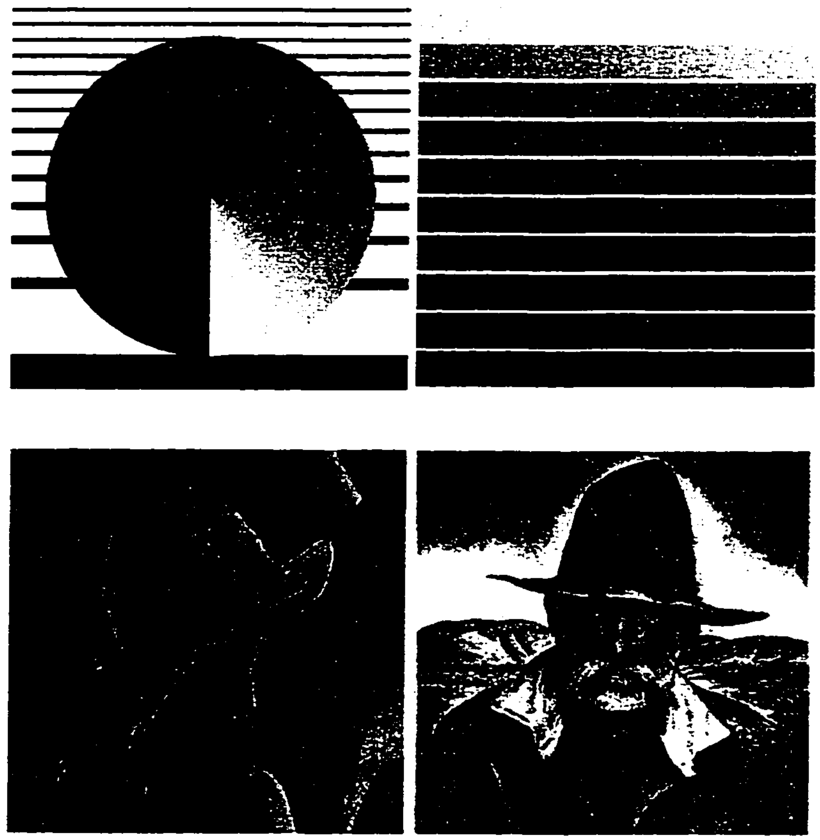


Figure 4.1: Original images used in this chapter. The top two images are computer generated continuous tone grey-scale images. The bottom left image is a grey-scale photograph of the ubiquitous Lenna and the bottom right image is a lithograph of a sheep herder by Peter Hurd [Hurd68].

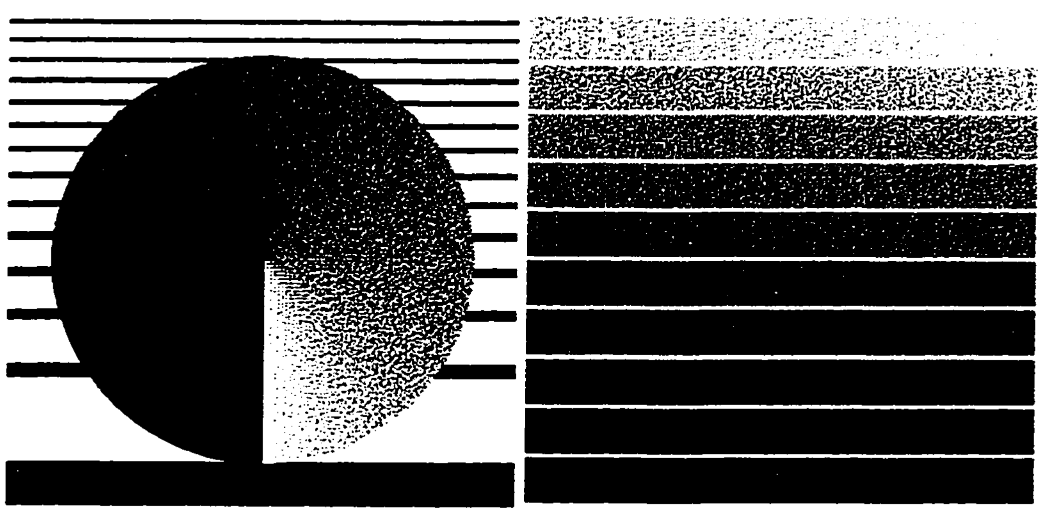


Figure 4.2: Importance function is intensity. The function returns higher values in regions of darker intensity. Hence, more display primitives are placed in these regions. The display primitive is a single black pixel and the number of display primitives is the average number for these images as calculated by equation 4.1.

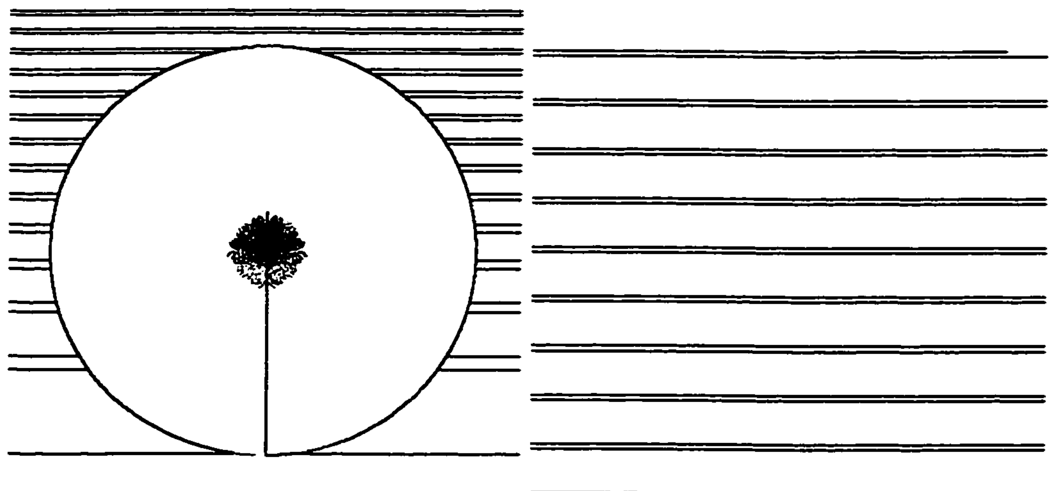


Figure 4.3: Importance function is variance. The function returns higher values in regions of higher variance (*i.e.* larger intensity difference between a pixel and its eight neighbors). Hence more display primitives are placed on edges, or differences in intensity.

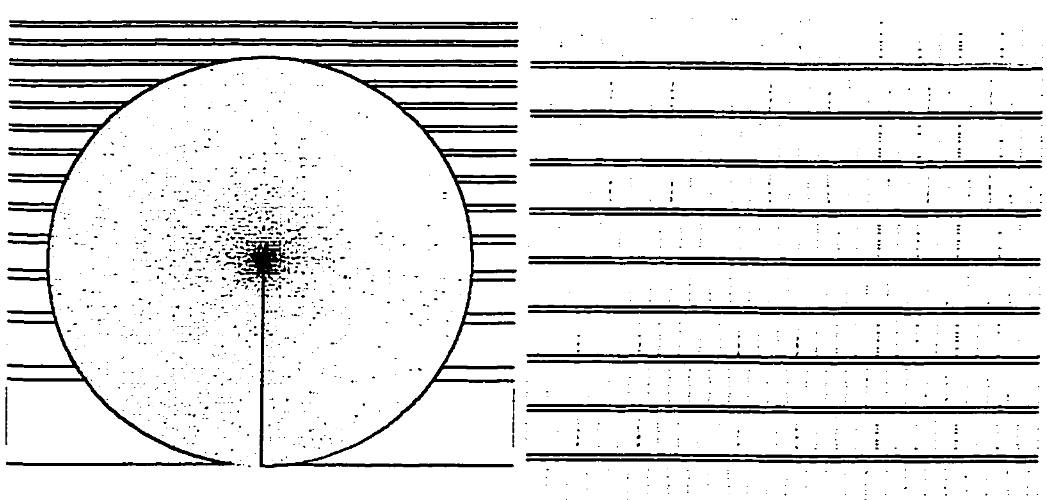


Figure 4.4: Importance function is gradient. The function returns higher values in regions of higher gradient magnitude. Hence more display primitives are placed in regions of higher gradient magnitude.

than with less. The importance function defines both the global and the local distribution of display primitives throughout the halftoning process and consequently more display primitives are placed in regions with higher importance function values. Finally, the importance function can only be a function of grey-scale intensity and spatial proximity since the halftoning process only uses a two dimensional continuous tone grey-scale image as input.

4.2.1 Intensity

In traditional halftoning the primary objective has been the approximation of grey-scale intensities using black ink on white paper. Thus, intensity may be a natural choice for an important image attribute. In our first example (figure 4.2), the importance attribute is intensity g, at location (x, y). The importance function F at location (x, y) is:

$$F(x, y) = g(x, y)$$

Thus for an 8 bit grey-scale image the values of the importance function range from 0 to 255. A value of zero represents white and a value of 255 represents black. Hence, more display primitives will be placed in regions with larger intensity values (darker regions).

4.2.2 Contrast

A second common objective in traditional halftoning is the preservation of edges or contrast. Edges can be determined by finding regions of high contrast [Buch95]. Regions of high or low contrast can be located by analysis of differences in intensity. This can be done in various ways, including finding the maximum difference in intensity between a pixel and it's eight neighbors, calculating the gradient or calculating the variance. As an example the importance function F, can be variance, and is calculated as follows:

$$F(x, y) = \frac{\sum_{i=-1}^{i=+1} \sum_{j=-1}^{j=+1} |g(x, y) - g(x+i, y+j)|}{\text{Num Valid Neighbours}}$$

This sets the importance function to be the normalized sum of the absolute difference in intensities between a pixel and it's immediate eight-connected neighbors. As shown in figure 4.3 more ink is placed in regions of higher variance.

Another method for detecting edges or changes in contrast is using the strength of the gradient at a particular location. The gradient is a vector which points in the direction of maximum rate of change in intensity and is defined as:

$$\nabla f(x, y) = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

The partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ can be calculated in many ways. Gonzalez and Woods [Gonz92] claim that using sobel operators provide both a smoothing and differencing effect. Thus, from the

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

(a)

-1	-2	-1
0	0	0
1	2	1

(b)

-1	0	1
-2	0	2
-1	0	1

(c)

Figure 4.5: (a) 3x3 image region; (b) mask used to compute G_x at center point of 3x3 region; (c) mask used to compute G_y at center point of 3x3 region.

masks shown in figure 4.5 the partial derivatives are calculated as follows:

$$G_x = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)$$

and

$$G_y = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)$$

and the magnitude of the gradient is:

$$mag(\nabla f(x,y)) = |G_x{}^2 + G_y{}^2|^{1/2}$$

Figure 4.4 shows results of using the magnitude of the gradient as the importance function. The magnitude of the gradient can also be approximated as:

$$mag(\nabla f(x,y)) \approx |G_x| + |G_y|$$

The images shown in figures 4.3 and 4.4 are very similar because the importance functions used are both functions of the differences in intensity over a 3x3 region. Both are useful for edge highlighting as shown in figures 4.3 and 4.4.

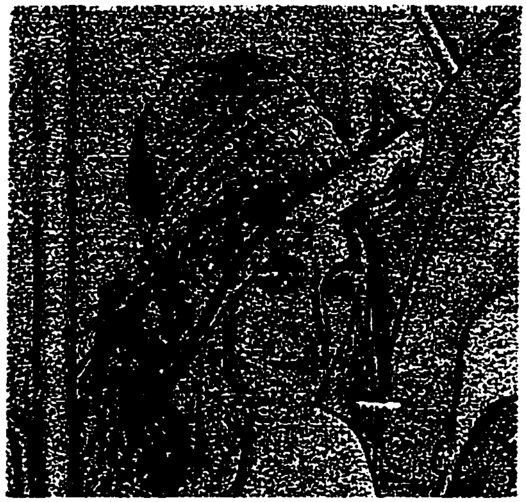
4.2.3 Combination of Importance Functions

As mentioned in chapter 3 the user can decide to define the importance function as a weighted combination of individual importance functions, calculated as follows:

$$F(x,y) = w_1 * f_1(x,y) + w_2 * f_2(x,y) + ... + w_n * f_n(x,y)$$

where $w_1, w_2, ..., w_n$ are weights where $\sum_{x=1}^{x=n} w_x = 1$ and $f_1, f_2, ..., f_n$ are importance functions. In figure 4.6(c) both variance (shown in figure 4.6(b)) and intensity (shown in figure 4.6(a)) have been

(a)



(b)

(c)

Figure 4.6: (a) Importance function is intensity; (b) Importance Function is pixel-to-pixel variance over eight neighbors; (c) Importance Function is intensity (70%) and variance (30%). The display primitive is a single black pixel and the number of display primitives is the average number for this image as calculated by equation 4.1.

combined to form the resulting importance function F, at location (x, y):

$$F(x,y) = 0.7 * \text{Intensity Fcn}(x,y) + 0.3 * \text{Variance Fcn}(x,y)$$

or

$$F(x,y) = 0.7 * g(x,y) + 0.3 * mag(\nabla f(x,y))$$

In this manner, any number of arbitrary importance functions can be combined to form a single importance function. Weighting the individual importance functions reflect the order of the importance among the different image attributes.

4.3 Number of Display Primitives

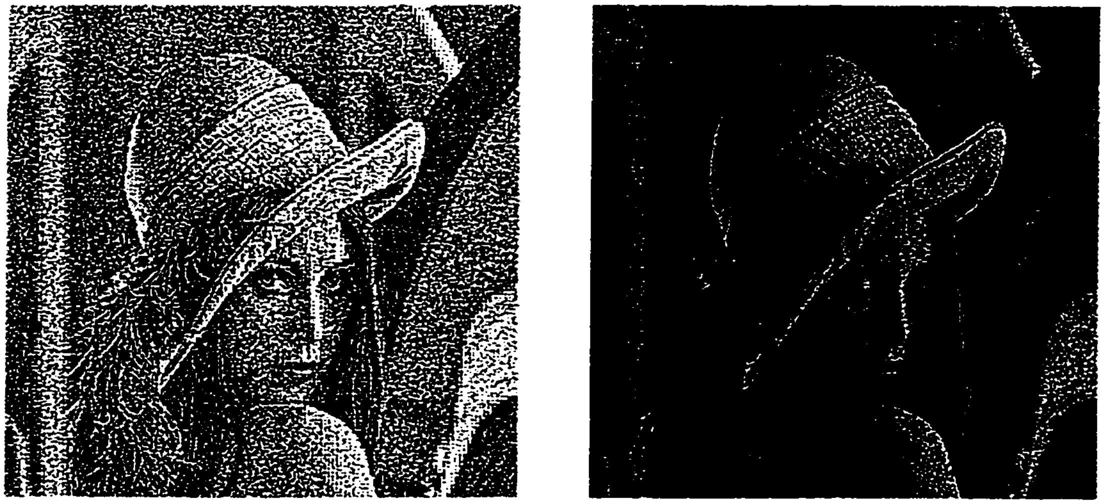
As a second parameter in importance driven halftoning, the user is able to control the number of display primitives. By varying the number of display primitives the user is able to control how many of the attributes (according to importance) are rendered. The user can specify an average number of display primitives, similar to the amount used in traditional halftoning methods, to obtain results similar to traditional halftoning. If the user specifies less than this average quantity the results are similar to limited resource renderings. Finally, if the user specifies more than the average number, this can lead to over-enhancement or even distortion (by the addition of too much ink) of the important attributes in the image. Specifying an excessive number of display primitives can be used for placing emphasis on features or for the creation of alternate, possibly artistic results.

4.3.1 Average Number of Display Primitives

The user can choose an average number of display primitives. The average number is usually calculated based on intensity as follows (see section 3.4.2):

$$\text{Average No Display Primitives} = \frac{\text{Average Intensity} \times \text{No Of Pixels In Image}}{\text{No Of Pixels In Primitive} \times \text{Maximum Intensity}} \qquad (4.1)$$

This calculation assumes black is the Maximum Intensity and black pixels are set. Importance driven halftoning ensures that using intensity as the importance function will preserve relative intensity in the halftoned image. This is ensured, because each region is a weighted ratio of four regions. This means, that general contrast in the image is preserved or enhanced, but the absolute intensities may be different. Thus, depending on the number of display primitives the entire image may appear darker or lighter than the original. However, by using the average number of display primitives (black pixels) the average absolute intensity is closely approximated, thus keeping intensities relative and absolute. The image in figure 4.6(a) uses intensity as the importance function and an average number of display primitives (single pixels) are set.



(a) (b)

Figure 4.7: (a) Image halftoned using 50% of the average number of display primitives; (b) Image halftoned using 150% of the average number display primitives. Both images have intensity as the importance function and single black pixels as the display primitives.

(a) (b)

Figure 4.8: Both images have 25% of the average number of display primitives needed. The importance function is gradient. The display primitive is a single black pixel.

4.3.2 Insufficient or Excessive Number of Display Primitives

By specifying less than the average amount of display primitives importance driven halftoning can be used to exclusively render regions of the image with higher importance values. With intensity as the importance function, relative intensity will be preserved throughout the image. This means that contrast or differences in intensity will be relative to both the number of display primitives and to the other regions of the image. Thus, if there is a 10% difference in intensity between two regions there will be a 10% difference in the number of display primitives used to render these two regions. This means decreasing the number of display primitives below the average amount will result not only in a decrease in the overall intensity, but also in a decrease in perceived contrast. This is illustrated in comparing figure 4.6(a) with the images in figure 4.7. Figure 4.7(a) uses 50% of the average number of display primitives shown in figure 4.6(a). Figure 4.7(b) uses 150% of the average number of display primitives in figure 4.6(a). Altering the number of display primitives causes the entire image to appear lighter or darker than both the original image and the image in figure 4.6(a). However, both images in Figure 4.7 retain relative intensitv proportional to the number of display primitives available, so that contrast is preserved for the interpretation of different image components.

Draft versions of the image can be created by altering the importance function and restricting the number of display primitives, for example using gradient for the importance function can preserve high frequency or edge information. In figure 4.8 both images use a quarter of the display primitives needed for each image as calculated by equation 4.1.

Increasing the number of display primitives beyond the average amount specified by equation 4.1 renders all areas of the image with a proportional increase in the number of display primitives. Thus, if the importance function is intensity, and black is the maximum intensity value, the image becomes darker as shown in figure 4.7(b). This enhances contrast possibly at the expense of detail in the image depending on the excessiveness of display primitives. Specifying an excessive number of display primitives causes the darker regions to become completely black, hence loosing detail in these regions.

The dependency of the change in intensity on the number of display primitives is illustrated in figure 4.9. Figure 4.9(a) and (c) have the same number of black and white pixels, as do figures (b) and (d). The apparent differences between figures 4.9(a) and (c) and figures 4.9(b) and (d) come from simply distributing the black and white pixels differently. In figures 4.9(a) and (b) the importance function is directly proportional to intensity and the display primitive is a single black pixel. In figures 4.9(c) and (d) the importance function is inversely proportional to intensity and the display primitive is a single white pixel. Figures 4.9(b) and (c) have an excess number of display primitives and figures 4.9(c) and (d) have an insufficient number of display primitives. Thus, given an excessive number of display primitives the darker regions get proportionally darker than the lighter regions in figure 4.9(b). In figure 4.9(c) the lighter regions get proportionally more light than the darker regions. An excess of display primitives can be used to emphasize contrast. However, if

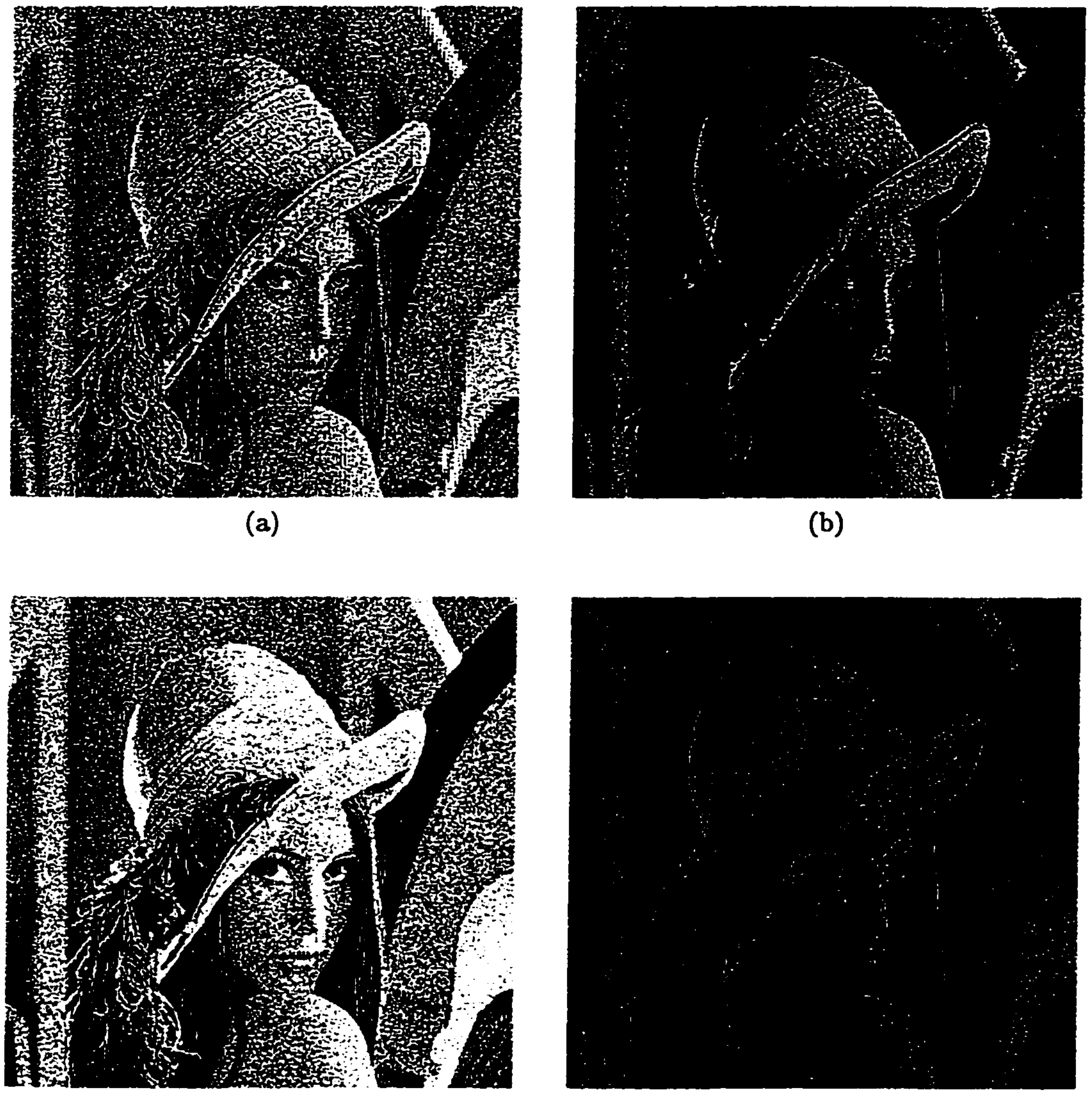


Figure 4.9: (a) Black ink, 80000 Display Primitives; (b) Black ink, 180000 Display Primitives; (c) White ink, 180000 Display Primitives - same amount of white ink as in (a); (d) White ink, 80000 Display Primitives - same amount of white ink as in (b). All images have intensity as the importance function.

the number of display primitives is insufficient (figures 4.9(a) and (d)), none of the regions become dark enough (figure 4.9(a)) or none of the regions are light enough figure 4.9(d). An insufficient number of display primitives results in a lessening of contrast. This illustrates that the distribution of display primitives is proportional to the number of display primitives and that varying the number of display primitives can not only lighten or darken the entire image, but can also be used to enhance or lessen contrast.

4.4 Type of Display Primitive

The third parameter of importance driven halftoning is the type of display primitive. The type of display primitive defines the local clustering of the ink. This can introduce local texture effects into the image. These local texture effects can in turn affect the global appearance of the image.

4.4.1 Individual Pixels

The first type of display primitive used was a single pixel. Using this as the display primitive resulted in images as shown in figures 4.2 to 4.4 and figures 4.6 to 4.9. These results are similar to traditional halftoning results. The evaluation of these results in this context will be further explored in chapter 5.

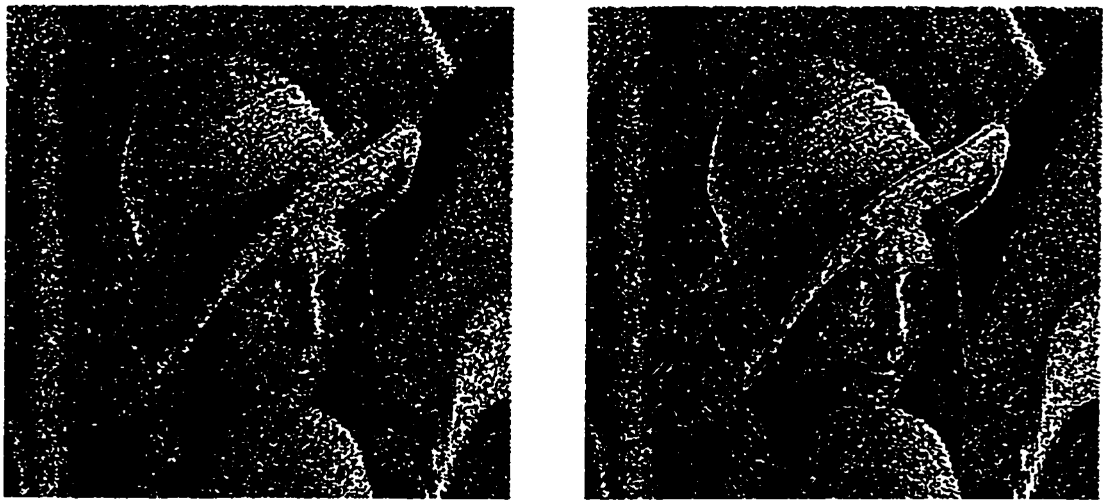
4.4.2 Line Segments

A second type of display primitive used was a line segment. The maximum length of the line segment is specified by the user and the line segment is centered on the determined, display primitive location (x, y). Once this is done there are many options for performing local adjustment to the placement of ink. As mentioned in the previous chapter, three types of adjustment to this display primitive have been experimented with: scaling, thresholding and rotation. These modifications to the local placement of ink produce both photorealistic and non-photorealistic results.

Scaling

Figure 4.10 illustrates two halftoned images using intensity as the importance function and black line segments of five pixels as the display primitives. Both images use a constant line length throughout the image. The image in figure 4.10(a) is aligned with the direction of the gradient and the image in figure 4.10(b) is rotated at 90^o to the direction of the gradient. In both images the content is quite distorted. The intensities of the image have been preserved, but the somewhat, seemingly random or white noise texture distorts the attributes of the image. Rotating the line segments 90^o to the direction of the gradient preserves the change in intensity and highlights edges as detected by the gradient. Improvement to intensity approximation can be achieved by scaling the length of the line segment as a function of intensity as follows:

$$Length(l) = ||g(x, y)|| \times \text{Max Length}$$



(a) (b)

Figure 4.10: (a) Importance function is intensity. The display primitive is a five pixel line segment aligned with gradient; (b) Importance function is intensity. The display primitive is a five pixel line segment rotated at 90^o to the direction of the gradient. The average number of display primitives as calculated in equation 4.1 were used for both images.

(a) (b)

Figure 4.11: Both images have intensity as the importance function. The display primitive is a 10 pixel line segment scaled to intensity and rotated at 90^o to the direction of the gradient. The average number of display primitives as calculated in equation 4.1 were used for both images.

(a) (b)

Figure 4.12: (a) Importance function is intensity. The display primitive is a 10 pixel line segment scaled to gradient and thresholded. This image uses the average number of display primitives; (b) Importance function is intensity. The display primitive is a 10 pixel line segment scaled to the gradient, thresholded, and rotated at 90^o to the direction of the gradient. This image uses 66% of the average number of display primitives.

where l is the line segment to be centered at location (x, y), $g(x, y)$ is the intensity at location (x, y) in the original image and the user has specified the maximum desired length of the line segment to be Max Length. Scaling the line segment to the intensity enhances contrast in the image altering the number of pixels set in regions of various intensities. Comparing the image in figure 4.10(b) with the image in figure 4.11(b) illustrates the effect of scaling the line segments. Both images have approximately the average number of black pixels set on white paper according to the average intensity as specified in equation 4.1. Both images in figure 4.11 have the average number of pixels set, have intensity as the importance function, and use 10 pixel line segments scaled to intensity and rotated at 90^o to the direction of the gradient as the display primitive. The line segment can also be scaled to other image properties such as variance or gradient depending on the desired effect. An example of this combined with other alterations will be shown in section 4.5.

Thresholding

Threshold matrix halftoning often yields adequate results because it approximates an image region using the underlying image information. Importance driven halftoning can also use the image information to threshold the line segment. To do this a threshold matrix is constructed along the length of the line segment. The matrix sets the thresholds starting from the center of the line segment and increasing outwards to the ends of the line segment as described in chapter 3, section 3.3.4. Thresholding the line segment allows the direct adjustment of the length of the line segment to a particular image area. This ability to adjust the length of the line segment aids in the preservation of edges and high frequency information. The images in figure 4.12 illustrate this local alteration.

(a) (b)

Figure 4.13: (a) Importance function is intensity. The display primitive is a line segment of length 10 scaled to intensity and thresholded, rotated at 90° to the direction of the gradient; (b) Importance function is intensity. The display primitive is a line segment of length 10 scaled to intensity, thresholded, rotated at 90° to the direction of the gradient in regions of high gradient, and rotated at 45° to the direction of the image axis in regions of low gradient.

In both images the line segment is length ten and is scaled to the gradient and thresholded. The image in figure 4.12(a) contains an average number of black pixels and the image in figure 4.12(b) contains about 66% of the average number of black pixels for this image. The image in figure 4.12(a) has the line segments oriented to the direction of the gradient and the image in figure 4.12(b) has the line segments rotated at 90° to the direction of the gradient. Comparing the image in figure 4.12(b) with 4.11(b) illustrates the effect of scaling the line segment. The images in figure 4.13 also show line segments that have been thresholded to the underlying image intensities.

Rotation

Rotation is the third explored method of locally adjusting the placement of ink. As seen in figures 4.10 to 4.13(a) rotating the line segment so that it lies 90° to the direction of the gradient preserves contrast, hence aiding in the preservation of edges and high frequency information. The line segments can also be aligned uniformly to create texture other than noise. However, since the edges and regions of high contrast must still be preserved, the rotation is thresholded so that regions with high gradient magnitude have the line segments oriented at 90° to the direction of the gradient, but regions with low gradient magnitude (semi-uniform intensity) have the line segments aligned in a common direction for the creation of an obvious texture. The image in figure 4.13(a) orients the line segments in the direction of the gradient and the image in figure 4.13(b) orients the line segments in the direction of the gradient in regions of high gradient magnitude and orients the lines at 45° to the direction of the image axis in regions of low gradient magnitude. Both of these images threshold the line segments to intensity and use intensity as the importance function. This change in orientation

(a) (b)

Figure 4.14: (a) Importance function is intensity. The display primitive is a 10 pixel line segment rotated 45^o to the direction of the image axis; (b) Importance function is intensity. The display primitive is a 10 pixel line segment rotated as a function of intensity.

can be seen by comparing the hat and the shirt regions of figures 4.13(a) and (b). Finally in figure 4.14 the line segments are a constant length of ten and the line segments are oriented at 90^o to the direction of the gradient in regions of high gradient magnitude. In regions of low gradient magnitude figure 4.14(a) orients the line segments at 45^o to the direction of the image axis and figure 4.14(b) orients the line segments as a function of intensity between 0^o and 90^o to the direction of the image axis. In the last image the amount of rotation for the line segments (angle Θ) is calculated in degrees as follows:

$$\Theta = ||g(x, y)|| * 90$$

4.4.3 Square Region of Pixels

A third type of display primitive that has been used is a square. The local placement of ink can be adjusted using a rectangle or square as the display primitive and thresholding this area using a traditional threshold matrix. This is done by simply traversing down the lowpass pyramid until a region the size of the matrix is reached. This area is then thresholded according to the values in the threshold matrix. There are two methods of performing this thresholding:

- A: Set the first n lowest regions of the matrix, where n is the number of display primitives left to be distributed at this point in the pyramid
- B: Threshold the region using the threshold matrix and set the first n regions that are required to be set (as defined by the original image intensity and the threshold value), where n is the number of display primitives left.

(a) (b)

Figure 4.15: (a) Importance function is intensity. The display primitive is a 4×4 square, thresholded by method A using a clustered 4×4 matrix given in section 4.4.3; (b) Importance function is intensity. The display primitive is a 4×4 square, thresholded by method A using a dispersed 4×4 matrix given in section 4.4.3.

(a) (b)

Figure 4.16: (a) Importance function is intensity. The display primitive is a 4×4 square, thresholded by method B using a clustered 4×4 matrix given in section 4.4.3; (b) Importance function is intensity. The display primitive is a 4×4 square, thresholded by method B using a dispersed 4×4 matrix given in section 4.4.3.

The images in figures 4.15 and 4.16 use the following 4 × 4 threshold matrices for cluster and dispersed thresholding, respectively:

Clustered

$$\begin{vmatrix} 14 & 12 & 13 & 16 \\ 5 & 4 & 3 & 10 \\ 6 & 1 & 2 & 11 \\ 9 & 7 & 8 & 15 \end{vmatrix}$$

Dispersed

$$\begin{vmatrix} 10 & 6 & 11 & 7 \\ 4 & 14 & 1 & 15 \\ 12 & 8 & 9 & 5 \\ 2 & 16 & 3 & 13 \end{vmatrix}$$

All four of these images were generated using the average number of black pixels and thus result in images very similar to traditional threshold matrix halftoning using clustered and dispersed threshold matrices. Method A sets the pixels within the matrix in the same order and hence has the same basic regular structure. This creates noticeable artifacts in the image. Method B sets the pixels in an order dependent on the underlying image. This causes a variation in the structure which not only reduces the introduction of artifacts, but also preserves edges and high frequency information. Halftoning the image using importance driven halftoning has the advantage of easily limiting the amount of ink used in the image. This will be illustrated in section 4.5.3.

Thus, by scaling, thresholding, and re-orienting the display primitive, as well as altering the type of display primitive importance driven halftoning can be used to create very different local and global effects and yet still preserve important image attributes.

4.5 Alteration of Multiple Parameters

The previous section showed how altering the three user controlled parameters individually can create various effects. This section will illustrate results from altering these parameters simultaneously.

4.5.1 Importance Function and Quantity of Display Primitives

Previously in this chapter the effects of changing both the importance function and the number of display primitives simultaneously were illustrated. Figure 4.7 showed the result of using intensity as the importance function and altering the number of display primitives both above and below the average number as calculated by equation 4.1. The results in figure 4.7 can be compared with figure 4.8 where gradient is the importance function and 25% of the average number of display primitives are used. These images show how very different results can be created through simultaneously altering the importance function and the number of display primitives.

4.5.2 Local Adjustment to Display Primitives and Limited Resource Rendering

By adjusting the structure of the display primitive and restricting the resources images can be created that appear similar to sketches of the original grey-scale image. In figures 4.17 (a) and (b) the importance function is intensity and the display primitive is a line segment using 50% of the average number of primitives. In figure 4.17 (a) the line segment is oriented at 90^o to the direction

(a) (b)

Figure 4.17: Both images have intensity as the importance function and a line segment as the display primitive. The line segments in both images are scaled to the gradient and thresholded. In the image on the left (a) the line segment is rotated at 90° to the direction of the gradient and in the image on the right (b) the line segments are rotated at 90° to the direction of the gradient in regions of high gradient and at 45° to the direction of the image axis in regions of low gradient. Both images use approximately 50% of the average number of display primitives.

Figure 4.18: Importance function is intensity. The display primitive uses a line segment scaled to the gradient and thresholded. The line segments are rotated at 90° to the direction of the gradient in regions of high gradient and at 45° to the direction of the image axis in regions of low gradient. Approximately 25% of the average number of display primitives are used.

Figure 4.19: Importance function is intensity. The display primitive is a line segment scaled to intensity and thresholded, rotated at 90° to the direction of the gradient. The average number of display primitives are used.

(a) (b)

Figure 4.20: Both images use intensity as the importance function. The display primitive is a square thresholded by method B using a clustered 4×4 threshold matrix. (a) 50% of the average number of display primitives are used; (b) 25% of the average number of display primitives are used.

of the gradient creating a non-uniform and less noticeable texture. This image illustrates how edges or outline is still preserved with less resources. By preserving relative intensity using line segments to enhance outlines the resulting image can appear similar to a sketch. In figure 4.17(b) the line segment is oriented at 45^o to the direction of the image axis in regions of more uniform intensity creating a more obvious texture in the image. By altering these multiple parameters there are senses of outline, relative-tone, and texture to these images. These are the primary components of non-photorealistic images (see section 2.4). Figure 4.18 uses half the display primitives of figure 4.17(b) resulting in the appearance of a rough sketch or draft version.

Traditional halftoned "like" results can be created by altering the local placement of ink and the number display primitives. In figure 4.19 the average number of display primitives are used and the display primitives (line segments) are scaled and thresholded. A more detailed comparison of these results will be made in chapter 5, section 5.2.

4.5.3 Ease of Ink Reduction in Traditional Techniques

Finally, results similar to traditional threshold matrix results can be created with compensation for hardware limitations. As mentioned in chapter 1, laser printers do not have the ability to independently set pixels especially white pixels surrounded by black, due to dot gain. Importance driven halftoning provides an easy method for reducing the amount of ink used to render the image even with traditional halftoning techniques. Figure 4.20(a) uses 50% of the number of display primitives as determined by equation 4.1 and figure 4.20(b) uses 25% of the average number of display primitives. Since the distribution of display primitives is relative to the number of display primitives this causes the dark regions to become relatively less dark than the light regions. This can be useful for displaying images on devices such as laser printers where the ink tends to bleed and the intensities near the dark end of the spectrum can sometimes be lost.

4.6 Chapter Summary

This chapter has shown how altering the three user controlled parameters can drastically alter the results obtained from the halftoning process. It has also shown how altering the importance function allows the user to place emphasis or priority on certain image attributes, as well as weight multiple important attributes to assess relative importance. Results using intensity, variance, gradient and combinations of these functions for the importance function were illustrated. Using these importance functions was shown to be useful in preserving both relative intensity and contrast or edges. Altering the number of display primitives has been shown to alter the overall look of the image by decreasing or increasing the amount of ink used to render the image. The ability to control the amount of ink can be useful in creating rough or draft versions of the image or for emphasizing or possibly distorting important attributes of the image. Reducing or increasing the number of display primitives when using intensity as the importance function was shown to not only alter the overall darkness or brightness of the image, but was also shown to be useful in the control over contrast. Over emphasizing the number of display primitives enhances contrast and reducing the number of display primitives lessens contrast. Reducing the number of display primitives is useful in creating reduced resource rendered images while still retaining important image attributes. Finally, the type of display primitive has been shown to be useful in creating an overall texture in the image. Three different types of display primitives were used: an individual pixel, a line segment, and a square 4×4 region of pixels. Varying the type of display primitive produced different results. The local placement of ink was further controlled by adjusting the structure of the display primitive to image properties. This was done using three techniques: scaling, rotation, and thresholding. Aligning display primitives semi-uniformly was shown to be useful in the creation of texture. Using line segments with various importance functions, such as gradient and variance create effects similar to sketches or outlines. Using square regions and thresholding produced results similar to traditional threshold matrix halftoning results. This is useful in creating non-photorealistic effects.

These three user controlled parameters work together to create alterations in tone, texture and outline of the image. Ultimately, having control over these properties of the image allows the user to create very different and interesting results.

Chapter 5

Evaluation of Importance Driven Halftoning

This chapter evaluates importance driven halftoning in the context of current assessment methods. Evaluation of this technique will be performed by assessing the advantages and disadvantages of the importance driven halftoning process and the results of this halftoning technique. The technique will be assessed by analyzing memory requirements and computational complexity compared to other halftoning techniques. The halftoned results from importance driven halftoning will then be evaluated. In the fields of both computer graphics and image processing there is a need for the assessment or evaluation of rendered or processed images. Both of these fields have adapted two main methods for image assessment: qualitative and quantitative measurements. This chapter will qualitatively assess the results of importance driven halftoning by visually comparing components of the image. The quantitative assessment of our results will be made using two measures proposed by Ulichney [Ulic87] and Veryovka *et al* [Very98].

5.1 Evaluation of Importance Driven Halftoning Process

Most traditional halftoning techniques were designed to require only one or two scanlines of memory due to printer memory limitations, as well as were designed to be fast and simple so that the halftoning process could be performed while the document is in the printer queue. With the availability of memory, other techniques have been developed with objectives other than memory restrictions and computational simplicity.

5.1.1 Memory Requirements

Error diffusion techniques such as Floyd-Steinberg's error diffusion typically require approximately two scanlines of memory. Traditional threshold matrix halftoning techniques only require enough memory to store a single pixel value and the size of the threshold matrix. Research is currently being done on the creation of large dither matrices. As the dither matrices become larger and larger more memory is required for these techniques. Optimization techniques tend to use more memory than both error diffusion and threshold matrix techniques. Analoui and Allebach [Anal92] created an optimized halftoning technique based on a model of the human visual system. They adapted error diffusion to ***look ahead*** to monitor the diffusion of error. To do this they build a binary tree of possible pixel settings and error diffusions. Thus, the memory needed is $2 \times \text{Size Of Scanline} \times (2^k - 1)$ where k is the depth of the binary tree.

Importance driven halftoning trades off memory for computational complexity. The amount of memory required is approximately 3/2 times the size of the image. This is required to build the pyramidal structure that is essential for the distribution of display primitives. Thus, importance driven halftoning requires more memory that traditional halftoning techniques, but tends to use less than most of the optimized methods of halftoning.

5.1.2 Complexity

As mentioned previously most traditional halftoning techniques were designed with simplicity in mind. Typical error diffusion techniques, specifically Floyd-Steinberg's error diffusion, perform five operations (size of error matrix + 1) per pixel and threshold matrix methods only perform one operation per pixel. Techniques that combine both error diffusion and threshold matrix methods tend to encompass the complexity of both. Finally, other optimization methods ([Mull92, Papp92, Zakh92]) use re-iterative halftoning processes. This means that the complexity of the whole process is multiplied many times before an optimized solution is reached. Analoui and Allebach's [Anal92] optimized halftoning technique could result in $(2^k - 1) \times (n \times m)$ operations where k is the depth of the binary tree and $n \times m$ is the size of the image. This is in addition to the evaluation of the stopping condition after each iteration.

Importance driven halftoning is more complex than the traditional methods, but less complex than optimized methods. To build the pyramid the following number of operations are needed:

$$log_4(n \times m) + \text{Importance Fcn Complexity} \times (n \times m)$$

where $n \times m$ is the size of the image. To actually set the pixels or display primitives at most $3/2 \times n \times m$ operations are needed. The complexity of this technique is comparable to Floyd-Steinberg's error diffusion.

5.2 Evaluation of Importance Driven Halftoning Results

There are two main methods of evaluating halftoned results: qualitative and quantitative. Qualitative assessment of images relies heavily on the human visual system and the observations of participants for assessment. Hence these assessments of images tend to be highly subjective. Quantitative measures provide a more objective assessment of images. However, these assessments are generally less complete since there does not currently exist a metric for evaluating all dimensions

Figure 5.1: Floyd-Steinberg error diffusion halftoning

Figure 5.2: Threshold Matrix Clustered Ordered Dither using a 4x4 threshold matrix as shown in section 5.2.1.

necessary to assess the image. In this section, a set of images are presented to the reader for perceptual comparison. As well, two quantitative metrics that measure different image attributes are presented in order to objectively compare results from our halftoning technique with results from other halftoning techniques.

5.2.1 Comparison with Current Techniques

Due to the lack of an complete metric, one of the current methods of assessment used in halftoning is visual comparison. Figure 5.1 is halftoned using Floyd-Steinberg's error diffusion with serpentine traversal of the image. Figure 5.2 is clustered ordered dither threshold matrix halftoning and figure 5.3 is dispersed ordered dither threshold matrix halftoning. These two figures use the following threshold matrices respectively:

Clustered

14	12	13	16
5	4	3	10
6	1	2	11
9	7	8	15

Dispersed

10	6	11	7
4	14	1	15
12	8	9	5
2	16	3	13

Figures 5.4 to 5.6 are images halftoned using importance driven halftoning. Figure 5.5 uses intensity as the importance function, a single black pixel as the display primitive, and an average number of display primitives. Figure 5.6 is the same as figure 5.5 except it uses 40% more display primitives. This excess number of display primitives, as mentioned in chapter 4 improves or enhances contrast. Figure 5.4 uses intensity as the importance function. The display primitive is a line segment scaled to intensity, thresholded, and rotated at 90^o to the direction of the gradient. This image also uses an average number of display primitives.

Figure 5.3: Threshold Matrix Dispersed Ordered Dither using a 4x4 threshold matrix as shown in section 5.2.1.

Figure 5.4: Importance Driven Halftoning using intensity as the importance function. The display primitive is a line segment scaled to intensity, thresholded, and rotated at 90° to the gradient. This image uses an average number of display primitives.

Figure 5.5: Importance Driven Halftoning using intensity as the importance function. The display primitive is a single pixel and an average number of display primitives are used.

Figure 5.6: Importance Driven Halftoning using intensity as the importance function. The display primitive is a single black pixel. The number of display primitives used is 40% more than the average number needed.

These images can be compared by analyzing the common components of halftoned images, such as grey scale intensity reproduction, edge or high frequency information preservation, and unwanted artifact introduction. Comparing importance driven halftoning with traditional techniques reveals that importance driven halftoning (figure 5.5) reproduces approximately the same number of grey tones as the traditional techniques in figures 5.1 and 5.2 and has less quantization bands than dispersed ordered dither in figure 5.3. Areas of uniform intensity or uniform increasing intensity (*i.e.* the shoulder) look slightly *more grainy* with importance driven halftoning than with Floyd-Steinberg's error diffusion, but less grainy than with clustered ordered dither or dispersed ordered dither. The high frequency information (*i.e.* edges) seem to be preserved in importance driven halftoning (figures 5.5 or 5.6) better than in both of the threshold matrix images (figures 5.2 and 5.3) as well as Floyd-Steinberg's error diffusion. This can be illustrated by analyzing the creases in both the ribbon and the top part of the hat. These attributes are preserved in figure 5.5, but seem to be non-existent in the other images. Both importance driven halftoning and Floyd-Steinberg's error diffusion tend to introduce unwanted artifacts in other areas, such as the cheek and top part of the background in figure 5.1 and the lips and vertical rail in the left part of the image in figure 5.5. The importance driven halftoning technique illustrated in figure 5.6 has enhanced the contrast in the image. It reveals similar properties of preserving the details in the hat and ribbon and has reduced the regular artifacts in the lips and rail, but introduces a more grainy effect to the entire image. Based on these observations we claim that importance driven halftoning adequately reproduces the grey-scale intensities, introduces similar unwanted artifacts as error diffusion (better than threshold matrix methods), and preserves high frequency information better than both error diffusion and threshold matrix methods.

5.2.2 Quantitative Measures

The results of importance driven halftoning will be analyzed using two quantitative metrics. The first metric was defined by Ulichney [Ulic87] and is an analysis of the presence of various frequency information and symmetry of signal and noise information in the image. The second metric has been proposed by Veryovka *et al* [Very98] to measure the high frequency information for the preservation of edges in the original image and the introduction of unwanted edge effects. This metric measures these *edge effects* at multiple resolutions, and thus attempts to account for the visual system's ability to perceive and form interpretations at multiple resolutions. First details of Ulichney's image frequency metrics will be outlined. The results from evaluating this metric on three levels of grey will then be presented. Second, details of the edge metrics proposed by Veryovka *et al* will be outlined and the results of evaluating these metrics on two test images will be presented.

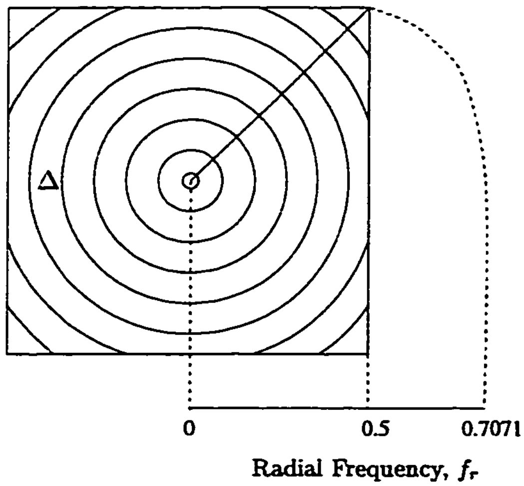


Figure 5.7: Segmenting of the spectral estimates. Each spectral estimate is partitioned into annuli of width Δ. From these, the samples are taken to calculate the radially average power spectrum as calculated by equation 5.1.

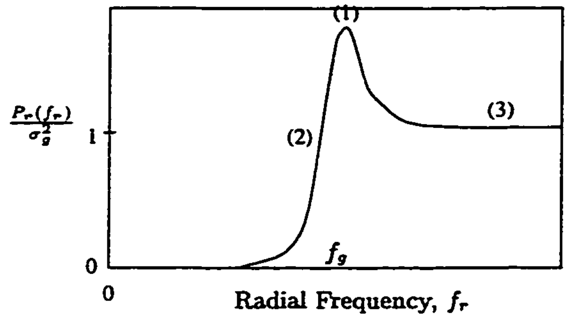


Figure 5.8: Spectral Characteristics of a Well Formed Dither Pattern. An ideal dither pattern will have very few frequencies below the principle frequency f_g, peak at the principle frequency and then decrease and level out after the principle frequency has been reached.

Image Frequency Metrics - Details

Ulichney's book [Ulic87] discusses two metrics for evaluating a constant fixed tone grey-scale region. One metric describes the distributions of frequencies within the uniform grey-scale region and is called the *power spectrum*. This is equivalent to describing the distribution of black pixels within the spatial domain. The second metric analyzes the symmetry of the distribution. As Ulichney [Ulic87] states "A desirable attribute of well produced aperiodic halftone of a fixed grey level is radial symmetry; directional artifacts are perceptually disturbing." This is a measure of ***anisotropy***.

The graph of the power spectrum can reveal the quality of the resulting dither pattern. The power spectrum is calculated by constructing spectral estimates in a similar fashion to the method used in Ulichney's book. The estimate $\tilde{P}(\mathbf{f})$ of $P(\mathbf{f})$ is constructed using Bartlett's Method [Bart55] of averaging periodograms. These periodograms are the magnitude squared of the Fourier transform of the sample output divided by the sample size. All of the following estimates have been produced using 10 samples each of 128×128 pixels from a halftone rendering of a fixed grey level. The samples are not placed near the edges of the halftoned image due to existence of artifacts at the edges of images that are not present throughout the rest of the halftoned image. Each spectral estimate $\tilde{P}(\mathbf{f})$ can be partitioned into annuli of width Δ for regular rectangular (see figure 5.7) or hexagonal grids. In this thesis we are only concerned with rectangular grids. Each annulus has a radial frequency f_r and $N_r(f_r)$ frequency samples. From this, the radially averaged power spectrum is calculated as follows:

$$P_r(f_r) = \frac{1}{N_r(f_r)} \sum_{i=1}^{N_r(f_r)} \tilde{P}(\mathbf{f}) \tag{5.1}$$

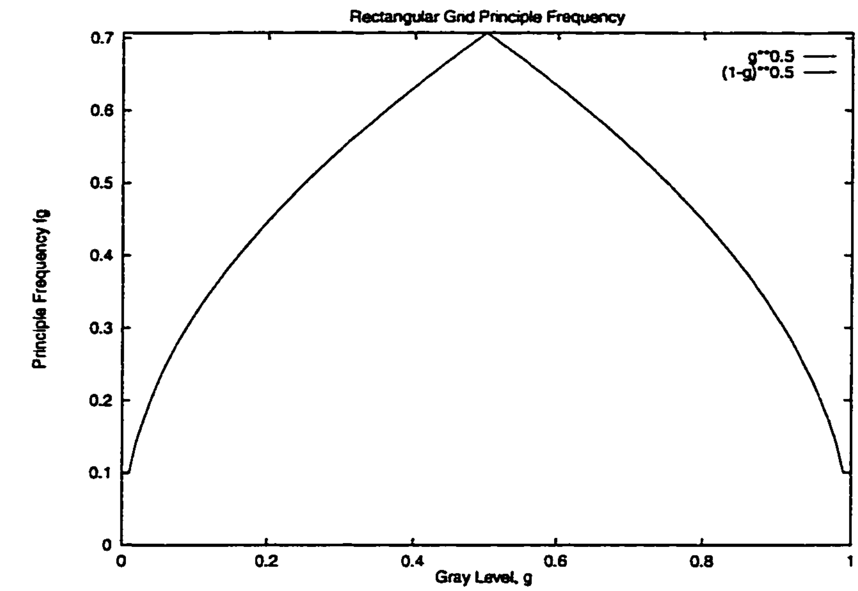


Figure 5.9: The Principle Frequency f_g is a function of the grey level and is calculated as in equation 5.2.

for each radial frequency r. Finally the power spectrum data is normalized by the variance, σ_g^2 which is calculated as $\sigma_g^2 = g \times (1 - g)$. This normalization is needed to render all plots on a relative scale since the spectral energy increases with σ_g^2.

Ulichney describes the spectral characteristics (graph of the radially averaged power spectrum) of a well formed dither pattern to be the following:

1. Low frequency cutoff at principal frequency
2. Sharp Transition region
3. Flat high frequency "Blue Noise" region

These characteristics are illustrated in figure 5.8. The principal frequency, f_g is calculated as follows (shown in figure 5.9)

$$f_g = \begin{cases} g^{0.5} & \text{if } g < 0.5 \\ (1-g)^{0.5} & \text{otherwise} \end{cases} \tag{5.2}$$

Anisotropy measures the radial symmetry of frequencies in a halftoned image of a fixed grey level. The radial symmetry can be qualitatively assessed using three dimensional plots of $\hat{P}(\mathbf{f})$. From the radially averaged power spectrum the sample variance can be calculated as follows:

$$s^2(f_r) = \frac{1}{N_r(f_r) - 1} \sum_{i=1}^{N_r(f_r)} \left(\hat{P}(\mathbf{f}) - P_r(f_r)\right)^2$$

From this the measure of anisotropy can be calculated as defined by Ulichney as:

$$\frac{s^2(f_r)}{P_r^2(f_r)}$$

This is a measure of the relative variance of frequency samples for each radial frequency r. This measure has also been described as the coefficient of variance or "noise-to-signal" ratio. The graphs will be plotted in decibels of this measure due to the range of values. Assuming $P(\mathbf{f})$ is perfectly radially symmetric, an anisotropy measure of $\frac{1}{10}$ or -10dB can be considered to be "background noise" (see Ulichney [Ulic87] p.59 for details) since only 10 periodograms are used. Hence, a measure of anisotropy close to this value indicates a halftone that is fairly radially symmetric.

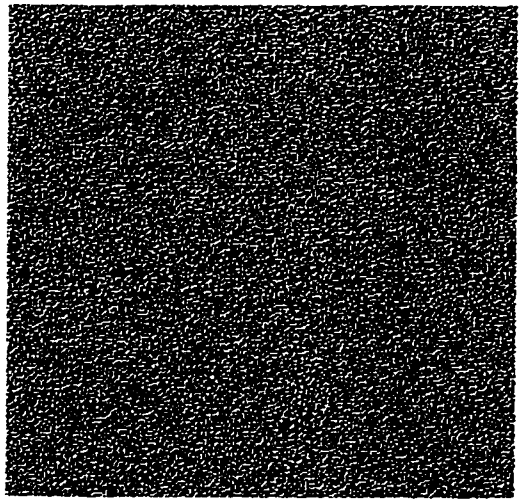
Image Frequency Metrics - Results

In figures 5.10 to 5.15 Floyd-Steinberg's error diffusion is compared with importance driven halftoning by analyzing both the radially averaged power spectrum and anisotropy measures for three fixed levels of grey.

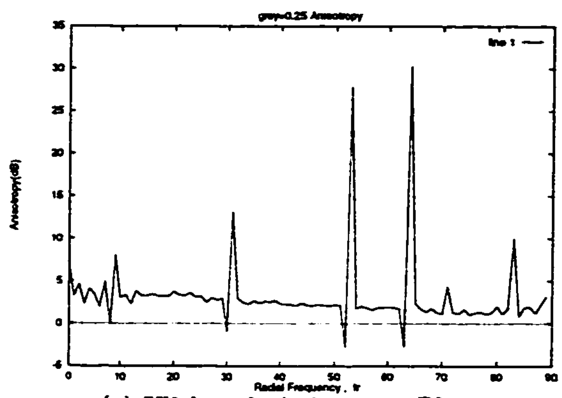
Figures 5.10 and 5.11 show the power spectrum and anisotropy plots for importance driven halftoning and Floyd-Steinberg's error diffusion for grey level 0.25. The principle frequency at this level of grey is 64. Both techniques have a low frequency cutoff at the principle frequency, however Floyd-Sternberg' error diffusion has slightly less low frequency information before f_g. Thus, both techniques at this grey level seem to have similar frequency distributions (see figures 5.10(b) and 5.11(b)). The anisotropy plots (see figures 5.10(a) and 5.11(a)) reveal that Floyd-Steinberg's error diffusion is slightly less radially symmetric than importance driven halftoning. This is shown by the anisotropy graph in figure 5.11(a) having generally lower values than the anisotropy graph in figure 5.10(a).

Figures 5.12 and 5.13 show the power spectrum and anisotropy plots for importance driven halftoning and Floyd-Steinberg's error diffusion for grey level 0.5. The principal frequency at this level of grey is 90. Comparing the power spectrums (see figures 5.12(b) and 5.13(b)) at this level of grey, show that importance driven halftoning has more energy concentrated in frequencies below the principal frequency cutoff than Floyd-Steinberg's error diffusion. However, importance driven halftoning appears to be much more radially symmetric than Floyd-Steinberg's error diffusion as seen by figures 5.12(a) and 5.13(a).

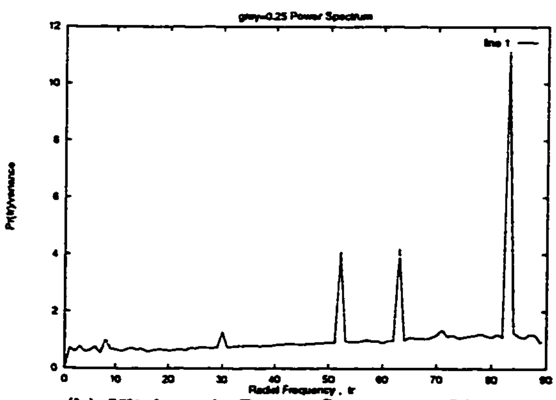
Finally figures 5.14 and 5.15 show Ulichney's metrics for grey level 0.75 and principal frequency 64. With this grey level importance driven halftoning has relatively the same amount of energy concentrated in frequencies below the principal frequency as Floyd-Steinberg's error diffusion (see figure 5.14(b) and 5.15(b)), however there seems to be a more definite peak and leveling out after the principal frequency with importance driven halftoning than with Floyd-Steinberg's error diffusion. The radial symmetry seems to be relatively the same as well, except, once again there seems to be definite spikes in the anisotropy with importance driven halftoning (see figures 5.14(a) and 5.15(a)).



Importance Driven Halftoning grey=0.25

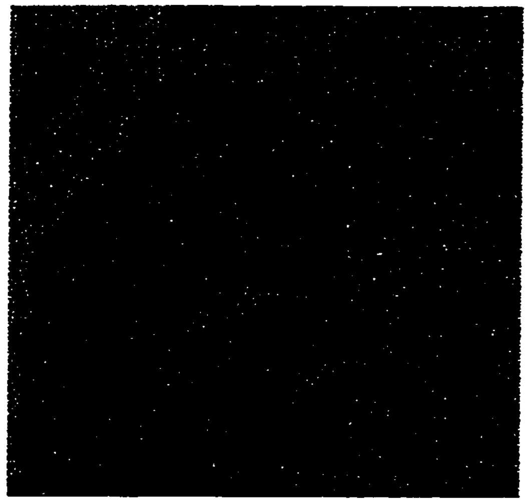


(a) Ulichney's Anisotropy Plot

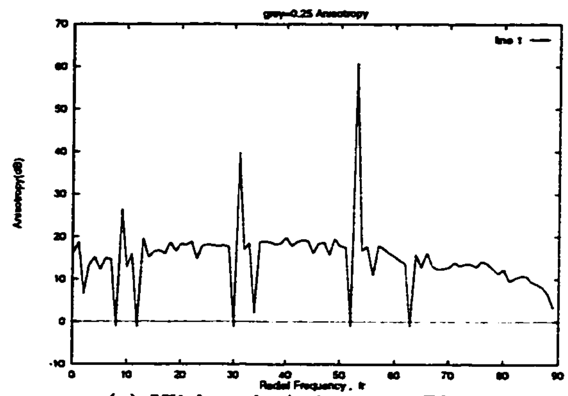


(b) Ulichney's Power Spectrum Plot

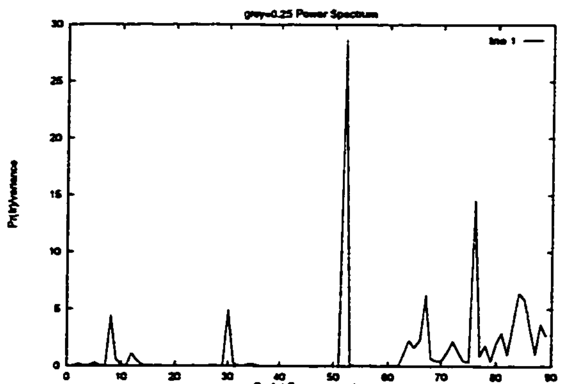
Figure 5.10: Importance Driven Halftoning, grey=0.25, $f_g = 64$



Floyd Steinberg's Error Diffusion grey=0.25

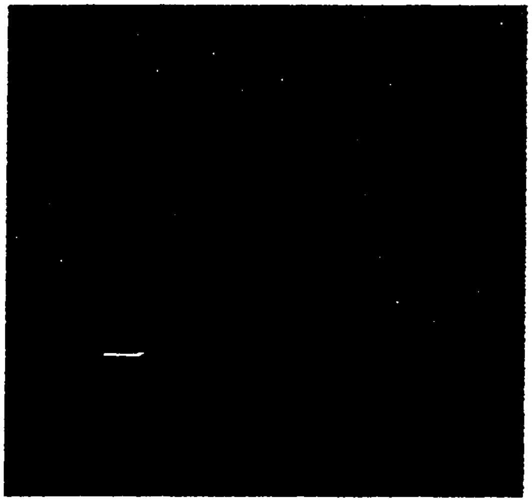


(a) Ulichney's Anisotropy Plot

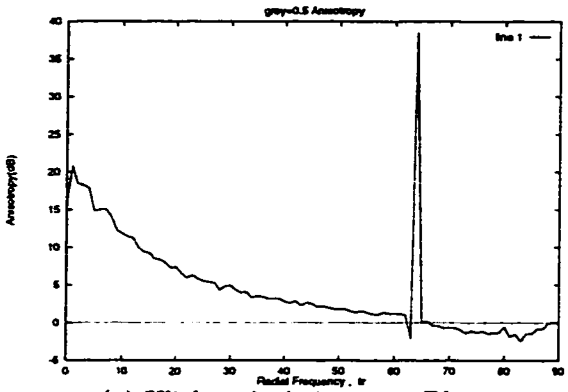


(b) Ulichney's Power Spectrum Plot

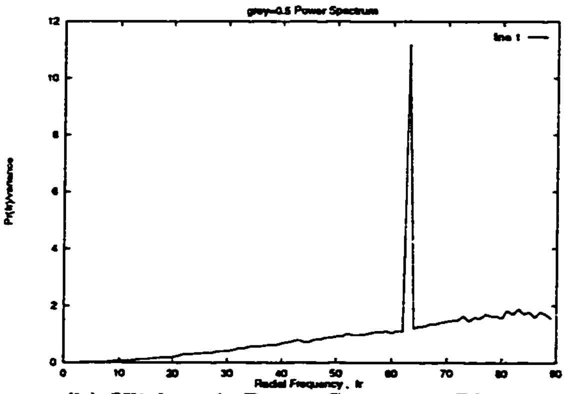
Figure 5.11: Floyd-Steinberg's Error Diffusion, grey=0.25, $f_g = 64$



Importance Driven Halftoning grey=0.5

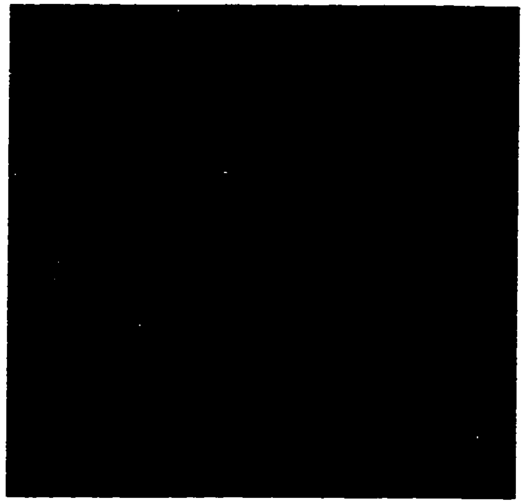


(a) Ulichney's Anisotropy Plot

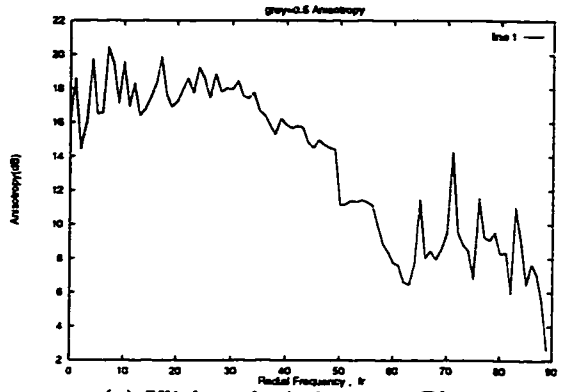


(b) Ulichney's Power Spectrum Plot

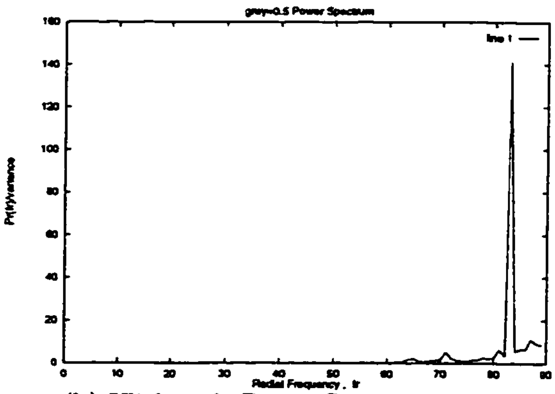
Figure 5.12: Importance Driven Halftoning, grey=0.5, $f_g = 90$



Floyd Steinberg's Error Diffusion grey=0.5



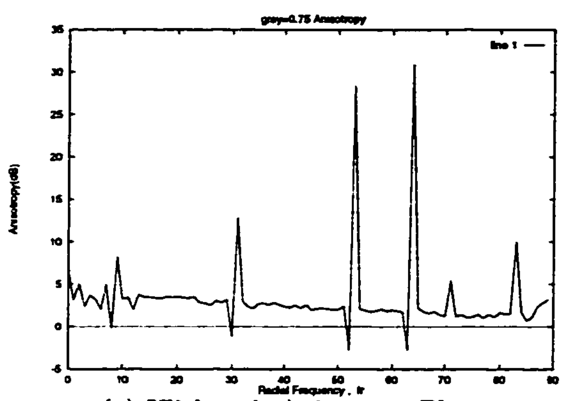
(a) Ulichney's Anisotropy Plot



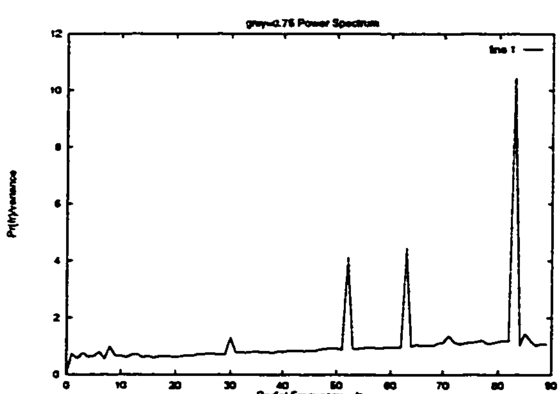
(b) Ulichney's Power Spectrum Plot

Figure 5.13: Floyd Steinberg's Error Diffusion, grey=0.5, $f_g = 90$

Importance Driven Halftoning grey=0.75

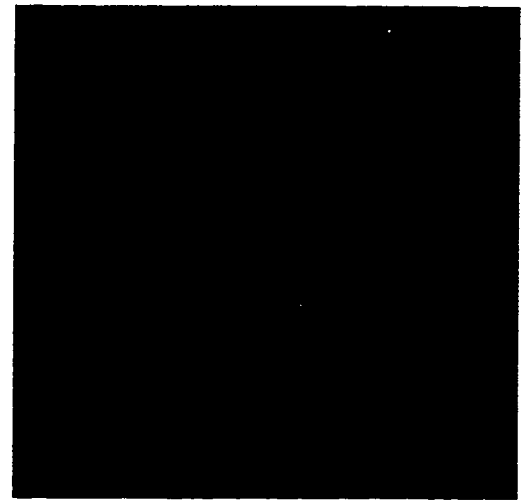


(a) Ulichney's Anisotropy Plot

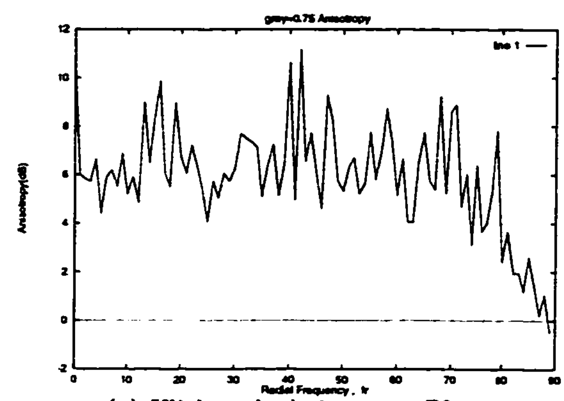


(b) Ulichney's Power Spectrum Plot

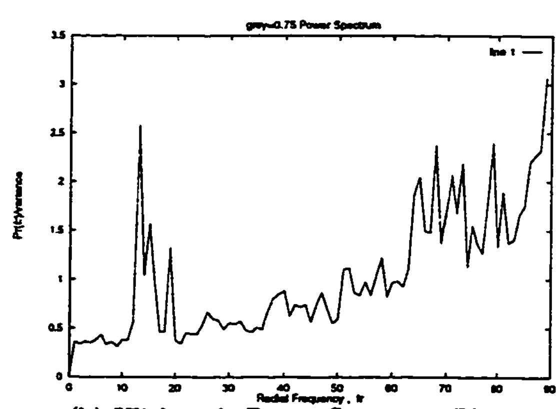
Figure 5.14: Importance Driven Halftoning grey=0.75, $f_g = 64$



Floyd Steinberg's Error Diffusion grey=0.75



(a) Ulichney's Anisotropy Plot



(b) Ulichney's Power Spectrum Plot

Figure 5.15: Floyd Steinberg's Error Diffusion grey=0.75, $f_g = 64$

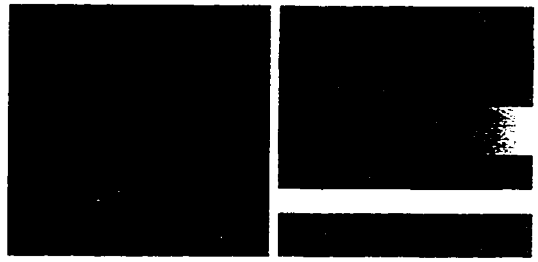


Figure 5.16: Original images for edge evaluation in chapter 5.

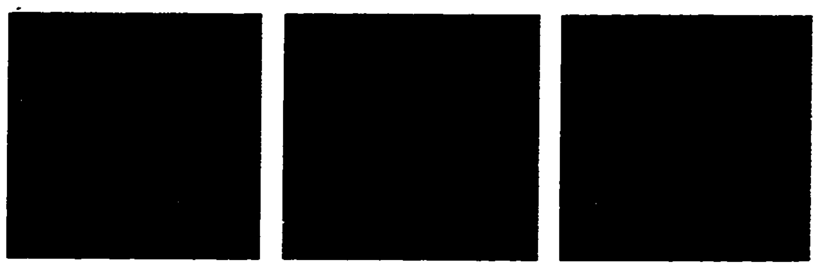


Figure 5.17: Left to Right: 4x4 Dispersed Ordered Dither; Floyd-Steinberg Error Diffusion; Importance Driven Halftoning

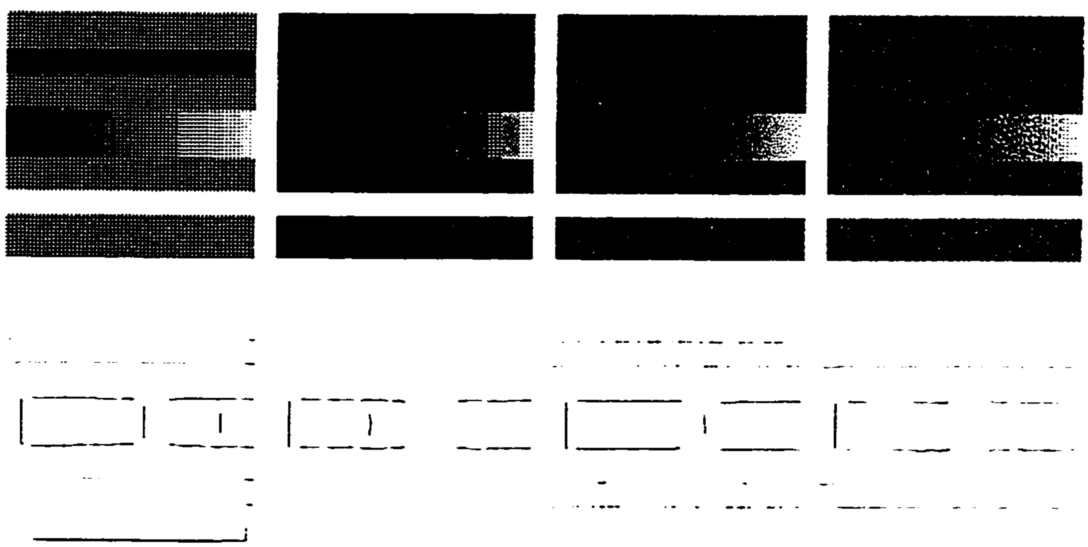


Figure 5.18: Top: Halftoned Images; Bottom: Edge errors detected; Left to Right: 4x4 Clustered Ordered Dither; 4x4 Dispersed Ordered Dither; Floyd-Steinberg Error Diffusion; Importance Driven Halftoning

Halftoning Technique	Values	scale 1	scale 2	scale 3	scale 4	scale 5
Dispersed dither	Min	26.719	8.871	0.484	0.619	0.220
	Average	101.298	16.273	5.439	5.627	5.285
	Max	120.443	31.937	20.822	18.303	16.862
Floyd-Steinberg	Min	85.166	1.774	0.171	0.560	0.015
	Average	103.979	14.406	5.257	5.728	3.022
	Max	120.443	33.336	26.964	23.674	21.853
Importance Driven Halftoning	Min	14.361	1.774	2.065	1.619	0.219
	Average	98.352	22.127	8.655	8.145	3.107
	Max	120.443	44.357	22.134	16.684	16.016

Table 5.1: This table reflects the strength of edge artifacts across five scales. The results of the artifact minimum, average, and maximum strength edges are shown.

Multi-scale Edge Metrics - Details

Ulichney's metrics are good for measuring frequency distributions within grey-scale images, but this does not account for another important attribute in halftoning, *edges*. Veryovka *et al* citeVeryovka98 proposed a metric for measuring edges at multiple scales. This metric not only measures the distortion of edges between the original image and the halftoned image, but also the lack of edges in the halftoned image that are present in the original image and the presence of edges in the halftoned image that did not exist in the original image. They call this a measure of *edge preservation* and *creation of edge artifacts*. Since edges are considered high frequency information this metric is able to detect high frequency information over multiple scales. This multi-scale analysis is done using a wavelet transform. Extracting extrema points from wavelet transforms and constructing scale based edge contours allow for the identification of discontinuities and the measurement of their significance. In their paper [Very98] they showed that these measures can be used to both measure fidelity of edge reproduction and reliably identify edge artifacts.

The two original grey-scale images used for the two measures of *edge artifacts* and *edge preservation* are shown in figure 5.16. The image on the left is a continuous tone computer generated ramp from black to 75% black. This image is used to indicate the introduction of artifacts into continuous tone regions - *edge artifacts*. The image on the right is a 50% black image with a band of white, a band of a continuous tone ramp from black to white and a band of black. This image provides different edges defined by different changes in intensity. It is used to indicate the destruction or preservation of edges - *edge preservation*.

Multi-scale Edge Metrics - Results

Table 5.1 shows the results of using the edge artifact measure from Veryovka *et al* [Very98] on the images in figure 5.17. The values reported in this table reflect the strength of the unwanted artifacts where a higher value indicates more artifacts. The table shows the minimum, average, and maximum values of the edge strength. By analyzing the three values an idea of both the average and the distribution can be formulated. On scale 1 all three metrics have the same maximum values. Im-

Halftoning Technique	Minimum	Average	Maximum
Clustered dither	-6.034	5.087	80.468
Dispersed dither	-5.679	5.463	79.723
Floyd-Steinberg	-5.779	3.289	83.600
Importance Driven halftoning	-6.624	4.148	71.746

Table 5.2: This table reflects the preservation of edges. A negative edge value represents unwanted smoothing of the edges during the halftoning process. A positive value represents an edge that has been exaggerated or enhanced during the halftoning process.

portance driven halftoning introduces edges that are weaker on average than both Floyd-Steinberg's error diffusion and dispersed ordered dither. By analyzing the three values, it can be concluded that importance driven halftoning introduces more weak edges than the other two methods, accounting for the drop in the average reported edge strength. At the second scale importance driven halftoning does worse on average than both of the other halftoning techniques. The minimum strength edge is as small as Floyd-Steinberg's, however the maximum strength edge is larger than both techniques, thus accounting for the increase in average value. On scales 3 and 4 the average values for importance driven halftoning is again higher than both the others. However, the maximum strength edge on both scales is lower than Floyd-Steinberg's error diffusion. The minimum edge strength for importance driven halftoning on these two scales are, however higher causing the increase in average values. On scale 5 the average value edge strength for importance driven halftoning is comparable to Floyd-Steinberg's error diffusion and is lower than dispersed ordered dither. The maximum and minimum edge strength values for importance driven halftoning are comparable to dispersed ordered dither and lower than Floyd-Steinberg's error diffusion.

Table 5.2 shows the results of measuring edge preservation on images shown in the top row of figure 5.18. This measure takes into account that the human visual system does not detect small shifts in the positioning of the edge, as long as the edge geometry is preserved. Also very small edges are either not noticed or are noticed as noise or artifacts rather than edges, so these have been discarded. The values presented in the table measure the edge distortions on scale 4. Thus a value of zero is the most desirable result. A negative value represents smoothing of the edge throughout the halftoning process and a positive value represents an exaggeration of the edge throughout the halftoning process. Importance driven halftoning has less distortion of edges on average than both threshold matrix techniques, however is not quite as good as Floyd-Steinberg's error diffusion. It is also shown that importance driven halftoning tends to exaggerate edges more than Floyd-Steinberg's error diffusion, but less than the other two methods. However, the maximally exaggerated edges are less exaggerated with importance driven halftoning than with any of the other techniques. The bottom row of figure 5.18 illustrates the results in table 5.2 by graphically depicting the edge distortions on scale 4. These images show edges that have either been smoothed out or exaggerated from the original image. As shown, importance driven halftoning distorts edges along the black and white bands more than dispersed ordered dither, but similar to or less than other techniques. The edge distortion along the ramp with importance driven halftoning is similar to other techniques. In fact, some may say, that in these regions importance driven halftoning distorts edges slightly less than both clustered ordered dither and Floyd-Steinberg's error diffusion.

5.2.3 Summary of Photorealistic Quantitative Evaluation

Ulichney's measure of power spectrum and anisotropy both show that photorealistic importance driven halftoning results are very similar to Floyd-Steinberg's error diffusion results. The distribution of frequencies around the principle frequency for the chosen grey-scale intensity seem to be quite similar. Importance driven halftoning seems to be slightly more radially symmetric than Floyd-Steinberg's error diffusion on average.

The measure of edge artifacts proposed by Veryovka *et al* reveals that on scales 1 and 5 importance driven halftoning performs as well as Floyd-Steinberg's error diffusion, however on scales 2 to 4 the average edge artifacts introduced are stronger and more apparent, but the strongest edges introduced are weaker with importance driven halftoning than with Floyd-Steinberg's error diffusion or dispersed ordered dither. The measure of edge preservation reveals that importance driven halftoning distorts edges less than threshold matrix methods, but slightly more than Floyd-Steinberg's error diffusion. It also reveals that on average importance driven halftoning exaggerates the edges more than Floyd-Steinberg's error diffusion, but less than threshold matrix techniques.

In summary, importance driven halftoning has results similar to Floyd-Steinberg's error diffusion results both with the frequency distributions as well as with the introduction of edge artifacts and the preservation of edges. These results correlate with the results of our visual assessment in section 5.2.1

5.3 Chapter Summary

In this chapter the memory and computational needs of the importance driven halftoning process have been presented. The results of this process have also been analyzed both perceptually and quantitatively. Visually comparing the results of importance driven halftoning with other traditional halftoning methods such as threshold matrix and error diffusion methods, show importance driven halftoning to produce images similar to Floyd-Steinberg's error diffusion results with a slightly better preservation of high frequency information.

Importance driven halftoning has been evaluated quantitatively using two metrics: Ulichney's metrics of image frequency information in regions of fixed grey level and metrics of multi-scale edge analysis proposed by Veryovka *et al.* Ulichney's *power spectrum* and *anisotropy* measures both show that the results from importance driven halftoning are comparable to Floyd-Steinberg's error diffusion. The measures of *edge effects* proposed by Veryovka *et al* have shown that importance driven halftoning has similar edge artifacts as Floyd-Steinberg's error diffusion. The measures also showed that importance driven halftoning distorts edges in a similar manner to Floyd-Steinberg's error diffusion but on average exaggerating the edges rather than smoothing them as is common in error diffusion techniques.

Thus, importance driven halftoning produces photorealistic halftoned results similar to results of Floyd-Steinberg's error diffusion, however with the flexibility and tailorability of parameters a wide variety of other effects can be created using this technique. Also, since this technique is able to use a generic importance function to partially determine the distribution of ink, it is adaptable to new knowledge of the human visual system.

Chapter 6

Conclusions

This thesis presented a new halftoning technique called, importance driven halftoning. An overview of the halftoning literature and the basics of the human visual system was presented in order to provide motivation for the design of this new halftoning technique. The details and parameters of importance driven halftoning were discussed and various results were presented. Finally, importance driven halftoning was evaluated using both simple visual comparison and two quantitative metrics. The technique was found to produce results similar to Floyd-Steinberg's error diffusion, while allowing the user to retain control over the parameters of the technique. This not only allowed for the creation of traditional halftoning results, but also proved to be useful in the creation of limited resource and non-photorealistic rendering results.

6.1 Summary of Objectives

Previous halftoning techniques have been designed around a specific image attribute or for a specific display device. Initially, halftoning techniques were developed solely to preserve the reproduction of grey-scale intensities. As these techniques evolved new focuses developed such as edge (high frequency information) preservation, lack of unwanted artifacts, contrast etc... Regardless of the focus, the problem of halftoning has remained the same, namely:

1. How much ink or black pixels should be used?
2. How should this ink be arranged?

The proper amount of ink must be used in order to create the illusion of tone and the pixels of ink must be arranged in such a manner as to preserve image attributes (detail, edges, contours) and yet not create unwanted or false attributes that were not present in the original image. This thesis did not focus on determining the best amount or arrangement of ink, but presented a technique which allows the user to control both of these parameters for the creation of specific desired effects.

6.2 Technique Summary

Importance driven halftoning was designed to allow the user to specify important image attributes and hence identify which attributes should be preserved throughout the halftoning process. Keeping in mind that the human visual system works on multiple scales, our new halftoning technique uses a multi-scale or multiple resolution representation. This representation helps identify the presence of important attributes at multiple scales and allows for the preservation of these attributes over multiple scales. Importance driven halftoning allows the user to have control over three parameters:

1. importance function
2. number of display primitives
3. local arrangement of display primitives

The multi-resolution structure used in importance driven halftoning in conjunction with the three user controlled parameters dictate how much ink (black pixels) is used and how the ink is arranged in the image.

6.3 Results Summary

In this thesis various results obtained from altering the three user-controlled parameters were presented. The parameters were altered both individually and in combination with each other to illustrate their contribution to the results.

Different importance functions were tried including intensity, variance, gradient, and combinations of these three. Using intensity as the importance function yields results similar to traditional halftoning results. Using variance and gradient as the importance functions emphasizes the edges or regions of high contrast in the image. Allowing the user to specify a generic importance function, not only causes the halftoning to be tailorable to specific attributes, but also to weighted combinations of these attributes. The user may construct a weighted importance function combining several different importance attributes in order to create a rated precedence for the preservation of multiple image attributes. Combining intensity and gradient or variance as importance functions yields results similar to traditional halftoning results with edge or high frequency information enhancement.

Altering the number of display primitives presented the possibility of creating both limited resource rendered images and images with emphasized attributes as well as images similar to those created using traditional halftoning methods. Using intensity as the importance function and altering the number of display primitives results in a change in intensity proportional to the number of display primitives. This means an increase in the number of display primitives causes the entire image to become darker, but the darker regions become proportionally more dark than the lighter regions due to the percentage difference in intensity and the dependence on the number of display primitives. This may cause an increase in contrast as the number of display primitives reach excessive amounts. Thus, by altering the number of display primitives various effects, including contrast enhancement can be achieved. Having the number of display primitives be user-controlled not only allows for the creation of traditional halftoned images but also permits the creation of images using restricted resources or images with over enhanced or distorted attributes. The limited resource option can be useful for displaying images on devices where ink bleeds (*i.e.* laser printers) and hence less ink is desirable. This option can also be useful in the creation of draft copies of documents where the use of maximum resources are expensive and not necessarily required. Specifying an excess number of display primitives for the over-emphasis or distortion of image attributes can be used when alternate, non-photorealistic effects are desired.

Finally, it was shown that altering the type and shape of the display primitive can cause the creation of local effects or texture. The ink may be locally arranged on a pixel by pixel basis, or arrangements of ink in line segments or squares (clusters) may be chosen. Results were shown using scaled, thresholded, and rotated line segments as a display primitive. These adjustments to the display primitive were made both uniformly across the image and according to underlying image attributes. These results can be compared with typical pen-and-ink images. By using a threshold matrix as a display primitive results similar to traditional threshold matrix halftoning were created with the added ability of having control over the amount of ink used.

6.4 Evaluation Summary

Importance driven halftoning was evaluated by making visual comparisons with results from traditional threshold matrix and Floyd-Steinberg's error diffusion methods. Comparisons were made of such image attributes as regions of uniform, constant, increasing, and decreasing intensities, edges, high frequency regions, and contrast. By visually making these comparisons, importance driven halftoning was found to produce results similar or comparable with Floyd-Steinberg's error diffusion, with a slightly better preservation of high frequency information.

Importance driven halftoning was also evaluated using two quantitative metrics. The first was a measure of frequency distributions and symmetry within the halftoned image proposed by Ulichney. The second was a measure of the reproduction of edges or high frequency information over multiple scales proposed by Veryovka *et al.* Using importance driven halftoning with parameters for creating results similar to traditional halftoning methods, Ulichney's metrics were computed on various fixed levels of grey for the analysis of frequency information. Ulichney's metrics evaluated the power spectrum and the anisotropy (symmetry). Once again, importance driven halftoning was shown to yield results similar to Floyd-Steinberg's error diffusion for both of these measures. The measure of edge effects proposed by Veryovka *et al* revealed that the strength of edge artifacts introduced into the image over multiple scales is similar to the edge artifacts introduced by Floyd-Steinberg's error diffusion. It was also shown that importance driven halftoning preserves edges in the image better than threshold matrix techniques, and similar to, but with less smoothing of edges than error diffusion techniques. These quantitative results reaffirm the results obtained from our visual assessment.

In general, importance driven halftoning yields photorealistic results similar to traditional error diffusion results. However, due to the flexibility and tailorability of this technique images similar to pen-and-ink renderings or other non-photorealistic renderings as well as renderings useful in situations where limited resources are a benefit can be created. Several examples of images generated in this manner were presented.

6.5 Further Research

In this thesis a new halftoning method called importance driven halftoning was presented. The flexibility and tailorability of this technique allows for the creation of many different effects. Further research on this technique may involve the following:

1. The ability to set both black and white ink within a single image depending on the image intensity for the purposes of guaranteed clustering of ink (clustered ordered dither type approach).

2. Ability to possibly re-distribute excess display primitives to other regions of the image while still having guaranteed relative importance between attributes.

3. Incorporation of an *error diffusion type* approach with importance driven halftoning. This may involve calculating error at varying resolutions and distributing error within each resolution, during the distribution or re-distribution of display primitives.

4. Ability to progressively add ink to the image. This could allow for the alteration of importance functions and type of display primitive as ink is added (over different quantities of ink). Altering the display primitive during this process could produce results similar to post-processing edge enhancement presented by Buchanan *et al* [Buch98b].

5. Explore the possibility of reducing ink in unimportant areas of the image as well as adding ink to important attributes in as similar manner as (see [Schl96]).

6. Extend this technique to be useful in halftoning of 3D scenes as mentioned in [Haeb93].

Glossary

Artifacts:
: Unwanted connected components or high frequency image information.

Banding:
: The appearance of regular, usually false contours.

Blue noise:
: An arrangement of pixels that has the frequency of noise, but is less grainy than white noise.

Blue noise mask:
: A filter or mask used to create an underlying structure of blue noise.

Color:
: Light reflected from an feature, object or image.

Continuous tone:
: Containing an infinite or non-discrete number of tones.

Contour lines:
: Lines defining the shape of a feature or object.

Dither:
: In halftoning, the arrangement of pixels within an area, also known as spatial dithering.

Dot-gain:
: Overlap of one pixel into another. Common in bleeding of ink on laser printers.

Error diffusion technique:
: Halftoning techniques which diffuse quantization error between one region of the image and another.

Eight-connected:
: Connected to all eight neighboring pixels.

Halftone:
: The process of creating the illusion of continuous tone using two discrete tones (usually black ink on white paper).

Intensity:
: Magnitude of tone.

Lithograph:
: A print made using an image fixed on a stone or metal plate with ink absorbing or ink repelling vehicles.

Mezzotint:
: An art-form from the 1600's used to create the illusion of continuous tone images on copper plates using one tone of ink.

Non-photorealistic:
: Of non-photographic quality. Usually drawn, sketched or painted.

Non-photorealistic halftone:
: A halftoned image that is similar to the original continuous grey-scale image, but will have added artifacts and texture effects.

Photorealistic:
: Of photographic quality.

Photorealistic halftone:
: A halftoned image that appears as visually similar to the original continuous tone grey-scale image as possible.

Pixel:
: A picture element.

Quantization error:
: The difference between the original image intensity and the intensity approximated by the binary image.

Stroke textures:
: Textures created using line segments in different orientations, lengths, and density.

Threshold matrix:
: A matrix defining the order and magnitude at which to set pixels.

Threshold matrix technique:
: A halftoning technique involving the tiling of a threshold matrix over the image to determine which pixels to set to black and which to set to white.

Tone:
: Tint or shade of a color.

Weber Ratio:
: A ratio between the background intensity and change in intensity defining the minimum discernible change in intensity at various background illuminations.

Bibliography

[Anal92] Mostafa Analoui and Jan P. Allebach. "Model Based Halftoning using Direct Binary Search". *SPIE - International Society for Optical Engineering*, Vol. 1666, pp. 96–108, 1992.

[Bart55] M.S. Bartlett. *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*. Cambridge University Press, New York, 1955.

[Baye73] BE Bayer. "An optimum method for two level rendition of continuous-tone pictures". *Proceedings of IEEE International Conference on Communications*, Vol. 26, pp. 11–15, 1973.

[Bill83] C. Billotet-Hoffmann and O. Bryngdahl. "On the Error Diffusion Technique For Electronic Halftoning". *Proceedings of Society for Information Display*, Vol. 24, No. 3, pp. 253–258, 1983.

[Buch95] J. W. Buchanan and O. Verevka. "Edge preservation with space-filling curve half-toning". *Graphics Interface '95*, pp. 75–82, May 1995.

[Buch96] J. W. Buchanan. "Special Effects with Half-Toning". *Proceedings of Eurographics 96*, pp. 97–108, 1996.

[Buch97] John Buchanan and Lisa Streit. "Threshold-diffuse hybrid half-toning methods". *The Eighth Western Computer Graphics Symposium*, pp. 79–90, 1997.

[Buch98a] John W. Buchanan and Lisa Streit. "Variable Shape Halftoning". *IASTED International Conference on Computer Graphics and Imaging*, June 1998.

[Buch98b] John W. Buchanan, Lisa Streit, and Oleg Veryovka. "Edge Enhancement Issues in Half-Toning". *Graphics Interface 98*, June 1998.

[Chu92] C.H. Chu and W. Watunyuta. "On Designing Dither Matrices Using A Human Visual Model". *SPIE - International Society for Optical Engineering*, Vol. 1666, pp. 134–143, 1992.

[Cole91] A.J. Cole. "Halftoning without dither or edge enhancement". *The Visual Computer 1991*, pp. 232–245, 1991.

[Dalt95] John Dalton. "Perception of Binary Texture and Generation of Stochastic Halftone Screens". *SPIE - International Society for Optical Engineering*, Vol. 2411, pp. 207–220, 1995.

[Elbe95] G. Elber. "Line Art Rendering via a Coverage of Isoparameteric Curves". *IEEE Transactions on Visualization and Computer Graphics*, Vol. 1, No. 3, pp. 231–239, 1995.

[Floy76] R.W. Floyd and L. Steinberg. "An adaptive algorithm for spatial grey scale". *Proc. Soc. Inf. Display*, Vol. 17, pp. 75–77, 1976.

[Fole90] J.D Foley, van A. Dam, and J.F Hughes. *Computer Graphics Principles and Practice*. Addison-Wesley Publishing Company, New York, 1990.

[Gonz92] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Publishing Company, New York, 1992.

[Grog92] Timothy A. Grogan and David Keene. "Image Quality Evaluation with a Contour-Based Perceptual Model". *SPIE - International Society for Optical Engineering*, Vol. 1666, pp. 188–197, 1992.

[Haeb93] Paul Haeberli and Mark Segal. "Texture Mapping As A Fundamental Drawing Primitive". *Fourth Eurographics Workshop on Rendering*, pp. 259–266, June 1993.

[Hilg94] D. Hilgenberg, T. Flohr, C. Atkins, J. Allebach, and C. Bouman. "Least-Squares Model-Based Video Halftoning". *SPIE - International Society for Optical Engineering*, Vol. 2179, pp. 207–217, 1994.

[Holl80] Thomas M. Holladay. "An Optimum algorithm for Halftone Generation for Displays and Hard Copies". *Proceedings of the S.I.D.*, Vol. 21, No. 2, pp. 185–192, 1980.

[Hsu94] S. Hsu and I. Lee. "Drawing and Animation Using Skeletal Strokes". *(SIGGRAPH '94 Proceedings)*, pp. 109–118, 1994.

[Hurd68] Peter Hurd. *Peter Hurd, The lithographs*. The Baker Gallery Press, P.O. Box 1920, Lubbock Texas, 1968.

[Jarv76a] J.F. Jarvis, C.N. Judice, and W.H. Ninke. "A survey of techniques for the display of continuous tone pictures on bilevel displays". *Computer Graphics and Image Processing*, Vol. 5, pp. 13–40, March 1976.

[Jarv76b] J.F. Jarvis and C.S. Roberts. "A new technique for displaying continuous tone pictures on bilevel displays". *IEEE Transactions on Communications*, Vol. COM-24, pp. 891–898, 1976.

[Knox93] Keith T. Knox. "Error diffusion: a theoretical view". *SPIE - International Society for Optical Engineering*, Vol. 1913, pp. 326–331, 1993.

[Knox94] Keith Knox. "Threshold modulation in error diffusion on non-standard rasters". *SPIE - International Society for Optical Engineering*, Vol. 2179, pp. 159–169, 1994.

[Knut87] D. E. Knuth. "Digital halftones by dot diffusion". *ACM Transactions on Graphics*, Vol. 6, No. 4, pp. 245–273, October 1987.

[Kolp92] B.W. Kolpatzik and C.A. Bouman. "Optimized Error Diffusion Based on a Human Visual Model". *SPIE - International Society for Optical Engineering*, Vol. 1666, pp. 152–164, 1992.

[Land95] J. Landsdown and S. Schofield. "Expressive Rendering: A review of Non-photorealistic Techniques". *IEEE Computer Graphics and Applications*, pp. 29–37, May 1995.

[Lin93] Qian Lin. "Halftone image quality analysis based on a human vision model". *SPIE - International Society for Optical Engineering*, Vol. 1913, pp. 378–389, 1993.

[Mall89] Stephane G. Mallat. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation". *Proceedings of IEEE*, Vol. 11, No. 7, pp. 674–693, July 1989.

[Mits92] Throphano Mitsa and Kevin Parker. "Digital halftoning using a blue noise mask". *Journal of Optical Society of America*, Vol. 9, No. 11, pp. 1920–1929, 1992.

[Mits93] Theophano Mitsa, Krishna Lata, and Jennifer R. Alford. "Frequency-channel-based visual models as quantitative quality measures in halftoning". *SPIE - International Society for Optical Engineering*, Vol. 1913, pp. 390–401, 1993.

[Mits95] T. Mitsa and P. Brathwaite. "Wavelets as a tool for the construction of halftone screens". *SPIE - International Society for Optical Engineering*, Vol. 2411, pp. 228–237, 1995.

[Mull92] Jeffery B. Mulligan and Albert J. Ahumada Jr. "Principled Halftoning Based on Human Vision Models". *SPIE - International Society for Optical Engineering*, Vol. 1666, pp. 109–121, 1992.

[Naim96] A.C. Naiman and D.T.W Lam. "Error Diffusion: Wavefront Traversal and Contrast Considerations". *(SIGGRAPH '96 Proceedings)*, Vol. 19, No. 1, pp. 78–86, 1996.

[Ostr94] V. Ostromoukhov, R. D. Hersch, and I. Amidror. "Rotated Dispersed Dither: A new Technique for Digital Halftoning". *Computer Graphics (SIGGRAPH '94 Proceedings)*, pp. 123–130, July 1994.

[Ostr95a] V. Ostromoukhov and R. Hersch. "Halftoning by Rotating Non-Bayer Dispersed Dither Arrays". *SPIE - International Society for Optical Engineering*, Vol. 2411, pp. 180–197, 1995.

[Ostr95b] V. Ostromoukhov and R. D. Hersch. "Artistic Screening". *Computer Graphics (SIGGRAPH '95 Proceedings)*, pp. 219–228, August 1995.

[Papp92] Thrasyvoulos N. Pappas and David L. Neuoff. "Least-Squares Model-Based Halftoning". *SPIE - International Society for Optical Engineering*, Vol. 1666, pp. 165–176, 1992.

[Perl95] K. Perlin and L. Velho. "Live Paint: Painting with Procedural Multiscale Textures". *Computer Graphics (SIGGRAPH 1995)*, pp. 153–160, August 1995.

[Pnue94] Y. Pnueli and A. Bruckstein. "$Dig_D^i urer$ - a digital engraving system". *IEEE Computer Graphics and Applications*, Vol. 10, pp. 277–291, 1994.

[Purg94] Werner Purgathofer, Robert F. Tobler, and Manfred Geiler. "Forced Random Dithering: Improved Threshold Matrices for Ordered Dithering". *IEEE. Proceedings, ICIP-94: 1995 IEEE International Conference on Image Processing*, Vol. 2, pp. 1032–1035, Nov 1994.

[Rodr94] Michael Rodriguez. "Graphic Arts perspective on digital halftoning". *SPIE - International Society for Optical Engineering*, Vol. 2179, pp. 144–149, 1994.

[Sait90] T. Saito and T. Takahashi. "Comprehensible Rendering 3-D Shapes". *Siggraph 1990*, Vol. 24, No. 4, pp. 197–206, 1990.

[Sali94] M. Salisbury, S. Anderson, R. Barzel, and D. Salesin. "Interactive Pen-and-Ink Illustration". *(SIGGRAPH '94 Proceedings)*, pp. 101–108, 1994.

[Schl96] Stefan Schlechtweg and Thomas Strothotte. "Rendering Line-Drawings with Limited Resources". *Proceedings of GRAPHICON '96 - Sixth International Conference and Exhibition on Computer Graphics and Visualization in Russia*, Vol. 2, pp. 131–137, July 1996.

[Schu94] Mark Schulze. "Blue noise and model-based halftoning". *SPIE - International Society for Optical Engineering*, Vol. 2179, pp. 182–194, 1994.

[Sloa92] K.R. Sloan and Y. Wang. "Exploiting texture effects in dithering". *SPIE - International Society for Optical Engineering*, Vol. 1666, pp. 144–151, 1992.

[Sloa93] K.R. Sloan and D. Campbell. "Texture as information". *SPIE - International Society for Optical Engineering*, Vol. 1913, pp. 344–354, 1993.

[Snyd85a] Harry L. Snyder. "Image Quality: Measures and Visual Performance". *Flat Panel Displays and CRT's*, pp. 70–90. Van Nostrand Reinhold, New York, 1985.

[Snyd85b] Harry L. Snyder. "The Visual System: Capabilities and Limitations". *Flat Panel Displays and CRT's*, pp. 54–69. Van Nostrand Reinhold, New York, 1985.

[Stol95a] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin. "Wavelets for Computer Graphics: A Primer, Part1". *Computer Graphics and Applications*, Vol. 15, No. 3, pp. 76–84, 1995.

[Stol95b] Eric J. Stollnitz, Tony D. DeRose, and David H. Salesin. "Wavelets for Computer Graphics: A Primer, Part 2". *Computer Graphics and Applications*, Vol. 15, No. 4, pp. 75–85, 1995.

[Stol96] Eric Stollnitz, Tony Derose, and David Salesin. *Wavelets for Computer Graphics*. Morgan Kaufmann Publishers Inc., San Fransico, California, 1996.

[Stre98] Lisa M. Streit and John W. Buchanan. "Importance Driven Halftoning". *Eurographics 98*, August 1998.

[Thur94] Stefan Thurnhofer. "Nonlinear Detail Enhancement of Error-Diffused Images". *SPIE - International Society for Optical Engineering*, Vol. 2179, pp. 170–181, 1994.

[Ulic87] R. Ulichney. *Digital Halftoning*. MIT Press, Cambridge, Massachusetts, 1987.

[Ulic88] R. A. Ulichney. "Dithering with blue noise". *Proc. IEEE*, Vol. 76, No. 1, January 1988.

[Ulic93] R. Ulichney. "The void-and-cluster method for dither array generation". *SPIE - International Society for Optical Engineering*, Vol. 1913, pp. 332–343, 1993.

[Velh91] L. Velho and J. de Miranda Gomes. "Digital halftoning with space filling curves". *Computer Graphics (SIGGRAPH '91 Proceedings)*, Vol. 25, No. 4, pp. 81–90, July 1991.

[Velh95] L. Velho and J. Gomes. "Stochastic Screening Dithering with Adaptive Clustering". *Computer Graphics (SIGGRAPH '95 Proceedings)*, pp. 273–276, August 1995.

[Very98] Oleg Veryovka, Alain Fournier, and John Buchanan. "Multi-scale Edge Analysis for halftoned images". *Proceedings, SPIE—The International Society for Optical Engineering: Human Vision, Visual Processing, and Digital Display VIII*, January 1998.

[Wert39] M. Wertheimer. "Laws of Organization in Perceptual Forms". *A Source Book of Gestalt Psychology*. Harcourt Brace, New York, 1939.

[Will83] L. Williams. "Pyramidal parametrics". *Computer Graphics (SIGGRAPH '83 Proceedings)*, Vol. 17, No. 3, pp. 1–11, July 1983.

[Wink94] G. Winkenbach and D. Salesin. "Computer-Generated Pen-and-Ink Illustration". *(SIGGRAPH '94 Proceedings)*, pp. 91–100, 1994.

[Wink96] G. Winkenbach and D. Salesin. "Rendering Parametric Surfaces in Pen and Ink". *(SIGGRAPH '96 Proceedings)*, pp. 469–475, 1996.

[Witt82] I.H. Witten and R.M. Neal. "Using Peano curves for bilevel display of continuous-tone images". *IEEE Computer Graphics and Applications*, Vol. 2, No. 3, pp. 47–52, May 1982.

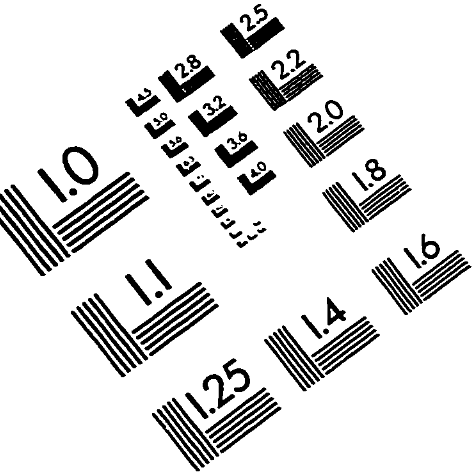
[Wong95] Ping Wah Wong. "Error Diffusion with Delayed Decision". *SPIE - International Society for Optical Engineering*, Vol. 2411, pp. 198–206, 1995.

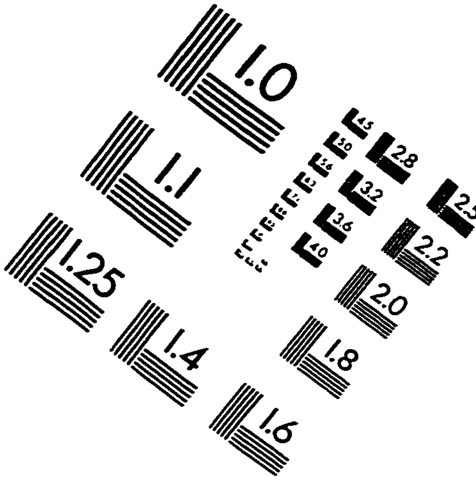
[Yao95] Meng Yao and Kevin J. Parker. "Dot gain compensation in blue noise mask". *SPIE - International Society for Optical Engineering*, Vol. 2411, pp. 221–227, 1995.

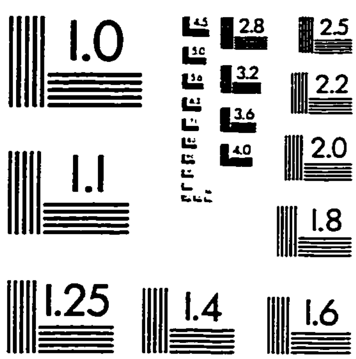
[Zakh92] Avideh Zakhor, Steve Lin, and Farokh Eskafi. "A New Class of B/W Halftoning Algorithms". *SPIE - International Society for Optical Engineering*, Vol. 1666, pp. 122–133, 1992.

[Zhan93] Y. Zhang and R. E. Webber. "Space diffusion: An improved parallel halftoning technique using space-filling curves". *Computer Graphics (SIGGRAPH '93 Proceedings)*, pp. 305–312, August 1993.

IMAGE EVALUATION TEST TARGET (QA-3)

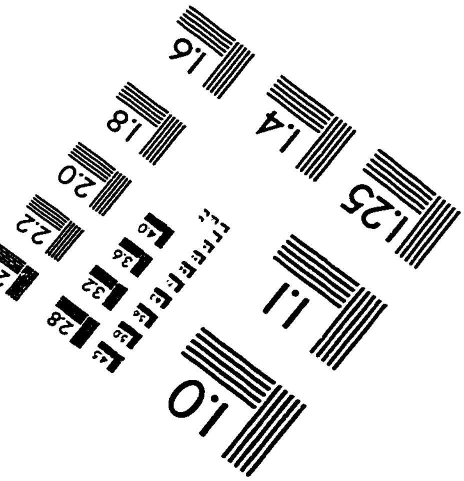


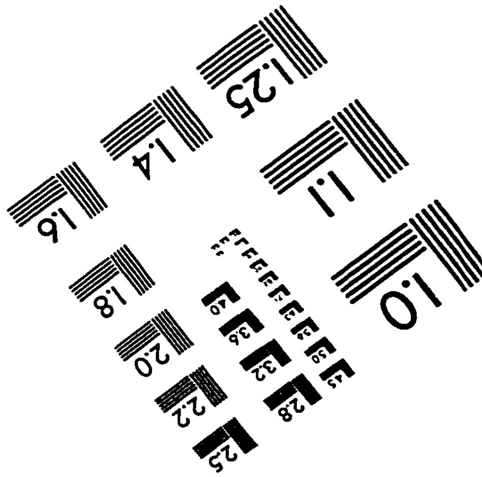




150mm

6″





APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989