8-1-2000

Adobe Acrobat image compression: An Investigation into the effects of compression in Acrobat 4.0 on image reproducibility for digital printing

Kelly Thornton

Adobe Acrobat Image Compression:

An Investigation into the Effects of Compression in Acrobat 4.0 on Image Reproducibility for Digital Printing

by

Kelly E. Thornton

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the School of Printing Management and Sciences in the College of Imaging Arts and Sciences of the Rochester Institute of Technology

August, 2000

Thesis Advisor: Professor Frank Romano

School of Printing Management and Sciences
Rochester Institute of Technology
Rochester, New York

Certificate of Approval

Master's Thesis

This is to certify that the Master's Thesis of

Kelly Elizabeth Thornton

With a major in Graphic Arts Publishing/ Electronic Publishing
has been approved by the thesis committee as satisfactory
for the thesis requirement for the Master of Science degree
at the convocation of

August, 2000

Thesis Committee:

Frank Romano
Thesis Advisor

Marie Freckleton
Graduate Program Coordinator

Frank Romano
Chair

Adobe Acrobat Image Compression:

An Investigation into the Effects of Compression in Acrobat 4.0 on Image Reproducibility for Digital Printing

I, Kelly Elizabeth Thornton, hereby grant permission
to the Wallace Memorial Library of R.I.T.
to reproduce my thesis in whole or in part.

_____

Kelly E. Thornton

August, 2000

Acknowledgements

I would like to extend a special thanks to my advisor, Professor Frank Romano, who was a great help throughout the process of writing this thesis. I would also like to thank Professor Marie Freckleton, Program Coordinator, for her advice and assistance in this project.

Table of Contents

List of Tables

List of Figures

Glossary of Terms

DCT—(Discrete Cosine Transform) A function of JPEG compression which converts spatial information to mathematical frequency information in a matrix.

Display List—an ordered list of objects in a page description file. The order of objects in a display is the order in which they are displayed or imaged by the screen/printer.

CCITT—(International Coordinating Committee for Telephony and Telegraphy) compression method is appropriate for black-and-white images made by paint programs and any images scanned with an image depth of one bit.

GIF—(Graphic Image Format) a file format option for storage of digital images.

LED—(Light Emitting Diode) used to image photoconductors in some electrophotographic printing devices.

Rasterization—the process of converting graphics described at a high level into bitmaps for rendering on a monitor or digital device

RLE—(Run Length Encoding) a method of encoding binary images as a string of codes representing successive runs of image or non-image pixels for the purpose of compression.

ZIP—a compression method that works well on image with large areas of single colors or repeating patterns such as screen shots and simple images created with paint programs, and for black-and-white images that contain repeating patterns.

Abstract

Trends in printing are moving toward more convenient and composite workflows. The printing industry as we know it is undergoing major changes in the way it provides printing services. Technology is manifesting itself in the industry through digital printing.

In an industry that is becoming increasingly digital and focused on generating and receiving information efficiently, time is an issue and file size is a concern. Digital workflow is being adopted at an incredible rate, and information must be stored and delivered in the smallest, most convenient formats. Electronic file formats have evolved to serve these purposes. One technology that enables the transition to digital workflow is a format called PDF, Portable Document Format.

There are many aspects of PDF that should be understood to render a quality file that in turn renders a quality digital print. One such aspect is file or data compression. Acrobat 4.0, the software that produces PDFs, is brand new technology which lends even more potential to successful digital workflows. Acrobat 4.0 offers a number of different ways to reduce the size of a file through methods of compression in Distiller, a program that translates a PostScript file into a PDF. A user can choose the type of compression for their file within Distiller. However, Distiller job options have changed since Acrobat 3.0.

A choice of compressions is offered to satisfy requirements for different types of graphics. Unless a person is very knowledgeable about data compression, or has researched the Distiller parameters for compression, that individual may be confused as to which setting will result in the best printing of their file.

The aim of this project is to investigate each compression setting in Acrobat 4.0 Distiller by using and comparing the effects of each on a digitally printed file. The goal is to define the visual results these compressions will yield according to the components of a given file.

The results of the research will be valuable in terms of dispelling confusion regarding the compression settings in Acrobat 4.0. Use of this research could improve the reproducibility and overall quality of images in PDF form for digital printing.

Chapter One

Introduction

Background and Significance

Use of digital workflow in the printing industry is inevitable. With the power of the internet growing daily, companies are moving towards digital workflow to maintain a competitive edge. The printing industry is not excluded from this digital revolution.

In an industry that is becoming increasingly digital and focused on generating and receiving information efficiently, time is an issue and file size is a concern. The distribution of information is crucial to the graphic arts. Today, more than ever, information must be stored and delivered in the smallest and most convenient formats. Electronic file formats have been evolving to meet these needs.

Adobe Systems offers software that makes the transition to digital workflow simple with two powerful products: PostScript and Acrobat.

PostScript

PostScript is a page description language. It is a code that explains what is on a given page. PostScript uses a "grid coordinate system"[1] to describe the elements of a page, identifying each by their mathematical location on the page. PostScript has become a standard; almost all printers or output devices are driven by this language. Therefore, PostScript is device independent—it is used on all different platforms, with different resolutions and capabilities.

> Like any spoken language, PostScript has a vocabulary. There are commands that control the size of type, the tint of colors or the position of pictures. Virtually every application program

> running in every desktop computer outputs PostScript and virtually every printer of every type, including some sophisticated prepress and even press systems, accept PostScript-coded files.[2]

PostScript is essentially coded text that describes images, graphics, and text on a page. It requires a RIP(Raster Image Processor), an interpreter to display the information that is coded. A RIP interprets the coded pages of PostScript and translates them into the finished and printed product. PostScript code was really never meant to be "seen" by the user. It was intended to be "a behind-the-scenes language for communication between a computer and printer."[3]

> PostScript refers to both the page description language . . . and the interpreter that converts the page description into pixels or bits to control a raster-based output device. As a general purpose programming language, PostScript contains procedures, variables and control constructs that may cause unpredictabilty. The PostScript file has a main fault: it cannot be viewed. [4]

The RIP

"A RIP is the essential element in any form of raster-based imaging which includes computer-to: paper, film, plate, cloth, plastic, metal, and perhaps epidermis."[5] A RIP is the "middle man", a device between the computer and printer, processing a file into a bitmap so it may be printed.

The RIP performs three functions:

1. Interpretation of the page description language from the application program.
2. Display list generation
3. Rasterizing (screening, color transforms, and making the page bitmap)[6]

Portable Document Format

PDF, by Adobe Systems, is a type of "document converted to a special coded file that can be displayed and/or printed on a Macintosh or Windows-based PC without the original application program."[7] Adobe Acrobat's PDF can significantly reduce the size of a file—a very important factor in digital workflow. "Introducing PDF to your workflow can streamline your printing process."[8] "PDF is quickly setting the standard for consistency and reliability in the print production workflow. This translates in to real benefits, including bet-

ter communication, fewer delays, and lower production costs."[9]

Portable Document Format was created to solve the problems arising from transporting electronic documents over a network like the internet.

> By creating an electronic document that carries all the needed components—fonts, graphics, and even a program to view and print the document—portable document software could eliminate the . . . distributing, and storing paper copies, while adding the ability to find text and link multiple documents do information would be more accessible and more dynamic.[10]

It is only recently that the benefits to print are being realized and taken advantage of.

Adobe's PDF is on its way to becoming a de facto standard for the printing industry much like its predecessor, PostScript. "PDF was a third version of the PostScript file format."[11] PDF takes PostScript to higher level than simple page description by *distilling* or translating PostScript to a different format which shrinks the file size through compression and removes variability from the file.[12] These features, among others, make a PDF much easier to transport and store. Even though a document that has been distilled in to a PDF is essentially "clean code", PDF can not be sent straight to a RIP yet. A PDF must be converted back to a PostScript file before it can be sent to most digital printing presses. This is somewhat inconvenient seeing as the file was made into a Postscript before it was distilled. This set-back will most likely be remedied in the near future. Scitex and Agfa already have RIPs that can handle PDF directly. The combination of PDF and digital printing brings the possibility of true digital workflow to the printing industry. Having the capability for digital workflow helps ensure the future of print in this technological and digital society.

The Statement of the Problem

True digital workflow uses completely digital means and eliminates many steps from a traditional processes. Digital file formats help to simplify workflow. Adobe Acrobat 4.0 is a very useful tool for their implementation. Acrobat allows for the creation of PDF—resulting in documents that are completely portable and platform independent.

The reason for the portability of these documents is PDF's ability to compress a file and reduce its size. Compression simply means the removal of data or reduction in the amount of data in an electronic file. Acrobat 4.0 offers a user options for the compression of a file: "automatic" settings (defaults) or individually adjusted settings to fit the file. Options include compression and quality for three different types of images and graphics: color, grayscale and monochromatic. When an individual creates a PDF, he or she typically wants the smallest possible file with the best possible quality. People are often skeptical about the quality of a printed PDF, equating the reduction in file size to a reduction in quality.

Acrobat 4.0 is the latest version of Adobe's PDF software. Distiller offers the same basic features, but 4.0 has streamlined the dialogue box. Figures 1 and 2 show the basic differences between Acrobat Distiller 3.0 and 4.0 compression options. Acrobat 4.0 offers a variety of downsampling while 3.0 does not(downsampling is reduction of the number of dots per inch in an image). Next, 4.0 has combined the choices of compression into one button instead of two and offers JPEG or ZIP compressions if the user wants the choice. "Automatic" compression is still offered. Last, 4.0 gives the user a choice of quality for a file. The settings range from minimum to maximum, referring to the amount of compression being applied, minimum meaning the most compression applied, maximum meaning the least.

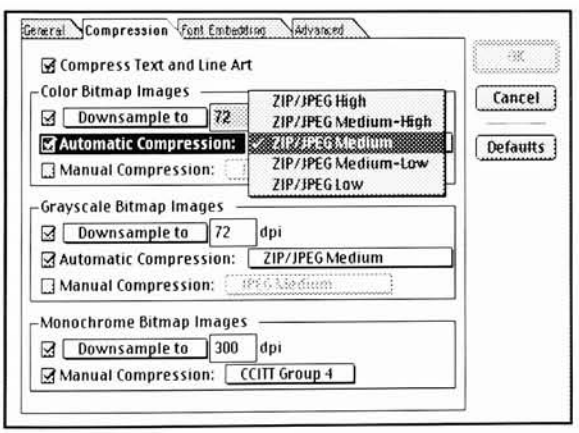


Figure 1 (Acrobat Distiller 3.0)

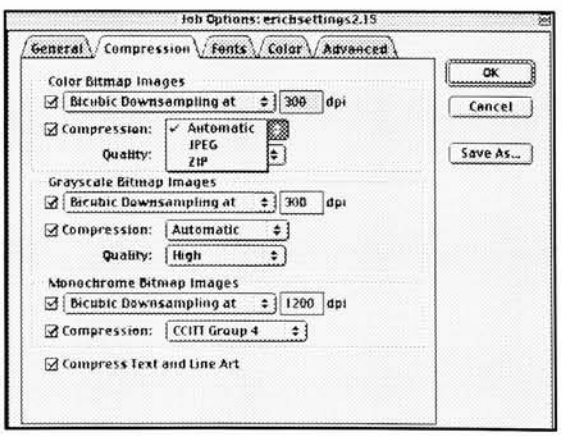


Figure 2 (Acrobat Distiller 4.0)

If compression is not suited for a file or is over-applied, the quality of a file can suffer. Many people might assume that the automatic settings are appropriate for their file, but there may be a setting which produces better results. There seems to be an assumption that anyone using the program will know which settings to use for their particular file and will know what to expect when the job is printed or will take the time to look it up.

The goal of this project is to test all the compression settings in Acrobat 4.0 and attempt to dispel confusion regarding their effects. The aim is to define what the individual settings mean and what effect they have on digitally printed images and to determine which compression settings are needed for various types of images and graphics for a quality outcome.

Reason for Interest

The printing industry is undergoing major changes. The onset of digital printing systems is revolutionizing the way information is reproduced. PDF technology has the potential to be a catalyst for change. Today, PDF is being utilized by those in the printing industry with a pioneering spirit. Perhaps someday PDF will be a standard.

PDF can make a job simple and hassle-free when the file is prepared correctly (a good PostScript file). If there are mistakes made in the distilling process, problems may arise when the file goes to print. This is why it is so important to understand the job options in Acrobat Distiller, particularly compression. Not everyone, though well-trained in the graphic arts, is an expert in data compression. Somehow demonstrating the effects of the compression settings could make the program more user-friendly.

Endnotes for Chapter One

[1] Michael H. Bruno. *Pocket Pal. A Graphic Arts Production Handbook.* (Memphis: International Paper. 1997.), 106

[2] ibid, 106

[3] Mattias Andersen, et al. *PDF Printing and Publishing. 2nd Edition. The Next Revolution After Gutenberg. Adobe Acrobat for Digital Prepress Systems.*(Torrance, CA:Micro Publishing Press. 1998.), 4.

[4] ibid, 4

[5] Michael H. Bruno. *Pocket Pal. A Graphic Arts Production Handbook.* (Memphis:International Paper. 1997.), 107

[6] ibid, 107

[7] Adobe Systems Incorporated. "PDF for Prepress Workflow and Document Delivery." *Adobe Printing Solutions.* 1997, 17.

[8] Adobe Systems. "Creating high-quality PDF files with Adobe Acrobat for print & press" *How to Create Adobe PDF Files For Print and Press.* 1999, 4.

[9] ibid, 2.

[10] Mattias Andersen, et al. *PDF Printing and Publishing. 2nd Edition. The Next Revolution After Gutenberg. Adobe Acrobat for Digital Prepress Systems.*(Torrance, CA:Micro Publishing Press. 1998.), 20.

[11] ibid, 21.

[12] ibid, 21.

Chapter Two

Background Theory

This project deals with the creation of files that are portable and considerably smaller in size than an original application file. Why are such files necessary? The internet has been the driving force behind the development of portable documents. However these types of files are becoming more and more important to the printing industry, an industry which for years has depended on traditional and mechanical, not technological, processes. Two reasons that the printing industry is beginning to notice and utilize the technology of the portable document format is the growing use of digital printing systems and the popularity of print on-demand. As digital printing is accepted by printing companies, PDF will become established in the printing industry. PDF is an ideal file format for digital printing systems.

> As we make the transition from analog to digital production,we face a formidable puzzle of digital components and procedures, with many of the pieces awkwardly arranged in precarious unpredictable sequences. . . . Whether we screen print or output to digital-color devices, the efficiency and accuracy of the graphics production depends largely on how we manipulate and move digital files throughout the process.[1]

This project will investigate the compression of Acrobat 4.0 to show how it affects image reproducibility in digital printing systems. Therefore the topics of digital printing, Adobe Acrobat and PDF, digital image formation and compression will be discussed to give a the reader a complete overview.

Digital Printing

The Digital Revolution

The growth of the internet is driving the expansion and creation of markets. The perceived "threat" of technology to the printing industry stems from the internet. One particular

area of internet technology that takes business from print is electronic commerce—business transactions occur on-line rather than on paper. Technology threatens the printing industry, but ironically, printers will use it to counterattack the threat. Digital printing will ensure the future of print.

Traditional printing is usually referred to as the use of lithographic, flexographic, gravure, and screen processes. Companies have begun to accept new technology and shifting their processes from traditional to electronic/digital—either partially or completely. Companies are gradually realizing that to stay competitive in today's markets, they must adopt the technologies available to them.

> The evolution of distribute print on-demand and communication networks is changing the way the world thinks about print and publishing. As requests for shorter and shorter print run-lengths increase, digital on-demand color printing equipment is coming to the forefront to meet the industry's demands.[2]

What is Digital Printing?

"Digital Printing is a printing method that transfers digitized images and text from the computer directly to a digital printing press, using lasers or LED's to etch or electronically image special plates or drums."[3] Digital printing can be divided into two categories: direct to an image carrier and direct to paper. This can also be classified as computer-to-plate and computer-to-print (computer-to-press).[4] Systems are considered truly digital when they are variable, meaning that the image area can be changed for each impression made.[5]

Direct to Image Carrier		Direct to Paper	
Re-imageable Plate	Fixed Image Plate	Plain Paper	Special Paper
Electrophotography	Direct-to-Plate	Ink Jet(continuous)	Electrostatic
Ion Deposition	DTP on Press	Drop on demand	Electrophotographic
Magnetography	Direct-to-stencil(on press)	Thermal Transfer	Diffusion Transfer

Table 1 (Frank Romano. *Pocket Guide to Digital Prepress*)

The following are some characteristics of digital printing systems. Digital printing offers very *quick turnaround (on-demand)*—jobs can be completed in five minutes if the run is short enough, *flexibility*—changes can be made to the job on the fly, and *short runs*—the ultimate short run of one reproduction can be made. There are, of course, limitations such as the quality of a digital print compared to an analog print. It is not quite as good—*yet*.

Digital Technology

> Digital printing presses combine the imaging lasers or LED's with new types of plates, drums and inks . . . . Not only do digital presses let you image your documents directly on the press itself, but they also automate document management and make-ready, eliminating the time consuming and costly preparation and calibration of film, plates and water-based inks. [6]

There are five basic types of technology employed in digital printing systems: Electrophotography, Ion Deposition, Magnetography, Ink Jet, and Thermal Transfer. These can also be referred to as electronic printing systems—the image area can be changed from impression to impression. The digital printing system utilized in this project's experiment, the Indigo E-Print, is an electrophotographic process.

> Electrophotographic printing systems use drums or metal plates coated with photoconductors that are charged overall with a corona charge, exposed to an image by light which discharges the charge in the non-image areas, the image is toned by powder or liquid toners in the charged areas, transferred to a substrate and fused or fixed by heat, solvent vapor or other fixing method. [7]

Electrophotography uses the same principles as electrostatic copiers. A charge is used to attract toner to an image area, and then the toner is fused to make it stay on the paper. The most familiar technology to grow out of electrophotography is Xerography.

Electrophotography seems to be the most widely used form of technology in digital printing systems today. The Indigo E-Print, Xiekon DCP, and T/R Systems are all examples of popular electronic printing systems employing its technology.

Applications of Digital Printing

There are new markets being created every day for digital printing systems. The most popular applications are *variable information printing, on-demand printing, and short-run four color (process) printing.*

Variable is the fastest growing market application for digital printing. How many times have you received materials from Reader's Digest or Publisher's ClearingHouse personally addressed to you? "Kelly Thornton, you may have won One Million Dollars!" Variable printing allows for personalized and direct mail, arguably the most effective form of advertising with the most returns. "The ability to print variable information is a unique, exclusive, and important characteristic of digital printing systems."[8] "The ability to customize individual pages with images and text in a single run also opens up numerous possibilities for direct mail campaigns, personal marketing approaches and targeted documents." [9]

On-demand printing is also a growing market. On-demand makes it possible to have a book or publication just in the time you need it. No more ordering far in advance. Using electronic databases to store documents reduces warehousing and inventory costs and keeps documents at everyone's disposal. Another nice feature of on-demand is the combination of printing and finishing in-line in one process that is offered. [10]

On-demand is a market that was created by digital printing systems. Traditional printing (analog processes) does not have the capability of producing an entire book on a moment's notice. Think of all the time that would be spent getting the plates made, doing makeready and bringing the press up to color. It is not feasible or economical. On-demand has a distinct advantage over analog printing in this department.

Short-run four color printing jobs are being snapped up by those companies with digital printing systems. According to author Frank Romano, "Almost 56% of commercial, book and office printing including duplicating and copying is in the category of run lengths from 500 to 5000 impressions, only 2.8% of all this printing is done in four or more colors."[11] This means that half the printing jobs out there belong to digital printing systems. There will continue to be jobs with long run lengths which will call for traditional printing processes. It would not be economical to do a long run on a digital press. However, digital printing systems offer an alternative source for printing short runs which are *not*

economical on the traditional presses.

Digital Image Compression

"Compression is nothing more—or less—than efficient coding designed to correct the overrepresentation that occurs in digital data-handling systems." [12] "If you have ever used a TIFF file or downloaded a GIF file . . . you have experienced the use of image compression." [13] Dealing with image compression requires an understanding of graphics—how they are created and how they are stored.

Digital Images

When an image, a representation of a person, animal or thing is digitized, converted to digital information, it is known as a digital image. "Based on the accuracy of the representation, we can classify images into three categories: black and white images, grayscale images, and color images." [14]

> Despite the advantages, there is one potential problem with digital images, namely, the large number of bits required to represent them. Fortunately, digital images, generally contain a significant amount of redundancy. Image compression research aims to reduce the number of bits required to represent an image by removing these redundancies. [15]

"Graphic file formats are based upon the method by which graphic programs create, store, and display images. Graphic programs and their resulting file formats can be subdivided into one of two categories—raster and vector."[16] A raster can also be referred to as a bitmap.

A bitmap is the most basic type of graphic. The word bit is a contraction of the words *binary* and *digit*. There are two binary digits: 0 and 1. Bitmaps are representations of graphics or images through the use of bits (0 and 1). 0's are used to represent image area, and 1's represent non-image area. This means that bitmaps only have two tones: black and white. These binary digits make up the smallest elements of a picture, or pixels. Pixels are joined and built up in rows to create the overall picture. [17]

> Black and white bitmaps need only one bit to describe each pixel—zero or one. A single bit per pixel doesn't provide

> enough information to specify a color or shade of gray. Images containing 256 grays or colors require 8 bits per pixel, and photographic-quality full-color images require 24 bits per pixel. . . . Depending on resolution, color, and grayscale, bit maps require significant volumes of storage. [18]

Vectors are graphics or images that are "mathematically described objects or paths." [19] According to Frank Romano in the *Pocket Guide to Digital Prepress*, vectors are like lots of electronic rubber bands describing the outline of an image, allowing you to enlarge, reduce, rotate, reshape, and refill objects with no loss in quality. Vector graphic programs are "smart," keeping track of every adjustment that is made to an object/ image. This type of graphic is very convenient for printing as well. Since an object is mathematically described, the printer can render it at the highest possible resolution instead of trying to find where all the pixels must be placed. [20]

Digital Image Formation

Images can be made digital through a process called scanning. Scanning takes samples of a picture, stopping at intervals in an image to collect data. The number of intervals is referred to as the sampling rate. The higher the sampling rate of a scanner, the more information is collected from an original. For each sample taken, a digital value is assigned. For a black and white original, a byte, 8 bits, is assigned to each sample. A byte allows 256 levels of brightness in the original photograph to be captured. In color scanning, 3 bytes are assigned for each sample, due to the three RGB components in each pixel. [21]

The dynamic range of an original, meaning how many tones it contains, will dictate how many bits per pixel are required. Logically, a larger tonal range requires more bits per pixel. "The digital value of each pixel represents some measure of the brightness, color, or spectral attributes sensed at that point in the image." [22] A scanner takes a certain number of samples in a horizontal direction in consecutive rows until the whole original has been covered. The number of samples taken in a row by the number of rows sampled creates an array in which to store the information digitally. The larger the dynamic range of an original, the higher the sampling rate has to be in order to capture the necessary information.

Digital images can also be created by software programs referred to as drawing programs. Original art in vector form can be created in Adobe Illustrator or FreeHand by

MacroMedia. No scanning is involved in using these programs because they are automatically digital, having been created in the computer's native language without the need for code translation.

Digital Image Storage

Once an image is made digital in order to be used, it must be stored somehow for easy access. Images and graphics are stored in graphic file formats. There are a number of file formats used in the market today, but there are three that are the most widely used: TIFF, EPS, and JPEG. [23] Most graphic file formats offer some sort of compression to reduce the size of an otherwise large file.

TIFF, Tagged Image File Format, is a format for storing pixel-based images. TIFF offers a compression called LZW which can reduce the file size by half and retain the quality of the images.

EPS, Encapsulated PostScript, is used for bitmapped and vector-based images. EPS is adaptable to a file, with the ability to have 48 bits per pixel if required. EPS therefore can represent high-quality photographs without any problem.

JPEG, named for the Joint Photographic Experts Group, is a format created specifically for use with pixel-based images, continuous tone photographs in particular. "Experimentation has show that in general it is possible to compress an image to approximately one-tenth its original size using JPEG" [24] before a noticeable difference in the appearance of the image occurs. JPEG's compression techniques can result in the loss of data and quality if used to great extent, but it considerably reduces the size of a file.

The Need for Image Compression

The need for image compression becomes obvious when an individual considers the number of bits that result from an image being sampled through the scanning process.
Consider the following example: a color image digitized at 24 bits per pixel and displayed at a resolution of 640x480 pixels in a computer requires nearly a megabyte of storage. A typical floppy disk has only 1.4 megabytes of storage space! The size of this file is very large and calls for some method to reduce it so that it is portable and easily stored.

Human perception plays a large role in digital image compression. There are limits to what the human eye can perceive and discern. Haltoning is an example of how the human eye can be tricked by what it is seeing. Halftoning creates the illusion of a variety of tones in an image by using dots of varying size to create images. The eye is tricked into perceiving shades or tones and cannot pick up the dot structure unless the screen frequency of the halftoning drops below a certain number or the image is scrutinized at close range.

Continuous tone images contain too much data for the human eye to perceive. This data tends to be redundant and can be reduced to an extent without the eye noticing any difference in detail. Redundant data can be removed, the large file size can be scaled down, and the image will still appear acceptable to the human eye. "Humans who are end receivers do not need—or cannot use—all the information captured during digitization."[26] Smaller files are more convenient than voluminous ones. Compression is logical, convenient, and necessary.

History of Compression

"The need to efficiently represent information, particularly text, has been with us since man learned to write. . . . Data compression for other forms of information has a much shorter history." [27] It wasn't until the 1950s that technological advances brought about digital systems for images and following that, compression of digital images.

The 1970s saw telephone transmissions go digital, when they became coded for efficiency. This is a form of compression. The 1980s saw the introduction of FAX technology, which employed a form of digital compression. By the mid 1980s "business and consumer applications for high-quality, digitized photographs, and all type of continuous tone images began to appear." [28] Now as we approach the new millennium, a user has a choice of many compressions for reducing the size of a file for transportation over the web or storage.

Methods of Image Compression

There are a number of image compression methods that have been developed over the years to reduce the size of large amounts of data. Compression techniques can be divided into two general categories: lossy and lossless compression.

Lossy. "Lossy compression techniques involve some loss of information, and data that have

been compressed using lossy techniques cannot be recovered or reconstructed exactly." [29] If loss of detail is not a concern, then lossy compression is a good choice, yielding very high compressions. Examples of the use of lossy compression would be images used on web pages or photos that appear in newspapers that have been transmitted over the web or a network where the images do not appear very sharp. The higher the compression, the faster the processing time, the lower the image quality.

Lossless. "Lossless compression techniques, as their name implies, involve no loss of information. If data have been losslessly compressed, the original data can be recovered exactly from the compressed data." [30] Lossless compression does not compress as aggressively as lossy methods, so file size is not reduced by as much. This type of compression is used to encode a file and is best used with text data, computer-generated data, and some image information. Also if the compressed data should reconstruct exactly, lossless compression should be used. [31]

According to Roy Hoffman in *Data Compression in Digital Systems*, compression techniques are used for two types of data. **Symbolic data** is a representation of something a human observer would recognize like letters or numbers. ASCII code is an example of symbolic data. **Diffuse data** is not something that is recognizable to the human observer. Its information has not yet been converted to something familiar. "The meaning and structured properties have not yet been extracted." [32] Image information is diffuse data. When an image is scanned and sampled, it is unrecognizable until a computer translates the sampled information and displays it. Diffuse data, data that is spread out, needs more compression than symbolic data which is more organized from the start.

Symbolic Data. Symbolic data compression algorithms process data and provide three kinds of encoding: variable to fixed, fixed to variable, and variable to variable. Variable to fixed length algorithms "replace sequences of consecutive identical symbols with three elements: a single symbol, a run length count, and an indicator that signifies how the symbol and count are to be interpreted." [33] Run Length Coding is a popular type of variable to fixed encoding. It is most effective when the sequences to encode are abundant. "In run-length coding, a start pixel represented by a specific source symbol is considered the start of the run, and the length of the run is used to replace the run." [34] For example, a line that read $$$$$$$$$$$$$$$$$$$, would be coded as 20$.

Fixed to variable algorithms can be divided into two categories: statistical and dictionary coding. Overall, these types of coding begin with fixed information and output variable code, "assigning shorter codewords to frequently occurring, more-probable symbols, and assigning longer codewords to infrequently occurring, less-probable symbols." [35] In statistical coding, "the number of bits assigned to each symbol is determined by the frequency of the occurrence of each symbol." [36] Two common compression techniques that use statistical coding are Huffman coding and Arithmetic coding. Dictionary algorithms are also very common.

> By constructing a dictionary, a message, which is a sequence of input symbols, can then be encoded as a sequence of reference entries to the dictionary. A simple technique is looking up the English word "terminology" in an English dictionary. If this word appears at the yth word of page x, then the word "terminology" can be represented by a reference entry (x,y). [37]

The dictionary is used as a reference. 'Since most images consist of repeating sequences of pixels, a dictionary based coding method . . . can be applied to an image file."[38] Variable to variable coding is a combination of variable to fixed and fixed to variable. This method is used to make up for inadequate compression of methods that focus on maximizing speed.

Diffuse Data. "Diffuse Data is compressed by applying lossy compression, which throws away non-essential information, in combination with lossless compression for efficient coding." [39] Diffuse data compression methods have been "created using what is described as perceptual coding techniques that exploit the limitations of human . . . eyes." [40] Most often a file with diffuse data is compressed using a perceptual coding algorithm first to get rid of unnecessary data and then a symbolic data algorithm in order to code effectively. Diffuse data can vary so much that the methods used to compress it must be flexible to accommodate it.

Efforts have been made to standardize compression methods. The standard recommendation for grayscale, color photographic quality images is JPEG, the Joint Photographic Experts Group. One of the project's focuses is on image reproducibility for digital print-

ing systems. Therefore, we will take a closer look at the JPEG compression algorithm and its methods of compressing diffuse data.

> Lossless image compression exploits both spatial redundancy, between pixels in a region of a picture, and symbol redundancy, in the output bit stream produced by the coder. Lossy compresion exploits irrelevancy to remove information from the picture that is unimportant to the human eye. [41]

To reduce the number of bits in a digital image JPEG takes advantage of both statistical and spatial redundancy.

JPEG Compression

The JPEG compression algorithm follows a series of steps to carry out its functions.

Image—Pixel Blocking—DCT—Quantization—RLE(or)Huffman coding

"Rather than thinking of JPEG as a fixed tool which performs a certain operation, it should be considered a set of tools which can be used in different ways to achieve desired end results."[42]

JPEG compression first breaks an image into 8x8 blocks (independent of the resolution of the image) and then applies something called DCT, Discrete Cosine Transform. "The purpose of the discrete cosine transform is to take the spatial information in the matrix (the 8x8 blocks) and converts it to frequency information." [43] "Spatial representation of an image is the only method which appears visually correct to the human eye, however, it does not lend itself well to compression because of the random distribution of luminosity values." [44] DCT organizes the image according to its luminosity values. So, instead of pixels, you have values representing that information organized from the largest to smallest in that matrix—the lowest frequency are in the upper left-hand corner, highest in the lower right-hand corner. Discrete Cosine Transform is lossless.

Quantization, however, is lossy. "Quantization is designed to drive small, non-essential frequency components to zero by scaling each number to the nearest multiple of a value found in the quantization table." [45] This explains where data is "lost" in JPEG compression. By trying to match the quantization tables as closely as possible, original information will be thrown out.

The last step of the JPEG compression is the use of one of the symbolic data encoding techniques. According to Daniel Goodenow in his thesis, *A Reference Guide to JPEG Compression*, Run Length Encoding is applied to compress all the zeros in the final matrix; these would be the luminosity values after quantization. Last a statistical algorithm like Huffman coding is applied.

With all the steps complete, the result is a file of much less magnitude than the original. JPEG is able to handle color and grayscale images at just about every resolution. A user may specify the level or amount of JPEG compression is applied.

Adobe Acrobat

A designer prepares a job to be sent to a printer. On a disk the designer has included the application file of the job and a folder with all the images and fonts used in that file. This person has checked to make sure that all the images are linked in the file and that the file is trapped and separated. The result of all this work is a file that *will print,* on a disk with hardly any memory left on it. Isn't there an easier way to prepare and send a job to be printed?

> Adobe Acrobat software, which is based on an imaging model similar to that of PostScript, was developed so that users could view and manage documents on-screen in a device—and application—independent manner. . . . documents can be created in virtually any application, on any platform, and easily converted to PDF, where they retain the full range of high-quality typography, graphics, images, and color. [46]

The original intention of PDF was to reduce the use of paper, create a paperless world. Acrobat PDF's are standardized containers for information that deliver a job in a way that includes all the pieces and ensure that it will print out correctly. According to Alexander Hamilton of *Electronic Publishing*, Acrobat is entering its adulthood where its true usefulness and functionality will be realized. Even though PDF was originally targeted at corporate America to reduce paper, it is now "finding a home with Adobe's core audience in publishing, prepress, and commercial printing."[47]

Acrobat 4.0 is made up of three distinct programs: Distiller which creates the PDF files, Reader which allows PDF's to display or print allowing very minor editing, and Acrobat, formerly known as Exchange, which allows more editability of a PDF file.

Large printing companies such as R.R. Donnelly are beginning to incorporate PDF into their printing workflow. Bob Schaffel, manager for emerging technologies at R.R. Donnelly says that they are *gradually* adopting PDF because it "is not quite ready for full-blown production" use.[48] Adobe has had several version of Acrobat on the market and have just released version 4.0 (April 1999). With version 4.0 Adobe has attempted to listen to the consumer and fix the problems that PDF previously had. There are still problems that are holding some users back, but they will most likely be resolved in the near future.

John Duebert, President of Acquired Knowledge, a company that educates users about the PostScript language, thinks that Acrobat 4 will finally establish PDF as a standard for a number of industries.[49]

Distiller

A PostScript jobs need a RIP to interpret its language. Distiller is in essence a mini-RIP. This program "distills" or filters PostScript code, removing redundancy, and making a display list just like a RIP. "This streamlines files and reduces their size. The resulting code is called "clean code" because it slips cleanly through the RIP and to the imagesetter or output device. When a user creates a PDF with Distiller, they are shown a window with the "distilling time", or the time it took the program to create the PDF. This helps gives the user an idea of the size of their file. Today, Distiller is the primary engine of the PDF workflow and one of four programs in the Adobe Acrobat family for managing PDF files."[50] It is very important to know the options available in Distiller and the parameters of this program in order to create successful PDFs.

Distiller offers a menu called Job Options where a user can configure the settings for a PDF file to their exact specifications. There are five windows within the Job Options: general, compression, fonts, color, and advanced settings. The settings we are concerned with in this project are the compression settings.

Acrobat 4.0 has done something to help a user prepare files that was not previously avail-

able. Before an individual ever delves into setting compression parameters, Distiller assigns a label to the file according to its final purpose and destination. Acrobat 4.0 allows the file to be *screen* optimized, *print* optimized or *press* optimized. As soon as one of these labels is chosen, Distiller assigns the file default settings. Screen optimized files are meant for viewing and displaying on a screen, such as on a website. This type of file has the lowest quality settings to create the smallest PDF possible. Print optimized files are intended for printer, digital copiers, or in some cases, proofs. This settings keeps all fonts, leaves color unchanged and compresses, but not to the same degree as the first. [51] Press optimized is intended for high-end print use. File size is not too much of an issue in this case, therefore compression is low. The settings accompanying these labels can be altered by the user in the job options windows in Distiller, but it is nice to see what settings the program engineers thought were appropriate for specific types of files.

Acrobat 4.0 Compression

"Depending on the settings you choose, compression and resampling can significantly reduce the size of a PDF file with little or no loss of detail and precision." [52] The compression offered by Acrobat 4.0 are divided into three categories: color, grayscale, and monochrome. Earlier, it was determined that there are two types of graphics: bitmap and vector. "Distiller applies ZIP compression to text and line art; Zip or JPEG compression to color and grayscale bitmap images; and ZIP, CCITT Group 3 or 4, or Run Length compression to monochrome images." [53] The following is more specific information about each type of compression offered by Acrobat 4.0.

> • Zip is a compression method that works well on images with large areas of single colors or repeating patterns such as screen shots and simple images created with paint programs, and for black-and-white images that contain repeating patterns. Acrobat provides 4-bit and 8-bit ZIP compression options. If you use 4-bit ZIP with 4-bit images, or 8-bit ZIP with 4-bit images, ZIP is lossless . . . . However, using 4-bit ZIP with 8-bit data can affect the quality, since the data is lost.
>
> • The JPEG compression method is suitable for grayscale or color images, such as continuous tone photographs that contain more detail than can be reproduced on-screen or in print. . . . Because JPEG eliminates data, it can achieve much smaller file sizes than ZIP compression.

> Acrobat provides five JPEG options, ranging from Maximum quality (the least compression and the smallest loss of data) to Minimum quality (the most compression and the greatest loss of data). The loss of detail that results from the Maximum and High quality settings are so slight that most people cannot tell an image has been compressed; at Minimum and Low, however, the image may become blocky and acquire a mosaic look. The medium quality setting usually strikes the best balance in creating a compact file while still maintaining enough information to produce high-quality images.
>
> • The CCITT (International Coordinating Committee for Telephony and Telegraphy) compression method is appropriate for black-and-white images made by paint programs and any images scanned with an image depth of 1 bit. CCITT is a lossless method.
>
> Acrobat provides the CCITT Group 3 and Group 4 compression options. Group 4 is a general-purpose method that produces good compression for most types of monochrome images. CCITT Group 3, used by most fax machines, compresses monochrome bitmaps one row at a time.
>
> • Run Length is a lossless compression option that produces the best results for images that contain large areas of solid white or black. [54]

Acrobat 4.0 offers many choices for compression of a file; however, there is a problem with the compression in 4.0 that is under repair. " Acrobat Distiller 4.0 does not compress color and grayscale graphics correctly. . . . 4.0 does not distinguish between high and medium compression, or between low and minimum compression."[55] If a user knows how each compression works, an educated selection can be made as to the compression of the file. If a user selects "automatic" compression, Distiller decides the best compression for the grayscale and color images in the file, depending on which optimization the user has cosen

> "JPEG is usually applied to 8-bit grayscale images and to 8-bit, 16-bit, and 24-bit color images when the images have continuous, smooth tones; ZIP is applied to 2-bit, 4-bit, and 8-bit grayscale images, to 4-bit color images, and indexed 8-bit color images, and to 16-bit and 24-bit color images when the images have sharp color changes." [56]

Creation of PDF

There are three ways to create a PDF document at this point. First a user may convert an application file to a PDF by using a print driver which appears in the Chooser menu called PDF Writer. Secondly, a file may be saved as a PDF from certain application programs such as Illustrator, an Adobe program, or Microsoft Word. The last and most effective method of creating PDF documents is to use the two step method: create a Post Script file and then use Acrobat Distiller. The user has the most control over the outcome of the quality of the file when using this method. Before considering how you will go about making your PDF file, you should first determine the final purpose of the file—is the job destined for high-quality print, a quick and dirty job, or maybe the web? The job's final destination has an impact on which method should be chosen.

Using PDF Writer or PDF straight from an application file is fine if a job is intended for easy transport over the web or other network. These processes, while the quickest and most convenient, don't afford the user a lot of control over the quality of the document and resolution of images may suffer. For high-resolution graphics or end-use involving a high-resolution printing system, the two step method gives the user very good control over all the elements of the file: color, fonts, and compression.

For a step-by-step explanation of how to create these documents, please see Appendix A.

Endnotes for Chapter Two

[1] Mark Coudray "PDF: The Cure for Dawdling Digital Files." *Screen Printing*, 88.13(December 1998):68.

[2] Ronald Recene. *An Investigation into the Use of the World Wide Web as an Interface for Distributing Electronic Documents to and from a Remote Digital Color Printing Site.* (Rochester: Rochester Institute of Technology. 1996): 5

[3] Robert Shuster. *An Introduction to Digital Color Printing.* (Randolph, MA: Agfa Educational Publishing, 1996.) 24

[4] Frank Romano. *Pocket Guide to Digital Prepress.* (Albany:Delmar Publishers. 1996.): 283

[5] Michael Bruno. *Printing in a Digital World. A Summary of Digital Printing Systems.* (Pittsburgh: Graphic Arts Technical Foundation. 1994.): 14

[6] Robert Shuster. *An Introduction to Digital Color Printing.* (Randolph, MA: Agfa Educational Publishing, 1996.) 3

[7] ibid, 2

[8] Michael Bruno. *Printing in a Digital World. A Summary of Digital Printing Systems.* (Pittsburgh: Graphic Arts Technical Foundation. 1994.): 16

[9] Frank Romano. *Pocket Guide to Digital Prepress.* (Albany:Delmar Publishers. 1996.): 291

[10] Robert Shuster. *An Introduction to Digital Color Printing.* (Randolph, MA: Agfa Educational Publishing, 1996.) 20

[11] Frank Romano. *Pocket Guide to Digital Prepress.* (Albany:Delmar Publishers. 1996.): 291-292

[12] Roy Hoffman. *Data Compression in Digital Systems.* (New York: Chapman & Hall. 1997.): 4

[13] Gilbert Held. *Data and Image Compression. Tools and Techniques. Fourth Edition.* (New York. John Wiley & Sons LTD. 1996.): 311

[14] Weidong Kou. *Digital Image Compression. Algorithms and Standards.* (Boston: Kluwer Academic Publishers. 1995.): 1

[15] Majid Rabbani. *Digital Image Compression Techniques.* (Bellingham, WA: Spie Optical Engineering Press. 1991.): 3-6

[16] Frank. Cost. *Pocket Guide to Digital Printing.* (Albany:Delmar Publishers. 1997.): 3-5

[17] Frank Romano. *Pocket Guide to Digital Prepress.* (Albany:Delmar Publishers. 1996.): 162

[18] ibid, 163

[19] ibid, 163

[20] Frank Cost. *Pocket Guide to Digital Printing.* (Albany:Delmar Publishers. 1997.): 36

[21] Roy Hoffman. *Data Compression in Digital Systems.* (New York: Chapman & Hall. 1997.): 91

[22] Frank Cost. *Pocket Guide to Digital Printing.* (Albany:Delmar Publishers. 1997.): 36

[23] ibid, 37

[24] Roy Hoffman. *Data Compression in Digital Systems.* (New York: Chapman & Hall. 1997.): 6

[25] ibid, 76

[26] ibid, 6

[27] ibid, 7

[28] Khalid Sayood. *Introduction to Data Compression.* (San Francisco: Morgan Kaufmann Publishers, Inc. 1996.): 4

[29] Roy Hoffman. *Data Compression in Digital Systems.* (New York: Chapman & Hall. 1997.): 3

[30] ibid, 4

[31] ibid, 9

[32] ibid, 56

[33] Weidong Kou. *Digital Image Compression. Algorithms and Standards.* (Boston: Kluwer Academic Publishers. 1995.): 14

[34] Roy Hoffman. *Data Compression in Digital Systems.* (New York: Chapman & Hall. 1997.): 58

[35] ibid, 73

[36] Weidong Kou. *Digital Image Compression. Algorithms and Standards.* (Boston: Kluwer Academic Publishers. 1995.): 20

[37] Gilbert Held. *Data and Image Compression. Tools and Techniques. Fourth Edition.* (New York. John Wiley & Sons LTD. 1996.): 315

[38] Roy Hoffman. *Data Compression in Digital Systems.* (New York: Chapman & Hall. 1997.): 75-76

[39] ibid, 76

[40] ibid, 128

[41] Daniel P. Goodenow. *A Reference Guide to JPEG Compression.* (Rochester, NY: Rochester Institute of Technology, 1993.): 1

[42] ibid, 4

[43] ibid, 3-4

[44] Roy Hoffman. *Data Compression in Digital Systems.* (New York: Chapman & Hall. 1997.): 101

[45] Daniel P. Goodenow. *A Reference Guide to JPEG Compression.* (Rochester, NY: Rochester Institute of Technology, 1993.): 10

[46] Adobe Systems Incorporated. "PDF for Prepress Workflow and Document Delivery." *Adobe Printing Solutions.* (1997.): 17

[47] Alexander Hamilton. "PDF for Output." *Electronic Publishing*, 23.3 (March 1999): 26.

[48] ibid, 30

[49] Hevenor, Keith. "John Duebert. President. Acquired Knowledge." *Electronic Publishing*, 23.4 (April 1999): 24

[50] Mark Coudray "PDF: The Cure for Dawdling Digital Files." *Screen Printing*, 88.13(December 1998):69

[51] Adobe Systems Incorporated. "Setting Acrobat Distiller Options." *User Guide. Acrobat 4.0* (1998): 57

[52] Adobe Systems Incorporated. Acrobat Distiller Parameters. (1999.): 16

[53] Adobe Systems Incorporated. "Setting Acrobat Distiller Options." *User Guide. Acrobat 4.0* (1998): 61

[54] ibid, 61-62

[55] Adobe Systems Incorporated. "PDF File Sizes Often Larger from Acrobat Distiller 4.0 Than Acrobat Distiller 3.0x" http://www.adobe.com/suportservice/custsupport/soluti.../1adba.ht. 9/23/99

[56] Adobe Systems Incorporated. "Setting Acrobat Distiller Options." *User Guide. Acrobat 4.0* (1998): 64

Chapter Three

Statement of the Project's Goals

The purpose of this project is to determine the effects that a particular compression setting in Acrobat 4.0 will have on a specific graphic element. As noted previously, there is a flaw in Acrobat 4.0 in terms of compression—there is not a distinct difference between each JPEG compression setting as there should be; instead, the compression is the same for JPEG High and Medium and for JPEG Low and Minimum. For the purposes of this thesis, the compression settings will be tested for color and grayscale images. The compressed files will be output to a digital press, and the results will visually demonstrate the effects of the compressions. Since we can expect that some of the JPEG settings will result in the same compressions, we will not be analyzing every compression setting in Distiller; JPEG Maximum, Medium and Minimum will be used.

This project's goal is to show how compression in PDF files affects the reproducibility of images when printed digitally. The ideal digital workflow would contain PDF and a form of electronic printing. Since Acrobat 4.0 and digital printing systems are becoming increasingly popular, this study is very worthwhile. The tests done in this project could be useful to everyone from the desktop publishing compression novice to the experienced designers and production managers.

This project will demonstrate the effects of compression in Acrobat 4.0 Distiller through digitally printed files. The digital prints will offer a visual analysis of the compression settings for the three types of graphics used. This project also provides information about the various types of compression offered in Acrobat 4.0 to educate users and make Distiller a bit easier to use.

Chapter Four

Methodology

The purpose of this project is to test the compression settings in Acrobat 4.0 for their effects on image reproducibility in digital printing. This invloves testing of a number of PDF files with different types and levels of compression.

The Files

The first step in this project's experiment was to create the original files/images that would be distilled and compressed. File number one contains a very detailed color image, and file number two contains a grayscale image with high contrast. The compression that Acrobat offers for monochrome images are lossless methods. Therefore, a monochrome image will not be affected as noticeably as a grayscale and color image will be. Monochrome compression was omitted from the project due to this fact. The images were scanned using a UMAX PowerLook flatbed scanner and saved as TIFF files without compression. The images were imported into two different Quark XPress files. In order to create the PDFs, the files were converted to PostScript (see Appendix A). The PDFs were created with Distiller from these PostScript files using the compression settings only. Sixteen PDF files were made using the following settings:

1. The Press Optimized feature was chosen, but the compression settings were cleared.
2. Seven PDFs with the following settings for color image compression from File Nmber One:
   - JPEG/Maximum
   - JPEG High
   - JPEG/Medium

- JPEG Low
- JPEG/ Minimum
- ZIP/4-bit
- ZIP/8-bit

3. Seven PDFs for grayscale image compression from File Number Two:

- JPEG/Maximum
- JPEG High
- JPEG/Medium
- JPEG Low
- JPEG/ Minimum
- ZIP/4-bit
- ZIP/8-bit

Printing

All PDF files were printed on an Indigo E-Print digital printing press. The Indigo is a sheet-fed machine employing an electrophotographic process. It is a very popular digital printing system, which is the reason for its incorporation into this thesis project. It is worthwhile to note that the printing was also attempted on a Xiekon DCP 32-D. The print run was unsuccessful as the PostScript files would not rip due to a PostScript error. The same files printed with no problem to the Indigo.

Analysis

The files were analyzed throughout the entire experiment. First an analysis of the numerical data was taken. The original image sizes were noted, and each individual PDF was analyzed for file size and distilling time. The expectation was that as compression increases, file size and distilling time will decrease. Visual analysis was accomplished through an evaluation conducted with outside observers. Twenty people were chosen to evaluate the grayscale and color groups of PDFs. These people were all affiliated with the School of Printing at RIT, and therefore were knowledgeable and practiced in closely evaluating printed material. All analysis took place in a graphic arts viewing booth under standard lighting conditions.

Chapter Five

Results

The results of this experiment were analyzed in two ways—through noting changes in file size and through visual analysis.

File Size Analysis
The test PDFs were created using Acrobat Distiller. Before the visual analysis of the files ever took place, it was apparent that compression changed the files from their original states. Below is a chart showing the distilling time of the files and the numerical differences between the original files and the PDFs.

Color Group
original Photoshop file size—6.5 MB

Compression	**Distilling time**	**File Size**
JPEG Max	3 seconds	1.3 MB
JPEG High	3 seconds	812 K
JPEG Medium	3 seconds	812 K
JPEG Low	3 seconds	144 K
JPEG Minimum	2 seconds	144 K
ZIP 4-bit	3 seconds	392 K
ZIP 8-bit	5 seconds	1.6 MB

Grayscale Group
original Photoshop file size—4.4 MB

Compression	**Distilling time**	**File Size**
JPEG Max	2 seconds	604 K
JPEG High	2 seconds	604 K
JPEG Medium	2 seconds	604 K
JPEG Low	2 seconds	84 K
JPEG Minimum	2 seconds	84 K
ZIP 4-bit	2 seconds	200 K
ZIP 8-bit	3 seconds	840 K

As these results demonstrate, PDF does make a significant difference in the size of a file; the higher the compression, the smaller the file size. It is also evident that the higher the compression is, the shorter the distilling time is. Given these findings, one would most likely expect that of all the PDFs made, the ZIP 8-bit compression would probably match the originals the closest with JPEG Maximum coming in a close second. One could also make the assumption that JPEG Minimum would produce the PDF farthest from the original quality because it produces the smallest file size. JPEG High and Medium and Low and Minimum would produce the same results since they are using the same compression, and one can see that their numerical results match. For this reason, these files were omitted from the visual analysis. One can see that there is a distinct file size difference between JPEG Medium and Minimum PDFs. This would lead an evaluator to believe that there woul be visual differences as well.

Visual Analysis

On first perusal of the printed experiment, the following results were observed by the author before group evaluation.

- JPEG compression produces pretty good results in terms of matching the originals
- Zip 4-bit compression caused the most visual loss of detail/quality in a photo-quality image
- JPEG caused the grayscale image to lighten as the compression increased throughout the experiment
- The opposite seemed to occur in the color image, with the image darkening slightly as compression increased.
- It was much easier to see the effects of compression on the grayscale image than on the color image with increased compression.
- JPEG Maximum and Zip 8-bit produce almost identical results having the closest matches to the original quality.

The PDF prints of the grayscale and color images were evaluated by a group of twenty judges and compared to the original image. The judges were told which print was the original image, but were not informed of the settings of the PDFs; the PDFs were numbered on the back of the print:

Print 1= JPEG Maximum

Print 2= JPEG Medium

Print 3= JPEG Minimum

Print 4= ZIP 4-bit

Print 5= ZIP 8-bit

The judges were asked to analyze the PDFs by comparing them to the original Quark file and then to answer survey questions, which helped to further articulate the effects that

compression had on the images. The judges were asked to point out the print that most closely matched the original(had the highest quality of the PDFs), as well as the print that had the lowest quality and to describe the quality of these prints. They were asked to try to identify any significant differences between prints number 2 and 3, JPEG Maximum and Medium. Finally they were asked to rate the quality of PDF overall in comparison to the original image. The following is the compilation of the data from the survey.

- 70% of the judges surveyed chose number 1, JPEG Maximum quality compression, as the print which most closely matched the original images.
- 20% of the judges chose number 2, JPEG Medium quality compression, as the print which most closely matched the original.
- 10% of the judges chose number 5, ZIP 8-bit compression, as the print that most closely matched the original.

- 100% of the judges surveyed chose number 4, ZIP 4-bit, compression, as the pint that was the farthest from the original quality .

- 90% of the judges surveyed were able to point out a significant difference between the prints that were compressed with JPEG Medium and JPEG Minimum compression.

  The most noticeable difference in the color group was an area in the sky of the picture that lost detail with the increase of compression. The most noticeable difference pointed out for the grayscale group was the loss of contrast and the overall "lightening" of the image with the increase of compression.
- 10% of the judges surveyed were not able to point out a significant difference between JPEG Medium and Minimum.

- 40% of the judges surveyed believe that the overall quality of the PDFs was good or better.
- 60% of the judges surveyed believed that the overall quality of the PDFs was poor to fair.

As you can see, the majority of the judges felt that the PDFs were not adequate for production purposes. The judges felt that the PDFs are pretty reasonable, but they were not good enough for reproduction work. Some said that PDF would be acceptable for typical magazine reproduction work and for quick and dirty jobs. They commented that the end use was important to note when making the decision to use PDF, and that as long as the reproduction was monitored, PDF would probably do a decent job.

The results of this test were most likely influenced by the particular images that were chosen for the test and by the printing method/press that was used to print the experiment. While the results of the test may be used as a general guideline to using Acrobat PDF, an individual may have different outcomes depending on the image or printing method they use. Images can have such a variety of spatial, tonal, and resolution differences form one to another. This will definitely affect how an image appears after it is compressed. Digital printing presses are capable of producing excellent quality, but cannot *completely* measure up to an analog process. This may affect the outcome of such an experiment as well.

Chapter Six

Summaries and Conclusions

This project has shown that when it comes to using PDF an individual must be conscious of two things: the final use or purpose of their file and how the compression will affect the contents of that file. One thing that is certain is that when PDF is used, a file will lose detail. Whether or not the compression is detectable to the human eye is the question. Therefore, the first step in determining your compression settings for PDF is to evaluate your file. What is the end use of the file—production work, proof, quick job? Does it include graphics? If so, is it important to maintain the quality of the images? How far can I compress these images for this particular job?

The purpose of this project was to determine the effects that a particular compresion setting in Acrobat 4.0 will have on a specific graphic element. Two graphic elements were tested in this project—grayscale and color. The test results demonstrated the effects of Acrobat compression methods on these types of graphics. It was possible to define and describe the effects of the compression through a visual analysis accomplished with the help of outside observers. The input from the judges in the experiment was analyzed and compiled to "define" compression settings offered in Acrobat for grayscale and color images.

Acrobat 4.0 Compression Settings
(For 300 dpi. grayscale and color images)

JPEG Maximum quality—
For grayscale images: An acceptable outcome. There will be only minute differences in detail noticeable in this file comapred with the original. A loss of contrast will be noticeable, image will appear lighter than original.

For color images: PDF will most likely produce a very close match to the original. It will be very difficult to find distinct loss of detail.

JPEG Medium quality—

Grayscale images: A decent outcome. A loss of detail will be apparent after close scrutiny, most noticeable in shadow and highlight. Will probably notice further loss of contrast from the original.

Color images: An acceptable outcome. Very slight loss of detail may be apparent, mainly in area of solid color. Image slightly darker than the original.

JPEG Minimum quality—

Grayscale images: Barely acceptable outcome. Highlight area will show a significant loss of detail, almost completely gone. Overall image will be lighter and flatter than the original and the previous settings.

Color images: Decent outcome. Will show a slight loss of data in shadow and highlight areas. Image darker than the original, and slightly darker than JPEG Med.

Zip 4-bit—

Grayscale images: Very bad outcome overall. Solid areas become posterized. Highlights show an almost complete loss of data.

Color images: Also an unacceptable outcome. Posterized in certain areas; details drop out of the highlights.

Zip 8-bit—

Grayscale images: An acceptable outcome. Almost identical to JPEG Maximum; a little less contrast than the original and JPEG Max.—flatter, but otherwise acceptable.

Color images: Acceptable outcome. Produces very similar results to the original and matches JPEG Max. Appears slightly darker then the original; some detail change is noticeable in the highlight.

This experiment could be extended by testing a number of different color and grayscale images of a variety of tonal ranges. The amount of information in an image will affect how it appears when it is compressed. It would also be interesting to see the different

PDFs printed out on a variety of digital printing presses, and not just one. It might also be worthwhile to compare the PDFs that were printed on digital presses to the same PDFs printed on an analog press. The printing method could affect the visual outcome of the prints.

It seems that with advancement in technology, quality of products are compromised and sometimes sacrificed. This is true of the portable document. PDF depends on compression to reduce size and transfer time of files. This results in a loss of data which (depending on the type of compression) cannot be recovered. This translates to a loss of quality. There are many factors to consider when compressing a file. If an individual wonders which settings to use in Acrobat, the answer really is, "It all depends . . .". This is why knowing your file, it's purpose, and how compression works can be so helpful in tackling the often confusing issue of PDF.

Bibliography

"Acrobat 2.1's Compression Options for PDF Files." *Adobe Technical Solutions Database*. http://www.adobe.com/supportservice/custsupport/solutions/bd06.htm. 24 Mar. 1999.

Adobe Systems Incorporated. "creating high-quality PDF files with Adobe Acrobat for print & press." *How to Create Adobe PDF File for Print and Press.* 1999.

Adobe Systems Incorporated. "Converting Electronic Files to PDF." *User Guide. Acrobat 4.0.* (1998.)

Adobe Systems Incorporated. "PDF for Prepress Workflow and Document Delivery." *Adobe Printing Solutions.* 1997.

Adobe Systems Incorporated. "Setting Acrobat Distiller Options." *User Guide. Acrobat 4.0* (1998)

Adobe Systems Incorporated. *Acrobat Distiller Parameters.* (1999.)

Andersen, Mattias, et al. *PDF Printing and Publishing. 2nd Edition. The Next Revolution After Gutenberg. Adobe Acrobat for Digital Prepress Systems.* (Torrance, CA:Micro Publishing Press. 1998.)

Bruno, Michael. Pocket Pal. *A Graphic Arts Production Handbook.* (Memphis: International Paper. 1997.)

Bruno, Michael. *Printing in a Digital World. A Summary of Digital Printing Systems.* (Pittsburgh: Graphic Arts Technical Foundation. 1994.)

Callan, Jamie. "Encoding and Compression." *Text Based Information Systems.* http://ciir.cs.umass.edu/cmpsci646/Notes/Compression/sld001.htm. 24 Mar. 1999.

Cost, Frank. *Pocket Guide to Digital Printing.* (Albany:Delmar Publishers. 1997.)

Coudray, Mark. "PDF: The Cure for Dawdling Digital Files." *Screen Printing*, 88.13 (December 1998):68-72.

Davis, Mills. "PDF in Print Production." *Electronic Publishing*, 23.1 (January 1999): 42-50.

Duebert, John. "Bad PDF: Symptoms and Solutions." *Electronic Publishing*, 22.11 (November 1998): 63-64.

Eliezer, Caren. "Data Compression: Aslett Makes it a Smart Move." *The Seybold Report on Publishing Systems*, 26.14 (April 1997):17-23.

Goodenow, Daniel P. *A Reference Guide to JPEG Compression*. (Rochester, NY: Rochester Institute of Technology, 1993.)

Hamilton, Alexander. "PDF for Output." *Electronic Publishing*, 23.3 (March 1999): 26-36.

Held, Gilbert. *Data and Image Compression. Tools and Techniques*. Fourth Edition. (New York. John Wiley & Sons LTD. 1996.)

Hevenor, Keith. "John Duebert. President. Acquired Knowledge." *Electronic Publishing*, 23.4 (April 1999): 20-24.

Hoffman, Roy. *Data Compression in Digital Systems*. (New York: Chapman & Hall. 1997.)

"Introduction to JPEG." *Compression Frequently Asked Questions.* http:// www.cis.ohio-state.edu/hypertext/faq/usenet/compression-faq/part2/faq-. 24 Mar. 1999.

Jacobson, Wendy. "PDF 'State of the Union'." *Vue/Point. The Hard Copy*. Summer 1998:35-38.

Kou, Weidong. *Digital Image Compression. Algorithms and Standards.* (Boston: Kluwer Academic Publishers. 1995.)

Rabbani, Majid. *Digital Image Compression Techniques*. (Bellingham, WA: Spie Optical Engineering Press. 1991.)

Recene, Ronald. *An Investigation into the Use of the World Wide Web as an Interface for Distributing Electronic Documents to and from a Remote Digital Color Printing Site*. (Rochester: Rochester Institute of Technology. 1996)

Romano, Frank. *Pocket Guide to Digital Prepress*. Albany:Delmar Publishers. 1996.

Sayood, Khalid. *Introduction to Data Compression.* (San Francisco: Morgan Kaufmann Publishers, Inc. 1996.)

Schlapfer, Kurt., and Erwin Widmer. "Comparison of File Formats for the Exchange of Full Page Electronic Data." *TagaProceedings*, (May 1997): 287-299.

Shuster, Robert. *An Introduction to Digital Color Printing.* Randolph, MA: Agfa Educational Publishing. 1996.

"Standards Appendix 6." *Digital Advertisement Transfer Industry Committee*. http://www.datic.com.uk/appendice6.htm. 24 Mar. 1999.

Stoy, John. "Color Workflow with PDF" *Electronic Publishing*, 23.4 (April 1999): 26-34.

Appendices

Appendix A

Creation of PDF

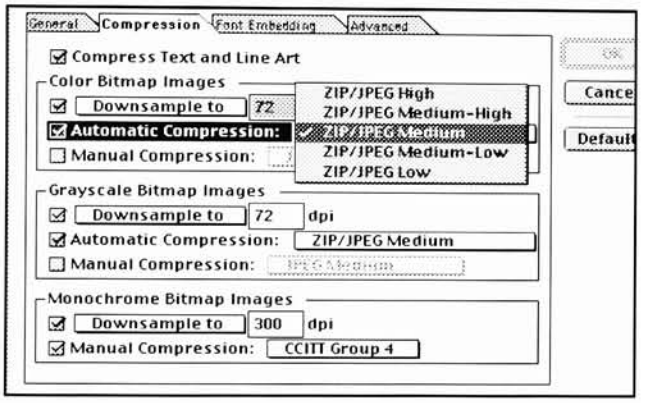
The Two Step Method

(Performed on a Macintosh G3)

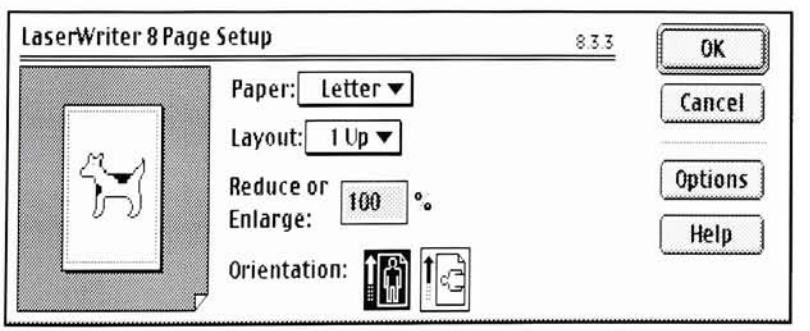
Step One—PostScript

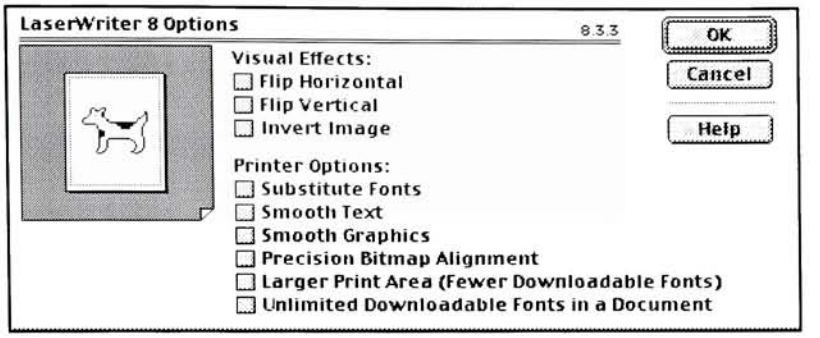
A user should be able to create a PostScript document from an application file in order to create PDF's using this method. It is a simple process to create PostScript files.

1. Go to Chooser and select a PostScript printer.

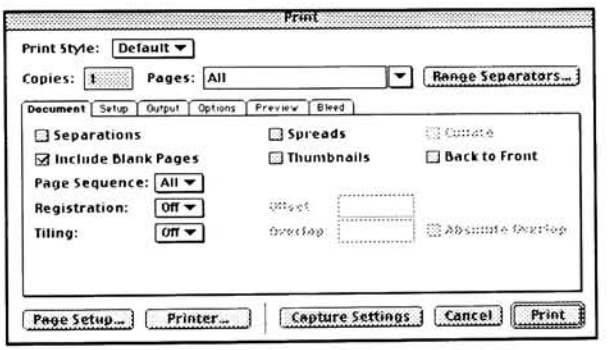


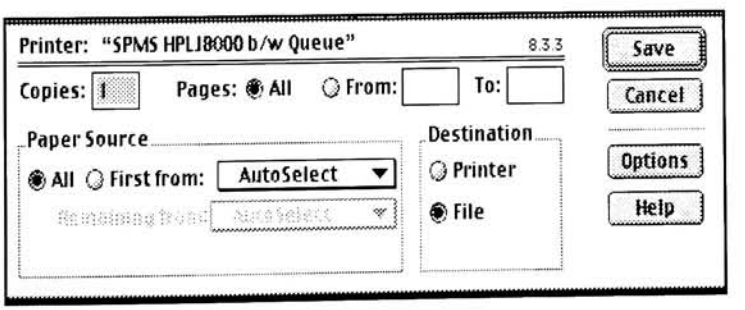
2. Go to Page setup; selections will differ according to output device. Click on options, go to PostScript options, and uncheck boxes.

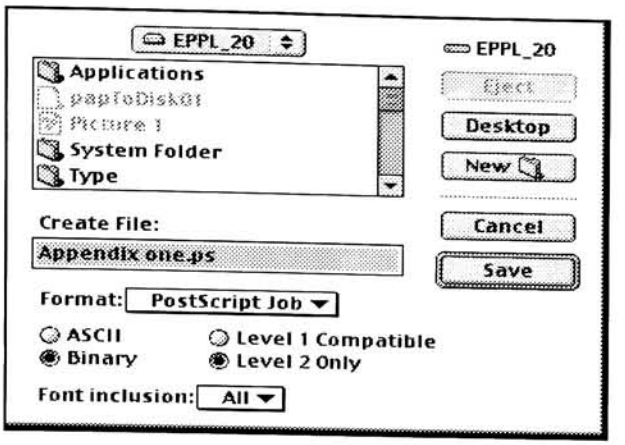




3. Go to Print dialogue box. Make sure Destination is File. Save as File. Choose PostScript level 2, Data Format—binary, font inclusion—all. Click on Save.





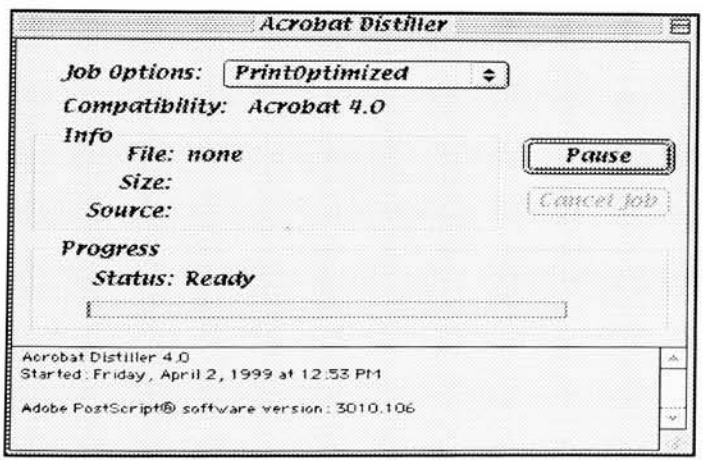


Now that a PostScript file has been made the user can open Acrobat and create a PDF. The way that Acrobat is able to make PDF files is through a piece of the program called Distiller.

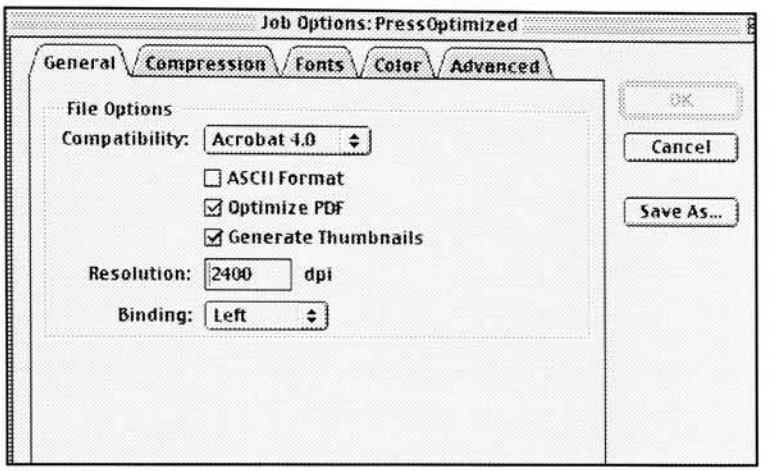
Step Two—Distiller

1. In Acrobat 4.0 folder, choose Distiller and open.

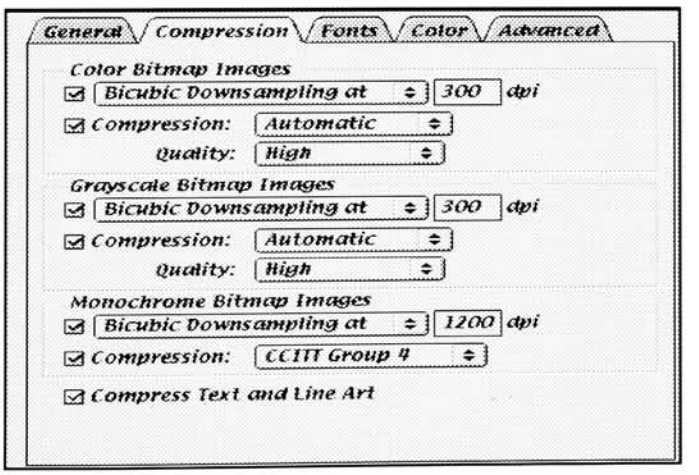
2. Distiller window will pop up. First choose how the file is to be optimized.



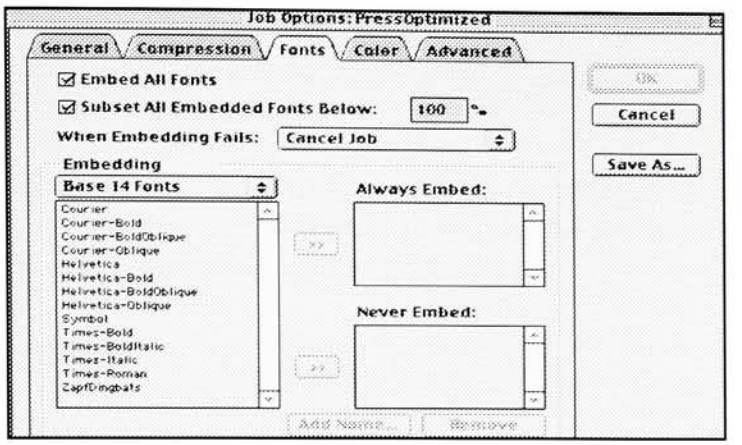
3.Under the File menu, choose job options. Here you may configure the settings for your PDF. First set general options:



4. Second set compression parameters. Acrobat will automatically downsample which will lower resolution of images. Uncheck if this is not a desired result.



5. Set Fonts parameters.



6. You may set the color and Advanced Settings paprameters for your file as well; the setings will depend on the type of file and purpose of the file.

7. Once settings are configured, go to File menu and open the PostScript file that is to be converted to a PDF. You will have the option to save the file as a PDF; choose save.

Following the two step method will result in a PDF that is customized to your particular needs.

Appendix B

PDF Image Evaluation

The images you are evaluating consist of an original QuarkXPress file and a series of PDF files utilizing a variety of image compressions. The files were printed on an Indigo E-Print digital press.

Please view the original QuarkXPress file closely and then evaluate the PDF prints based on that original. Please answer the following questions for each group of images focusing only on the quality of the image.

Group name

1. Which PDF seems to match the quality of the original file the closest? (please write in the number of the print)

Image #

2. Which PDF seems to have the lowest quality? (please write in the number of the print)

Image #

3. Do you see any significant differences between print numbers 2 and 3?

Yes No

Comments:

4. How would you rate the overall quality of the PDFs in comparison with the original? (please circle choice)

1—low quality
2—fair quality
3—good quality
4—very good quality
5—excellent quality

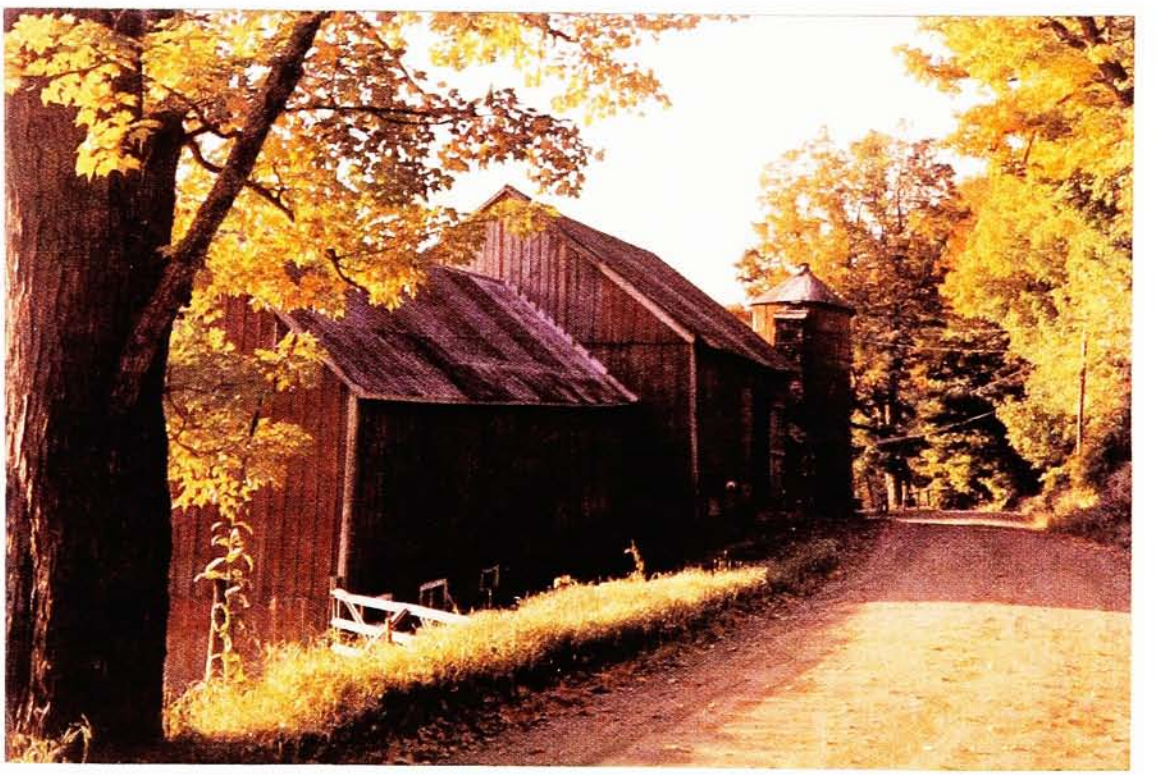
5. How would you describe the PDF with the highest quality?

6. How would you describe the PDF with the lowest quality?

Appendix C
Test Prints

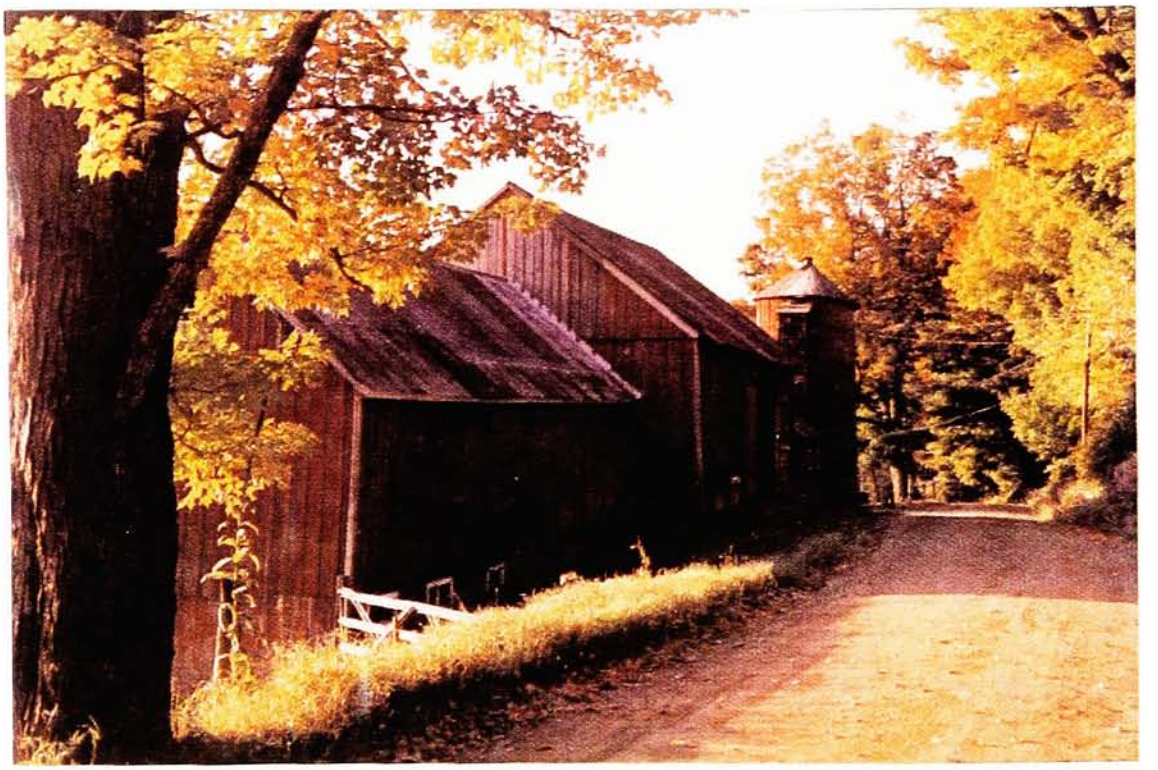
Color Image Test File

Barn.tiff



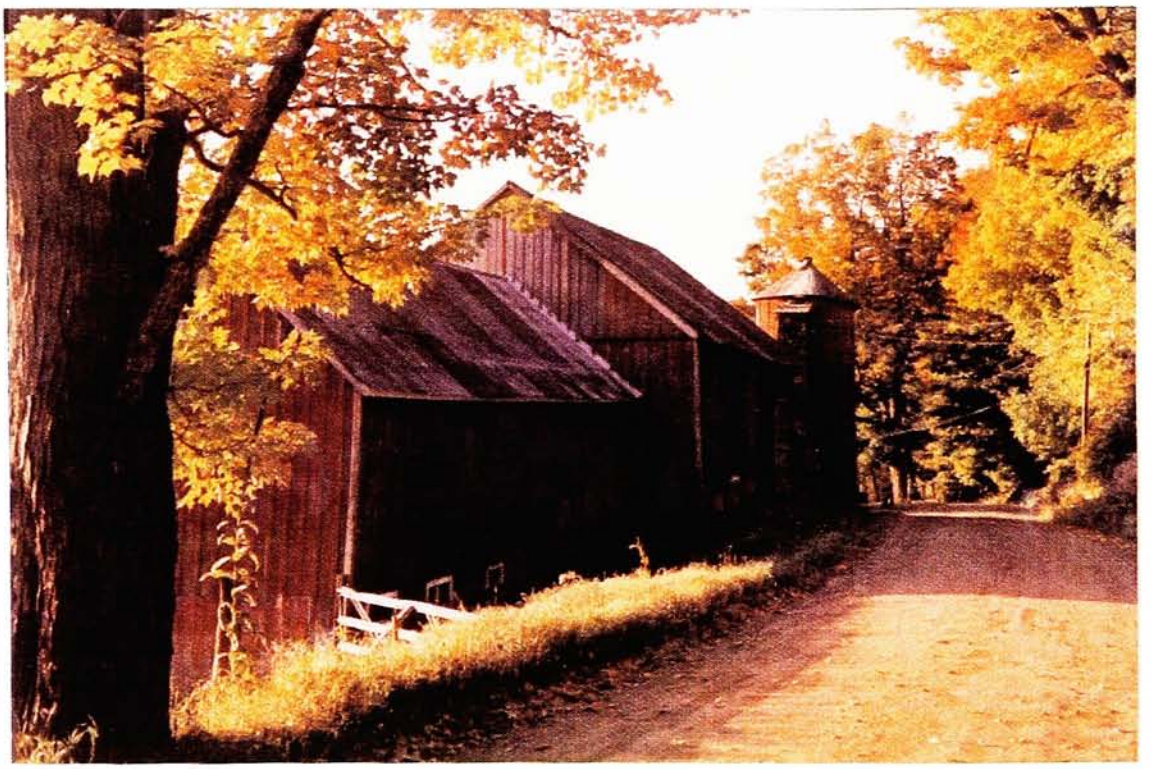
Color Image Test File

Barn.tiff
JPEG Maximum (quality) Compression



Color Image Test File

Barn.tiff
JPEG Medium (quality) Compression

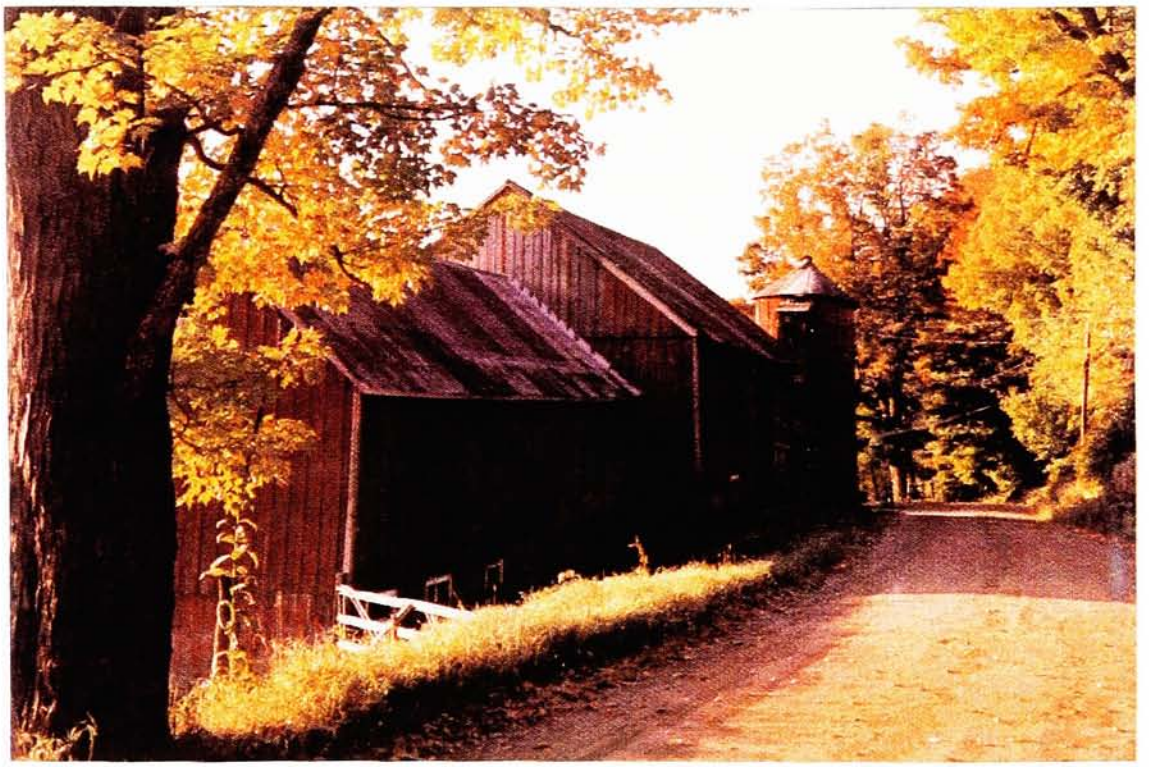


Color Image Test File

Barn.tiff
JPEG Minimum (quality) Compression

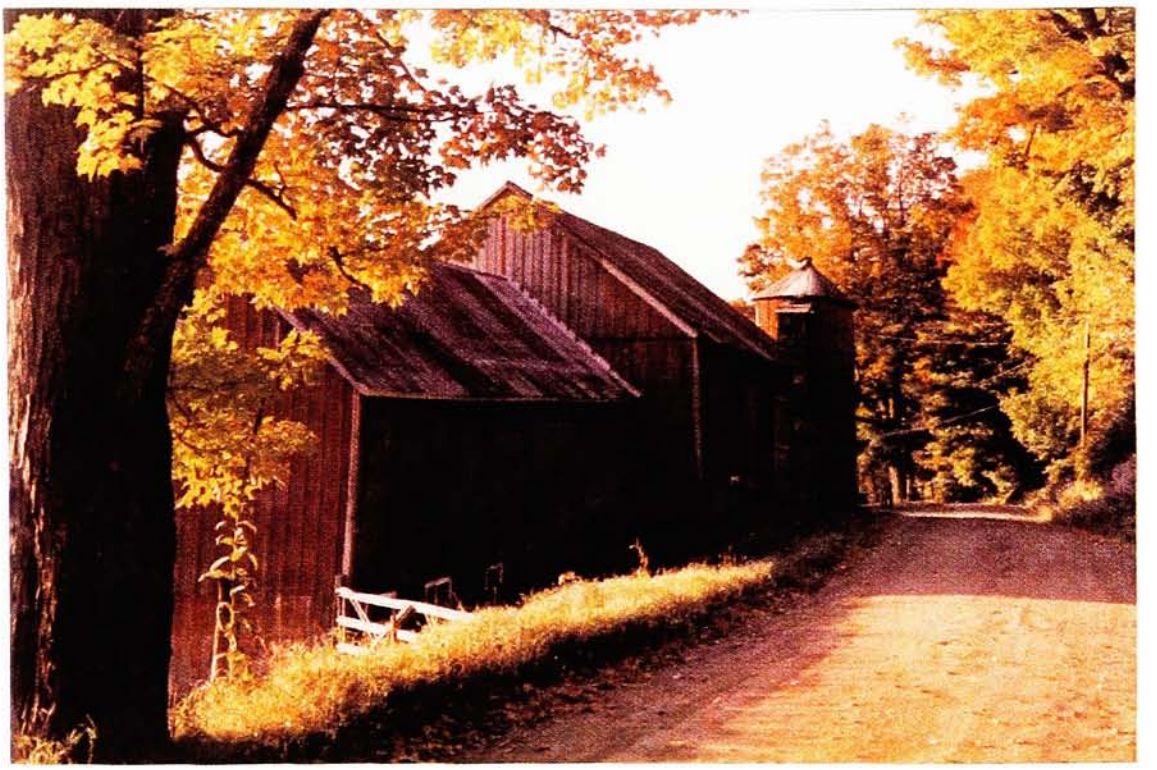
Color Image Test File

Barn.tiff
Zip 4-bit Compression



Color Image Test File

Barn.tiff
Zip 8-bit Compression



Grayscale Image Test File

Caragh.tiff
Original

Grayscale Image Test File

Caragh.tiff
JPEG Maximum (quality) Compression

Grayscale Image Test File

Caragh.tiff
JPEG Medium (quality) Compression

Grayscale Image Test File

Caragh.tiff
JPEG Minimum (quality) Compression

Grayscale Image Test File

Caragh.tiff
ZIP 4-bit Compression

Grayscale Image Test File

Caragh.tiff
ZIP 8-bit Compression