600 dpi for Document Imaging and Digital Copying

600 dpi for

Document Imaging

and Digital Copying

600 dpi (dots per inch, commonly used to refer to pixels per inch) is the standard for digital copying because digital copy quality must compete with analog copy quality. Copy quality must be 600 dpi because top management wants the copied documents, that represent top management in meetings and through the mail, to be crisp and sharp.

Conversely, top management considers document imaging to be a clerical task and a cost center. To reduce costs, the lowest resolution possible (200 and even 100 dpi) has been used in document imaging, even if the lower resolution made the documents harder to read.

Plummeting storage and transmission prices

Recent huge increases in transmission speed and recent huge increases in disk capacity have made resolution irrelevant as a cost factor for all but the largest document imaging efforts. If your project is less than 100 file cabinets (1 million pages) then the images (50 KiloBytes each) will fit on the single 50 GigaByte disk found in most desktop PCs (Personal Computers) and in some laptop PCs.

The 50 KiloByte (400 Kilobit = 400 thousand bytes = 8 bits per byte x 50 thousand bytes) compressed images will traverse a Gigabit ethernet network in one 1 thousandth of a second (1 thousand images can traverse the network in 1 second). On a Gigabit ethernet network, the images come up fast, and each one of 1 thousand people can view one image per second over the network. If your project serves less than 1 thousand users, you do not have to plan a complex design if you have Gigabit ethernet; COTS (Commercial Off-the-Shelf) equipment will easily scale to meet your needs.

Gigabit ethernet (1 billion bit ethernet = 1,000 Megabit ethernet) is now just slightly higher in cost than the common 100 BaseT (100 Megabit ethernet) and soon the two will be equal in cost. In the last generation of networking, 100 BaseT eventually became equal in cost to 10 BaseT (10 Megabit ethernet).

Digital copiers are merging with, and replacing, the scanners used in document imaging for many applications. The only feature missing in digital copiers, that exists in document imaging scanners, is the ability to handle physically damaged paper and to scan both sides simultaneously. Scanning both sides simultaneously is very important for physically damaged paper because the flipping mechanism in digital copiers depends on the use high quality sheets of paper for reliable operation.

Today, it is relatively easy to design scanners to scan at 600 dpi, so it is likely that document imaging scanners will soon offer increased resolution (600 dpi). Some of the document imaging scanners currently scan at 600 dpi and process the image down mathematically to 300 or 400 dpi for storage. Unfortunately, these processed images are not as sharp and crisp as images managed at 600 dpi, end-to-end, throughout the imaging process.

Technology marches on

(But not to our drummer)

The switch to 600 dpi has come about rather quickly. It is a result of advances in technology. Advances in technology are not tied to document imaging, or to the workplace in general. Technology rushes forward at its own pace, affecting the workplace in a somewhat random fashion.

Soon color will also be standard. No one though color monitors would become standard. No one thought flat panel monitors would become standard. In the near future, documents will merely be scanned. The scanner will decide if the images are black and white, grayscale, or color. The scanner will scan in color if the images are color. The scanner will scan in grayscale if the documents are handwritten or have shades of gray. The scanner will scan in high-contrast black and white if the documents are high-contrast black and white documents.

Because of the low cost, a losslessly compressed high contrast black and white image may be created along with all color and grayscale images. This high contrast image may be kept as a security copy to backstop possible problems with the more complex grayscale and color image formats.

As color copiers replace black and white copiers, color document imaging will appear. In general, digitizing in any profession leads to the merger of previously separate areas of expertise. No one expected typists to be able to select fonts and font sizes. It was unthinkable for a typist to be able to left and right justify an entire document with a single mouse click. (Mouse click? No one would have conceived of even a single mouse click, much less an entire world directed by mouse clicks.)

The switch, to digital, ends when all aspects of a process have been digitized. This can be seen as CDs and DVDs lead to digital movie projection in theaters. Star Wars II was filmed (recorded) using digital HDTV (High Definition TeleVision) and stayed digital through to its projection on some theater screens in May 2002.
[http://www.sonyusacinealta.com/]

Star Wars II will stay in the digital domain when it is delivered to consumers who view the movie using the new 100 Gigabyte Blu-Ray DVDs (blue light specials). The Blue-Ray DVD ROMs, announced on February 19, 2002, for HDTV videos, store 100 GigaBytes on two sides, two layers per side. The Blu-Ray DVDs can store: 2+ hours of HDTV video on one layer on one side, which holds 27 Gigabytes).
[http://www.matsushita.co.jp/corp/news/official.data/data.dir/en020219-4/en020219-4.html]

Summary

Document management is following the same path of integration and merger that digitizing has followed in many other areas.

Document imaging is now merging with copying, which is in turn merging with laser printing and document production. Soon there will be just document management.

Appendix

Discussing 600 dpi and 300 dpi, etc.

It is very much easier to manage everything at 600 dpi than it is to discuss the differences between several resolutions. A single resolution also eliminates the need for the controls necessary to make sure that the right documents are scanned at the right resolution. Standardizing on one resolution is like standardizing on a record storage carton size. It is much easier to do than to discuss the alternatives.

OCR

OCR (Optical Character Recognition) operates only on the raster image (pixels) of scanned documents. OCR is not attached to specific physical scanners. Document images can be OCRed at any time after scanning, including many years after scanning.

The OCR process is very sensitive to the size of pixels used (coarseness of scanning). The pixels used at 300 dots per inch are four times as large as the dots used at 600 dots per inch. With the larger pixels, the black pixels created when scanning adjacent characters (in a positive image) are more likely to touch than if smaller pixels were used. This makes the characters appear to run (bleed) together. Character that bleed (overlap) are much more difficult to read (for example, the text 'for example' condensed by 2 points per letter as follows: forexample). For the same reason, OCR results for lower resolution scans are often unacceptable.

Once a document is scanned at 100 dpi, the resolution cannot be increased. If the resolution is too low for effective OCR, then the only way to OCR the image, for full text retrieval or for republishing, is to rescan the image at a higher resolution. If the document was destroyed after scanning at 100 dpi, rescanning is impossible and OCR is impossible. Because 100 dpi scanning also makes the documents harder to read, rekeying the document will also have more errors.

Readability vs. Legibility

Legible means that a document could be read with more effort. Readability describes the extent to which the quality of a document contributes to productivity. Documents that are hard to read take longer to read. Alternatively, employees with deadlines read poor quality documents as quickly as the employees read higher quality documents, but the employees make more mistakes. The employees misunderstand the content of the documents more often. Data is entered incorrectly, and maintenance on equipment is performed incorrectly (if it is maintenance manuals that are being read).

Were the stoplights on the way to work red or green this morning?

No one remembers stoplights that are red (or green) on the way to work. You just stop (or go). It is the only way to drive to work. You do not notice the delays because it is the only way to get to work. Similarly, you do not notice how hard it is to read a poor quality document. You just do it because the document has to be read. You may make more mistakes, but you read the document.

The time required to complete a task, and the error rate for a task are taken as givens. These numbers are then used as the basis for ROI (Return on Investment) when the benefits of 600 dpi scanning are introduced. Because 600 dpi scanning is so inexpensive, the ROI can be done after the fact, rather than by designing expensive tests for ROI.

Makes as much sense as using a color monitor to view black and white documents . . .

Color highlighting on black and white documents is very useful as are color icons. And the same displays are used for all work, not just document imaging. Very few people do only document imaging anymore, just as very few people do only word processing in their job.

Previously, there were arguments over whether different fonts were necessary, over whether bold fonts were necessary, over whether different size characters were needed, and even over whether lower case was needed. When a feature has no apparent additional cost, and it has some use, it has been adopted.

Identifying digital copiers in you organization

Digital copiers scan the image of a document, process the image digitally, and then print the image on a laser printer. At one time a few color copiers were analog, but now all color copiers are digital. The easiest ways to identify a digital copier are: 1.) All of the pages of a document go through the suspected digital copier and then two or more copies come out. 2.) The suspected digital copier can be used as a laser printer from you computer. 3.) The suspected digital copier can send scanned documents to your computer.

DVD Burners

DVD Burners (DVD-R, Digital Versatile Disc – Recordable) are now under USD (United States of America Dollar) $500.00. If your project includes less than 100 thousand pages (10 file cabinets) then you can back up the entire project on one DVD. At less than USD $7.00 per blank DVD, it is no problem to make seven copies to assure disaster protection through spatial and managerial diversity of document storage. Pioneer Electronics:
[http://www.pioneerelectronics.com/Pioneer/CDA/Industrial/IndustrialProductDetails/0,1444,21838,00.html]
[http://www.pioneerelectronics.com/Pioneer/CDA/ServiceAndSupport/SelfHelpforIndWhitePapers/0,1470,60~6020~60201600,00.html]
[http://www.pioneerelectronics.com/Pioneer/Files/DVDR47_WhitePaper.pdf]
[http://www.pioneerelectronics.com/Pioneer/Files/DVDR_whitepaper.pdf] Apple was first with the affordable COTS DVD burners.
[http://www.apple.com/pr/library/2001/feb/19super.html]

CDs (Compact Discs) are fine for 10 thousand page (1 file cabinet) projects. But, like putting records in boxes, or batching documents for scanning, putting document images on DVDs (each providing storage for 80 thousand pages or 8 file cabinets) reduces the number of objects that must be managed. And, a big part of records management is cutting a job down to a manageable size.

Testing the speed of your network

Your network may not match its rated speed. If your network is not as fast as it is rated to be, your document management system may not work as fast as intended. To test the speed of your network, select a 12.5 MegaByte (100 Megabit) (125,000 KiloByte) file on your PC. Drag the file into a folder (directory) on a server or PC that is across the network from your PC. If the file appears on the server in 1 second, you have a 100 BaseT (Telephone twisted-pair wiring) (100 Megabit ethernet) network. If the file appears on the server in 10 seconds, you have a 10 BaseT (10 Megabit ethernet) network.

To test for Gigabit networking: select a 125 MegaByte (1,000 Megabits = 1 Gigabit) file on your PC. Drag the file into a folder (directory) on a server or PC that is across the network from your PC. If the file appears on the server in 1 second, you have a Gigabit ethernet network. If the file appears on the server in 10 second, you have a 100 BaseT (Telephone twisted-pair wiring) (100 Megabit ethernet) network. If the file appears on the server in 100 seconds, you have a 10 BaseT (10 Megabit ethernet) network.

If you do not have a big file, pick a word processing file (files with image are larger), select a large part of the document and keep copying and inserting your selection until the file is the size you are looking for.

If your test does not match the rated speed of the network, explain your method of testing to the network manager. The network manager may have an alternative test. The goal is to achieve the rated speed of the network, either by eliminating bottlenecks, or by lowering the nominal rated speed to the actual network speed. By using accurate data for planning, document management system plans can be made more effective.

1 Gigabit ethernet

In May 2002, Dell began offering Gigabit ethernet on the PC motherboard (from Intel).
[http://www.dell.com/us/en/gen/corporate/press/pressoffice_us_2002-05-20-aus-000.htm] Apple preceded Dell in standardizing on Gigabit ethernet for PCs by about a year. March 8, 2001: [http://www.apple.com/pr/library/2001/mar/08pmg4.html]

10 Gigabit ethernet

Cisco, and Nortel Networks are working on 10 Gigabit ethernet. Even though few people need 10 Gigabit ethernet today, work on 10 Gigabit ethernet [http://www.10gea.org/] provides some assurance that Gigabit ethernet is not at the cutting (bleeding) edge of technology. Cisco has 10 Gigabit ethernet installations at the Curtin University of Technology in Australia, at Kyoto University in Japan, and at Arkansas State University in the United States.
[http://newsroom.cisco.com/dlls/prod_031902.html] Nortel Networks has a 10 Gigabit ethernet installation at Northwestern University [http://www.nortelnetworks.com/corporate/news/newsreleases/2001d/12_05_01_omninet.html]

10 Gigabit ethernet has an application for voice over IP (Internet Protocol) (IP telephony). Each 10 Gigabit network connection can support 100 thousand telephone extensions. IP telephony is within the COTS PC scalability envelope for organizations with less than 100 thousand telephones using a single 10 Gigabit NIC (Network Interface Card).

Dots per (Linear) Inch vs. Dots per Square Inch

600 dpi images have 4 times as many pixels (commonly referred to as dots) as 300 dpi images for the following reason: If the resolution is 1 pixel per inch, there is 1 pixel per square inch. If the resolution is two pixels per inch, there are four pixels per inch, just like a window with 4 panes of glass has two panes of glass along each edge.

3 dpi images have 9 pixels per square inch. 4 dpi images have 16 pixels per square inch. 6 dpi images have 36 pixels per square inch (4 times the 9 pixels per square inch that 3 dpi images

have). Mathematically, resolution is given as a linear measure (dots per linear inch) but documents have a two dimensional area given in square inches so the volume of data for a scanned image is proportional to the document's area in square inches.

600 dpi images have an areal density of 360 thousand pixels per square inch.

400 dpi: 160 thousand pixels per square inch. 300 dpi: 90 thousand pixels per square inch. 200 dpi: 40 thousand pixels per square inch. 100 dpi: 10 thousand pixels per square inch.

600 dpi images are only slightly larger than 300 dpi images

600 dpi images are not really 4 times the size of 300 dpi images or 9 times the size of 200 dpi images.

All document images are stored as compressed images. Compression works by removing the redundancy from an image. This is similar to removing duplicates in a file. After the duplicates have been removed, all of the information is still in the file, but the file has less volume. When one scans a document image at 600 dpi, the image is the same as the image scanned at 300 dpi, there are just more pixels. The white areas are still white and the black areas are still black, there are just more pixels to represent the black and white areas. Only along the edge of black or white areas is there new information (more detail). This small increase of information is why compressed 600 dpi images are not exactly the same size as 300 dpi images.

Speckles (sometimes called salt and pepper) are random dots on the image caused by dust or dirt. 600 dpi scanning is more likely to pick up the dirt (or digital noise) and so de-speckling becomes more important at 600 dpi.

A compressed 600 dpi scan of an image is rarely more than twice the size of a compressed 300 dpi scan of the same image. Because the cost of disk storage is currently dropping at 50 percent per year, if storing 300 dpi images was feasible last year, storing 600 dpi images is feasible this year (1 year later and 50 percent less expensive per unit of storage).

1200 dpi

Yes, laser printers are now moving to 1200 dpi. Scanner will eventually follow.

1200 dpi scanning is necessary to reproduce all of the sharpness of documents printed so crisply on 600 dpi laser printers. (This requirement for doubled-resolution scanning is based on the Nyquist sampling theorem: Harry Nyquist, "Certain Topics in Telegraph Transmission Theory," Trans., AIEEE, Vol. 47, April 1928, pp. 617-644.) (The integration and merging effect of digitizing means that a paper published in 1928, on telegraphy, was relevant to document management in 2002.)

However, most analog copiers reduce the crispness of a 600 dpi image to 300 dpi. Many original documents were originally printed on older laser printers at 300 dpi. For copies and for older original documents, 600 dpi scanning is all that is required to record all the detail available in existing documents. For current documents, 600 dpi scanning is much better than 200, 300, and even 400 dpi scanning. By the time that the added costs of 1200 dpi scanning becomes unimportant (as the added cost of 600 dpi scanning has become unimportant for many applications) it is possible that the documents will be captured electronically, eliminating the need for scanning.

Note to Readers

Updates and More Detailed Descriptions:

When using the information in this article, please check [http://www.ArchiveBuilders.com] for updates. The version number of this article is just before the page number below. The website also has articles that provide more details on some of the terms and concepts in this article.

Comments

Please let us know how you like this paper, or if you had any questions. What would you like to see in the future? For more, and the most recent version of this article, please visit our web site at www.ArchiveBuilders.com.

Please send your comments via email to SteveGilheany@ArchiveBuilders.com. Tel: +1 310-937-7000. Fax: +1 310-937-7001. Also, please let us know where you saw this article.

Acknowledgements

Reprinted from Archive Planning, Volume 6, number 6, 2002, Archive Builders' analysis newsletter for document management.

See [http://www.ArchiveBuilders.com]

Note to Editors

Paper 22055v001

We will continue to update these articles as we get comments. Please contact us for the most current version before you publish. Also, please request permission to publish the article. Permission will be given freely for most purposes.

Steve Gilheany
Archive Builders
1209 Manhattan Ave
Manhattan Beach, CA 90266
Tel: +1 (310) 937-7000 Fax: +1 (310) 937-7001
[SteveGilheany@ArchiveBuilders.com]

Bio

Steve Gilheany, BA in Computer Science, MBA, MLS Specialization in Information Science, CDIA (Certified Document Imaging System Architect), AIIM Maser, and AIIM Laureate, of Information Technologies, CRM (Certified Records Manager, ARMA) has twenty years experience in document imaging and is a Sr. Systems Engineer at Archive Builders.

Author

Steve Gilheany is a Sr. Systems Engineer at Archive Builders. He has worked in digital document management and document imaging for twenty years.

His experience in the application of document management and document imaging in industry includes: aerospace, banking, manufacturing, natural resources, petroleum refining, transportation, energy, federal, state, and local government, civil engineering, utilities, entertainment, commercial records centers, archives, non-profit development, education, and administrative, engineering, production, legal, and medical records management. At the same time, he has worked in product management for hypertext, for windows based user interface systems, for computer displays, for engineering drawing, letter size, microform, and color scanning, and for xerographic, photographic, newspaper, engineering drawing, and color printing.

In addition, he has nine years of experience in data center operations and database and computer communications systems design, programming, testing, and software configuration management. He has an MLS Specialization in Information Science and an MBA with a concentration in Computer and Information Systems from UCLA, a California Adult Education teaching credential, and a BA in Computer Science from the University of Wisconsin at Madison. His industry certifications include: the CDIA (Certified Document Imaging System Architect) and the AIIM Master (MIT), and AIIM Laureate (LIT), of Information Technologies (from AIIM International, the Association of Information and Image Management, www.AIIM.org), and the CRM (Certified Records Manager) (from the ICRM, the Institute of Certified Records Managers, the official certifying body for ARMA International, the Association of Records Managers and Administrators, [www.ARMA.org]).

Contact:

SteveGilheany@ArchiveBuilders.com

Tel: +1 (310) 937-7000 Fax: +1 (310) 937-7001

For more information, courses, and papers:

[http://www.ArchiveBuilders.com]