

Taking a Fresh Look at Advanced Document Imaging Compression

Optimized PDFs Provide Benefits that Address Today's, as well as Tomorrow's IT Landscape

Foxit Software Inc.
41841 Albrae Street
Fremont CA 94538, USA

sales_ea@foxitsoftware.com
support_ea@foxitsoftware.com
foxitsoftware.com/enterprise-automation/

Sales: +1(347)-292-2714
Support: +1(718)-793-7878



Table of Contents

Taking a Fresh Look at Advanced Document Imaging Compression	3
Specific Benefits of Advanced Compression	4
How it Works	8
A Complementary Technology	9
Conclusion	10
About the Author	11
Footnotes	11

Taking a Fresh Look at Advanced Document Imaging Compression

When it comes to digital files, smaller has always been better. Smaller files use less storage, take less time to download, utilize less network bandwidth, and are easier to e-mail and share. However, in many cases, especially in document imaging applications, you are probably working with files that are 10-50 times larger than they need to be.

If you are not integrating advanced compression technology with your document capture process, your organization is spending extra money on storage, using up more network bandwidth, filling up inboxes faster, and forcing people to wait longer to download the files than they need do. PDF Compressor by Foxit is designed not only to alleviate these problems, but also to deal with a whole new set of issues related to file size—which are arising as we evolve into a world where mobile technology makes documents more portable than ever.

Historically, people have downloaded documents onto their PCs from websites or sent and received them as e-mail attachments. To transfer especially large files, they might have utilized an FTP connection. However, over the past five years there has been an explosion in the adoption of smart phones, tablets, and cloud storage. These new avenues of communication are expanding the way we consume, share, and store files.

Now, instead of just downloading files over wired internet connections, we are finding ourselves opening documents on mobile computers with cell phone networks, or uploading documents to a cloud server for sharing and collaboration.

This evolution into the age of portable documents has made the size of files more relevant than ever. Nobody likes to wait to wait an inordinate amount of time for a download, but the often varying speeds of mobile computing connections have compounded the issues of working with larger files. In addition, with mobile data plans often limiting the amount of data we can stream per month or costing more for additional data.com usage and cloud storage vendors charging per GB, minimizing file sizes is more important than ever.

In 2012, Ericsson, the Swedish telecommunications giant, announced it was working on a new compression technology designed specifically to “reduce the bandwidth requirements for video delivery by over 50 percent.”¹ The good news for organizations utilizing document imaging is that there is already technology on the market that regularly achieves compression rates on images of much greater than 50%—and can achieve it at a speed of thousands of pages per hour.

Foxit's PDF Compressor is a mature, proven application utilized by thousands of organizations worldwide to not only reduce the size of their PDFs, but also to create highly accurate full-text searchable images. PDF Compressor employs a range of technologies, including advanced image processing, gray scale thresholding, and mixed raster content (MRC) techniques, to achieve compression rates ranging from 10 to 1 on simpler, text-heavy black-and-white images, to more than 100 to 1 on graphics heavy color files, and OCR accuracy rates greater than 99% (when measured by characters).

Deployable as part of a capture process, PDF Compressor offers multiple benefits, including potentially reducing your total cost of storage for document images drastically—a savings that by itself will typically provide a cost justification for the software. PDF Compressor is complementary to existing capture and document management implementations—because it makes both applications work better. It also makes them more compatible with this emerging world of portable documents.

Specific Benefits of Advanced Compression

Let's take a closer look at some of the specific benefits that PDF Compressor can bring to a document imaging implementation:

1. Reduced storage costs:



In-House Storage



Cloud Storage

In recent years, the falling cost of hard drive space has caused many people to underestimate the true cost of storage. While it's true that the cost of hard drive space, when purchased in a large enough quantity, may be just pennies per GB, that is nowhere near the true cost of storage. When you factor in costs related to servers, back-ups, system maintenance, staff, and the facility used to power and house storage, the true cost of internal storage is

probably closer to \$10 per GB, per year,² although there are reports of internal IT departments charging as much as a \$30 per month per GB³. And while cloud storage is advertised at around \$.10 per GB per month by data hosting giants like Amazon and Google, this is simple file storage and not a document management system. Any organization storing mission critical files is typically inclined to utilize a hosted document management system. Depending on variables like user charges, volume, and features, the monthly cost for hosted document management can range from \$2 to \$50 per GB.

PDF Compressor can reduce file sizes and storage costs significantly. Let's say you are storing 1 TB worth of TIFF files. At an estimated cost of \$10 per GB per year for in-house storage (which probably on the low end), you would be paying approximately \$10,000 per year to store those documents. Based on tests run on sample documents by Foxit, applying PDF Compressor and converting those TIFFs to full-text searchable PDFs can reduce their size by approximately 90%.

1 TB of document storage becomes 100 GB, and your cost for in-house storage has dropped from \$10,000 per year to \$1,000. Granted, this represents "potential" savings and assumes the rest of the resources in the storage system can be fully re-allocated (possibly for storing more documents). However, even if only half the potential savings is actually realized, it's easy to see how PDF Compressor can pay for itself.

In the emerging world of cloud storage, the savings can be even more pronounced. Let's say the user above is paying \$10 per month per GB for cloud-based document management. That's \$120 per year per GB, or \$60,000 per year for .5 TB worth of document storage. (Because cloud storage typically includes back-up, it will not require as much raw storage to maintain the same volume of documents as in-house storage). Applying PDF Compressor, and its 90% compression rate for black-and-white TIFF files, can reduce the amount of cloud document management storage needed from 500 GB to 50GB, potentially decreasing annual cloud storage costs from \$60,000 to \$6,000.

Now, are you actually going to realize \$54,000 in annual savings on cloud storage by compressing a half-terabyte's worth of documents? Probably not, as user licenses are also often factored into the cost of cloud-based storage. Also, under most cloud vendor pricing models, you are going to pay significantly (probably at least twice as much) more per GB for lower volumes of storage than you do for a half-terabyte or more. Nonetheless, the cloud storage savings related to PDF Compressor can be significant and more than pay for the cost of the software. It could also be the deciding factor that enables your organization to finally move off a legacy document management application and into a more modern, easier to maintain, cloud environment.

These savings obviously increase or decrease depending on the volume of documents, as well as the type. For example, for color documents, for which compression rates can average 98%, the savings can be even greater.

2. Improved collaboration capabilities:



E-mailing images is a common and convenient way to share paper documents. According to the Radicati Group, in 2012, there were 89 billion business e-mails sent daily. According to the Osterman Research, 29% of corporate e-mails contain attachments—which mean there were approximately 25 billion e-mail attachments sent daily in 2012.

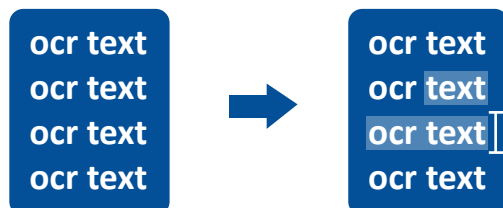
PDF Compressor is doubly useful when e-mailing document images. First, it ensures that scanned images are in a format - PDF- that most people can open without a special viewer. Second, it ensures that you are sending and receiving the smallest file sizes possible. This reduces strain on network bandwidth, e-mail servers, and desktop e-mail applications—as well as improves user experience by improving download and file opening speeds.

Bandwidth concerns become an even bigger factor when accessing files on mobile devices. Many data plans limit customers' bandwidth usage per month. Downloading a 6 MB uncompressed JPEG image of a document is going to use up 17 times more bandwidth than downloading a 357K version of the same file after it has been run through PDF Compressor.

A recent Nielson survey reported that in 2012, the average mobile data plan customer was utilizing twice as much data as in 2011⁴, with indications that data usage will continue to grow. Multiple devices sharing a single corporate data plan compound concerns about bandwidth issues.

The bottom line is that no one wants to unnecessarily use up bandwidth that might be needed for other tasks later. In addition, data overages can get expensive, or become inconvenient if the data plan provider slows down transfer rates after users reach their limits. Utilizing PDF Compressor will enable your organization to create files that require 10% or less the bandwidth of uncompressed files. This could apply to service manuals, forms, reports, legal and financial documents—any type of file an employee would need, or want, to e-mail and/or access on a mobile phone or tablet.

3. Fully searchable PDF files:



PDF Compressor includes state-of-the-art OCR technology for creating full-text searchable PDFs. The OCR results are included within the PDF, which make them portable with the file. In contrast, document images saved as TIFFs and JPEGs need to be tied to separate data files containing the OCR results.

PDF Compressor applies proprietary image processing to achieve OCR results that have been tested by Foxit to be 15% more accurate (on a character level) than results from a leading OCR engine. Having a highly accurate representation of the text associated with a document image is especially valuable in applications in areas like legal, where multiple document types are often intermixed without any pre-defined indexing system. Full-text search ability can be important in archiving because you often don't know what types of searches will be done several years down the road—especially for matters concerning e-discovery.

In regards to archiving, PDF Compressor also has the ability to output PDF/A files—an ISO standard designed to ensure long-term accessibility. PDF/A is widely accepted as an archiving standard in many European countries and is starting to gain traction in the North America, especially in regulated industries.

How it Works

PDF Compressor's has multiple adjustable settings. They fall into five major categories:

1 Image processing

PDF Compressor features a number of options to clean up images. It also offers the thresholding of gray scale and color images—which increases the contrast between foreground and background data—to create optimum bi-tonal files.

2 MRC segmentation

MRC involves the separation of a file into layers before it is compressed. MRC is utilized for achieving maximum compression on color images. MRC separates files into text and graphics layers and then compresses each layer differently.

3 Compression

PDF Compressor includes options for four different compression methods. JPEG and JPEG 2000 compression are available for color or gray scale documents and/or graphic elements of MRC documents. JBIG2 and Group 4 are available for black-and-white documents and/or textual elements of MRC documents.

4 OCR

On the most basic level, you can choose to turn OCR on or off. OCR data does not add much to the file size (maybe 5%), but does slow down the conversion process (see below for details). Up to 117 different languages can be considered when the OCR process is run. PDF Compressor offers four modes of OCR. To help decide which mode is right for your application, the software can produce reports on its confidence level for accuracy on a set of documents running under a specific mode. PDF Compressor also offers data on how long it takes to create a particular type of searchable PDF file.

5 Speed

In Foxit's testing, PDF Compressor's output is approximately 40,000 pages per hour per core without OCR and 3,000 pages per hour per core with OCR. PDF Compressor is designed to take full advantage of a quad-core CPU—by enabling each core to individually process pages—so those quoted speeds will quadruple with a quad-core processor. Specific OCR settings will also affect output rates.

PDF Compressor includes a Quick Set Up Wizard that asks a series of questions and uses the answers to automatically suggest your optimum settings. Sample output files are created, along with information on how long it took to produce each sample file. If you are satisfied with one of these samples, the settings can be saved as a profile. If not, the Wizard will ask additional questions and create a new set of samples.

A Complementary Technology

PDF Compressor is a great fit in any current document imaging implementation. It sits between your capture and document management and helps optimize the performance of both applications. It helps your capture software to output higher quality, more usable files, and your document management software runs more efficiently because it's utilizing less storage space and offering smaller files for download.

PDF Compressor can be integrated directly with leading capture application like Kofax Capture and EMC Captiva through pre-written connections. PDF Compressor can be integrated with other capture and document management applications on the command line level. It can also be configured as a standalone process, which automatically ingests and converts files from watch folders and outputs them into a file system.

PDF Compressor is a fit for both new and existing document imaging applications. For a new user of document management technology, it offers the opportunity to have a system optimized for storage, as well as for downloading to mobile computers. For a legacy user, it offers a relatively inexpensive way to modernize and refresh a system without having to invest in a whole new document imaging system.

PDF Compressor is also a great tool to help with moving legacy applications to cloud environments. If you have been hesitant to transition from an old document management system, even one for which a vendor is providing minimal support for, to a more modern cloud environment because of the perceived cost of storing all your legacy documents in the cloud— PDF Compressor can be used to significantly reduce cloud storage costs.

Implementing PDF Compressor will also better enable you to take full advantage of the most advanced scanner technology. Most high-speed document scanners are now able to produce high quality color images at the same speed with which they produce bi-tonal and gray scale images. However, many organizations do not take advantage of this capability to create and store the most accurate representation of their documents because of concerns about having to store the larger color files. With PDF Compressor's ability to reduce color file sizes by more than 95%, storage concerns about color should become moot and you can start taking full advantage of your scanners.

Conclusion

The bottom line is that PDF Compressor offers some advanced and mature technology that specifically meets many existing and emerging needs of organizations utilizing document imaging technology. In review, PDF Compressor can help you create smaller, full text-searchable PDFs, which are valuable for the following reasons:

1

Reduced storage costs

Due to the publicity of consistently falling hard drive prices, many people don't realize how expensive the true cost of storage really is. When you factor in costs related to items like servers, back-ups, system maintenance, staff, and facilities, storage costs add up. The savings are even clearer when users move their documents into cloud storage, where there is often a fee per GB used. PDF Compressor has the ability to reduce document image file sizes by more than 90%, which can potentially reduce storage costs by a similar percentage.



2

Smaller file downloads

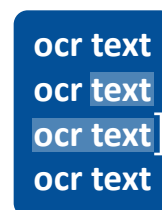
Creating smaller files improves performance when you are sharing files through e-mail, and it becomes especially important as users migrate to using more mobile devices—with bandwidth usage on mobile plans limited and larger files using up more bandwidth.



3

Accurately searchable files

PDF Compressor offers above industry-standard OCR results, which can be packaged with the PDF, or PDF/A files for long-term search ability.



About the Author

Ralph Gammon has served as editor and publisher of the Document Imaging Report newsletter since 2002. Prior to that he was an editor at Jameson Publishing for five years, focusing on the document imaging and ECM industries. Ralph remembers when PDF was considered a proprietary format that would never catch on. In addition to a newsletter, Gammon authors the popular Document Imaging Talk blog (<http://documentimagingreport.blogspot.com/>) and is a frequent Tweeter @DIREditor.



Footnotes

1. <http://www.infoworld.com/d/mobile-technology/new-video-compression-technology-willcut-bandwidth-use-in-half-says-ericsson-200747>
2. File Storage Costs Less In The Cloud Than In-House, by Andrew Reichman, Forrester Research, 2011.
3. <http://hardware.slashdot.org/story/10/07/29/1952235/internal-costs-per-gigabyte-whatdo-you-pay>
4. http://www.phonearena.com/news/Heres-the-amount-of-data-used-by-the-average-U.S.-subscriber_id32592