



White Paper: Data Capture and Document
Management Systems - 10 Tips and
Information Nuggets That Will Save
You Time, Money, and Hair



Choosing the Right Image Resolution – Part 1

by Phillip Cormier

When you think of the word ‘resolution’ what do you think of? Maybe you think of New Year’s resolutions, or maybe you think of resolution in the context of business law or mathematics. Or maybe you think about the British royal navy ships built in the 1600’s and 1700’s that bore the name of HMS Resolution. However, for the capture and document management world, most of us usually think of resolution in conjunction with image measurement which describes the detail an image holds. Simply stated, a higher resolution means more image detail. But determining how much detail is enough and understanding that there are trade-offs between detail, costs, and system performance are important points to consider when you are dealing with data capture and integrated document management systems.

Typically measured in DPI (Dots per Inch), resolution is a measure of how many pixels are contained in a given inch of linear space. A pixel is a single point in an image composed of varying values of red, green and blue. To put it simply, a pixel is one point of color. For example, a 300x300 DPI image will contain 300 pixels each horizontally and vertically for a total of 90,000 pixels per square inch of image. The higher the pixel density, or DPI, of an image, the more detail that image contains and generally speaking the better it will look. As a side note, it can be easy to confuse DPI with image size, such as the numbers often expressed for computer displays. A 1024x768 computer display is not an expression of its resolution, but rather the total number of pixels in the display.

But enough of the technical image resolution jargon. What does all of this mean to *you* and *your* scanning operation? Well, it means that determining the proper image resolution is not straight forward and that there are several things to consider: storage sizing, accuracy of processes such as Optical Character Recognition (OCR), document transmittal time, scanning time, and end user viewing quality. With each of these factors affecting the other it can be difficult to strike a balance.

Storage Sizing

Let’s start with storage sizing. Storage sizing can be a complex process, often involving some guess work. It is not always possible to accurately predict how many pages your system will acquire over the years. A good plan is do your best to accurately predict your volume, build in a buffer, and most importantly ensure that the hardware and software you choose is able to grow as your business grows. Image resolution plays an important role in this decision. The actual size of a file is affected by several factors including the resolution of the image, the actual amount of data the image contains, and the storage format. The following is a fairly standard example that you may run across: A 1 page letter sized text file, monochrome, with a medium amount of data, is saved as a TIFF Group IV image in 200x200, 300x300, and 600x600 DPI resolutions. The 200DPI image is 23KB, the 300DPI image is 33KB, and the 600DPI image is 99KB. You can see how the file size grows quickly as the resolution increases. Now, let’s say that the content management system will hold a maximum of 1 million images. The 200DPI image will require ~22GB of space, the 300DPI image will require ~31.5GB of space, and the 600DPI image will require ~97.5GB of space. Now, with disk space being increasingly affordable these

days, even 500GB isn't a really big deal. But you can see through this example how storage sizing variables are important things to consider when deciding what hardware to buy for primary and backup storage.

Scanning Time

Scanning time must also be considered when determining what resolution to acquire and store your images in. Scanning a 200DPI image on most modern scanners is a very fast process. However, scanning 600DPI, for instance, can sometimes make you pull your hair out because it can be very slow. The first factor to consider in scanning at higher resolutions is the actual scanner hardware. Incremental higher spend normally leads to incremental higher speeds. Some of the very high end scanners suck pages so fast that they can be scanned faster than you can see them being scanned, even at 600DPI! While, some lower end scanners can take minutes per page. Choosing the right scanner is normally just a matter of evaluating your budget and weighing the highest number of pages per minute you realistically will need to scan in a given hour. However, you will also want to evaluate the scanning time that is lost to software processes such as Optical Character Recognition (OCR). You will want to keep in mind that as resolution increases, OCR accuracy increases, but so does processing time.

Image Processing

Image processing, especially OCR and Zonal OCR, is another very important consideration when it comes to choosing the right resolution for your images. The DPI of the images to which these processes are being applied can have a huge affect on the accuracy of the results. In this application, higher is better – to a point. The sacrifice for processing higher resolution images is speed. All things equal, it can take a lot more time to analyze a 600DPI image versus a 300DPI image. There are no hard rules here, but years of experience dealing with the OCR products and API's from many companies shows that no less than 300DPI is acceptable for documents that will be exposed to these types of processes. There are of course many other factors that contribute to OCR/Zonal OCR accuracy but when it comes to resolution, understanding what DPI is being applied is right up there at the top of the pile.

In a nutshell, it is important to consider how long the scanning process will take at various resolutions. Let's say that you have 500,000 pages of backlog to scan, and you have hired a temp agency to do the work onsite for you. Let's also say that scanning and processing a page at 300DPI takes 10 seconds, and scanning the same page at 600DPI takes 18 seconds. The 300DPI backlog scanning will take 58 days, while the 600DPI will take 104 days. That is a big difference in time and cost! And that is just in pure processing time which doesn't include organizing, batching, break time, etc. Although these numbers are purely for example purposes since your actual hardware, processes, and number of scanners will dramatically affect the outcome, the point is to make sure that you think through the time and cost implications based on resolution and character recognition factors when determining your scanning processing needs. The goal always is to choose the lowest resolution that will effectively and efficiently work in the different steps of the document capture, routing, OCR and indexing process. We normally work with 300 DPI.

About UFC Inc.

UFC Inc. is a consulting, integration and solutions firm preferred by clients worldwide for our quality, innovation and integration expertise. UFC provides data capture, enterprise content management software, support and integration services - based on a flexible architecture and common set of applications for collecting, classifying, retaining, migrating, securing and accessing information – all at the lowest cost of ownership.

Unlike vendors that deliver generalized ECM products with centralized or consolidated architectures, or support few applications and data types, UFC delivers the most comprehensive solution, specifically tailored for the customer. The distributed nature of the solution along with UFC's extensive expertise and unique approach makes it ideal for companies with remote offices that have limited storage space, minimal IT infrastructure or technical support. Remote locations realize significant improvement in operational efficiencies, improved collaboration, a reduction in storage costs - without sacrificing centralized control or visibility of information. From capturing and processing vendor invoices, personnel information such as employment applications and human resource forms to capturing and storing engineering drawings, large contracts and correspondence, UFC provides their customers the ability to reduce paper transaction costs while increasing their data processing efficiencies.

Call us today to find out how we can help your organization at (248) 447-0100 or email us at sales@ufcinc.com.

Keywords:

document capture
document capture software
web based scan
web based scanning
scan capture software
document capture solutions
capture document
process capture software
capture documents
web based data capture

Quillix and QSX are registered trademarks of Prevalent Software, Inc. Microsoft, Windows, Windows 95/98/Me/NT/Vista/2000, SharePoint, .NET, and the Windows logo are trademarks of Microsoft Corporation in the United States and other countries. MuWave, UFC, UFC, Inc. and User Friendly Consulting and associated family of applications are registered trademarks of User Friendly Consulting, Inc. All other trademarks are used herein are the property of their respective owners.
©Copyright 2011 UFC Inc. All rights reserved.

