

D-Lib Magazine July/August 2008

Volume 14 Number 7/8

ISSN 1082-9873

A Format for Digital Preservation of Images

A Study on JPEG 2000 File Robustness

[Paolo Buonora](#)

Archivi di Stato, 00150 Roma, Italia
<paolo.buonora@beniculturali.it>

[Franco Liberati](#)

Università degli Studi di Roma "La Sapienza", 00149 Roma, Italia
<liberati@di.uniroma1.it>

Abstract

Digital preservation requires a strategy for the storage of large quantities of data, which increases dramatically when dealing with high resolution images. Typically, decision-makers must choose whether to keep terabytes of images in their original TIFF format or compress them. This can be a very difficult decision: to lose visual information though compression could be a waste of the money expended in the creation of the digital assets; however, by choosing to compress, the costs of storage will be reduced. Wavelet compression of JPEG 2000 produces a high quality image: it is an acceptable alternative to TIFF and a good strategy for the storage of large image assets.

Moreover, JPEG 2000 may be considered a format that can guarantee an efficient robustness to bit errors and offers a valid quality with transmission or physical errors: this point of view is confirmed by the case study results that we report in this article, concerning image quality after occurrence of random errors by a comparison among different file formats. Easy tools and freeware software can be used to improve format robustness by duplicating file headers inside or outside the image file format, enhancing the role of JPEG 2000 as a new archival format for high quality images.

Introduction: current trends

In recent years the JPEG 2000 format has been widely used in digital libraries, not only as a "better" JPEG to deliver medium-quality images, but also as new "master" file for high quality images, replacing TIFF¹ images. One of the arguments used for this policy was the "lossless mode" feature of JPEG 2000; but this type of compression saves only about the half of the storage

requirements of TIFF, so it is unlikely that this was the only reason that digital libraries moved in this direction. The only reasonable choice was the standard lossy compression, which offers a 1:20 (color) or 1:10 (grayscale) ratio. This provides a significant savings in terms of storage, considering that the quality of images in digitization projects has increased dramatically in the past few years: the highest standards for image capture are now very common in digital libraries.

Thus, the argument turned from the "mathematically lossless" concept to a softer "visually lossless" definition, and the question became: what do we lose in choosing the JPEG 2000 "lossy" mode? Let's focus on the following definitions:

"The image file will not retain the actual RGB color data, but it will look the same because screens and our eyes are so forgiving"²

"... many repositories are storing "visually lossless" JPEG 2000 files: the compression is lossy and irreversible but the artifacts are not noticeable"³

As mentioned above, some institutions began to store JPEG 2000 files in their digital repositories as the "archival format"⁴. This policy was sometimes officially declared, or in some cases was adopted *de facto*.

"The migration process involves creating a derivative master from the original archival master..." or, as shown in the example of the following migration rationale:

"<migration_rationale> Create JPEG 2000 datastream for presentation and standardize on JPEG 2000 as an archival master format. </rulibadmin:migration_rationale>"⁵

One of the most relevant and specific examples of format migration to JPEG 2000 was made at the Harvard University Library (HUL):⁶

"HUL chose to perform a migration of various image files to the JPEG 2000 format. There is great local interest at Harvard in the retrospective conversion of substantial numbers of existing TIFF images to enhance their utility by permitting the dynamic image manipulation facilitated by the JPEG 2000 format. The three goals that guided the design of the migration were:

- To preserve fully the integrity of the GIF, JPEG, and TIFF source data when transformed into the JPEG 2000 (JP2) format
- To maximize the utility of the new JP2 objects
- To minimize migration costs"

The Xerox Research Center, namely Robert Buckley, was involved in this strategy, producing studies about the integration of JPEG 2000 in the OASIS Reference Model and defining it as a *digital preservation standard*.⁷

Although Buckley's *Technology Watch Report* has been accepted and promoted by the Digital Preservation Coalition in the UK, many relevant experts in this

field still seem to show some skepticism and continue to take a "wait and see" position:⁸

"... some institutions engaged in large-scale efforts are considering a switch to JPEG 2000 ... However, the standard is not yet commonly used and there is not sufficient support for it by Web browsers. The number of tools available for JPEG 2000 is limited but continues to grow".⁹

Tim Vitale's opinion on JPEG 2000 was very clear in his 2007 report:¹⁰

"It is not an archival format ... Existing web browsers (mid-2007) are not yet JPEG 2000 capable. One of the biggest problems with the format is the need for viewing software to be added to existing web browsers ... There are very few implementations of the JPEG 2000 technology, more work needs to be done before general understanding and acceptance will be possible."

However, this is no longer the case: most common commercial, digital imaging programs now support JPEG 2000, not to mention JPEG 2000 support by some excellent shareware.¹¹ The real problem is that the JPEG 2000 format allows the storage of very large images, and no current programs can manage the computer memory in an intelligent way: this is the commercial reason for professional image servers and encoders, which are relatively costly,¹² or specific viewers for geographic images (generally free¹³), or browser plug-ins (free as well).

1. Image compression of continuous tone images

The primary objection to JPEG 2000 compression remains the possible loss of visual information. Our approach in arguing against this will not focus on how the wavelet approach works,¹⁴ but why it works, with some very basic elements of compression theory.¹⁵ In other words, preserving visual information deals mainly with how the images are perceived visually, and only secondarily deals with the mathematical aspects of the physical signal (materials, procedures, techniques).

Some would argue that images look the same as they did before compression simply because humans don't see very well, and that a deeper examination (or a better monitor) would reveal errors and losses. This is not true: even when JPEG 2000 images are enhanced by magnification, no human could perceive any errors or losses. A digital surrogate is not necessarily a bad copy of the original, and compression does not always mean loss of information.

Some people also may think that compression is the equivalent of the "sampling" of a signal; for example, if we choose 300 points per inch to represent an object, sub-sampling might take only 150 or 100 points instead, which creates the risk of losing some information essential for reconstructing that signal. Any sampling below the Nyquist rate produces aliasing effects: if we represent the signal as a wave, the sampling interval should match exactly the shape of the wave. Otherwise, original images are "misunderstood" and appear as artifacts. But compression is not a kind of sub-sampling made after the capture of an image.

We can either eliminate *redundant* information (a sequence of identical values), or we can have some kind of lossless compression, but below the physical-mathematical reality, we can operate on the human perception of it. Since we are dealing with the information that we perceive with our eyes, we can compress *irrelevant* information, i.e., what is less relevant to our senses. The human eye is less sensitive to colors than to light, so the chrominance signal can be compressed more than the luminance signal can, without any loss of perception.

This is very important with digital images of historical documents, as they are usually either color or grayscale images, i.e., "continuous tone" images. As opposed to a "discrete tone" image (as a printed or typed document in black and white), in a continuous tone image any variation of adjacent pixels is relevant: in other words, pixels are "correlated" with each other. We cannot retrieve a sequence of identical values to compress, and we need a more sophisticated strategy.

We can select a part of the image, an array of pixels, and calculate the average of the values; then, we can calculate the difference of any single value from the average. This is called "de-correlation" of the image pixels, and at the end of this process we will find that many of the differences from the basic average value are 0, or almost zero, so we can easily compress the image by assigning them the same values (quantization).¹⁶ When we separate the three color channels, each of them can be considered as a grayscale image, and we can use the "bit planes" technique.¹⁷ For example, let us take three adjacent pixels in a grayscale image, with very different values, in a decimal and in a binary code:

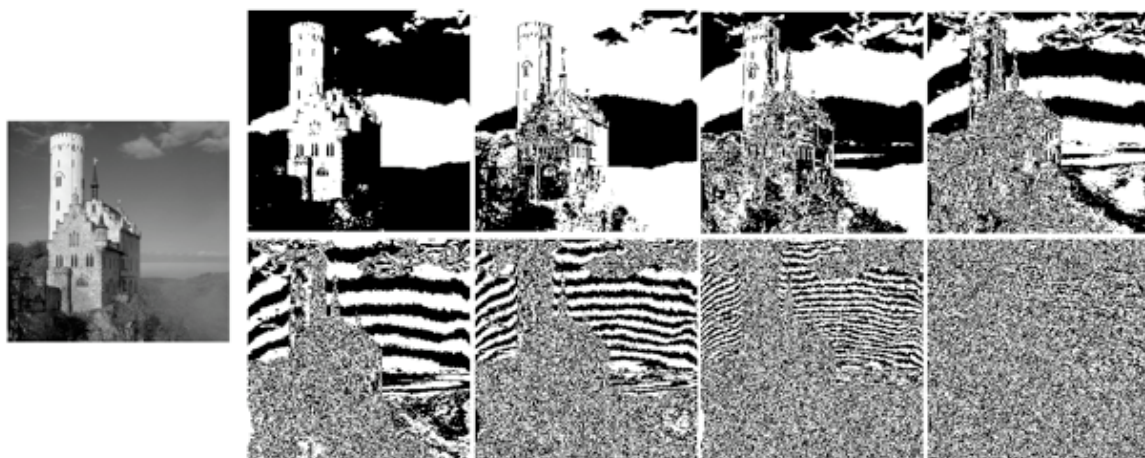


Figure 1: An 8-bit grayscale image and its bit planes.

$$10 = 000001010$$

$$3 = 000000011$$

$$-7 = 100000111$$

The image is at 8-bit depth, so we have 1+8 bits (the first represents +/- sign). At positions 2,3,4,5 (i.e. at bit-plane 2,3,4,5) we find only "0", and at position 8 find only "1": this is also expressed by saying that the relevance of the information or *energy* (low frequencies) concentrates at certain levels, and the other levels (high frequencies) can be easily compressed.¹⁸ This is very clear in

the following representation of an image in 8-bit planes: continuous tone variations between adjacent pixels are now turned in eight separate contexts, where it is now possible to compress adjacent values.



Figure 2: A corrupted JPEG file.

There are two main methods for de-correlating pixels: orthogonal transform and subband transform. The concept of "transform" is easy to understand from a geometrical point of view: a transform, as a reflection, "is a mathematical concept, but it is not a shape, a number, or a formula"; it is more "a way to move things in space"¹⁹ to operate the Hi/Lo frequency separation mentioned above. The DCT (discrete cosine transform) is the typical orthogonal transform that has been used in JPEG compression for many years.

In the JPEG compression, color images are decomposed in an YCbCr color space (Y is the *luma*, or the brightness in an image, Cb and Cr are blue and red *chroma* components, respectively): the luminance component is the most relevant to human eyesight, so it is less compressed than the other two components. JPEG can use DCT to break up an image to its spatial frequency components, and it compresses the low-frequency component first. This important - but optional - feature is called progressive encoding. Unfortunately, JPEG is generally used in a sequential mode, rather than in a progressive mode, so when data is corrupted, the encoding/decoding process fails and the rest of the image is lost.

2. JPEG 2000

The subband transform implies that we consider a signal in the frequency domain, applying different algorithms than the Fourier DCT, currently known as "wavelets",²⁰ as these mathematical functions are graphically represented as small waves moving up and down the x axis. The wavelet transform is bi-dimensional: it splits the array of values in the upper/lower and then the

right/left elements, calculating the average and concentrating low frequencies in the top-left side of the array. This process of subband decomposition is repeated again and again in a progressive mode. In the reverse decoding, we can see a blurred image that becomes increasingly sharp, because each subband level adds new details to the basic graphic information stored at the top-left of the image array.



Figure 3: Progressive subband decomposition.

The consequence of this process is a "multi-resolution"²¹ image: in the file we have the same image at different resolutions, from the high-resolution version to the thumbnail preview that we can use on our catalogue webpage.²² With the current availability of the ISO Standard multi-resolution format JPEG 2000, there is no reason for keeping three different formats of the image - i.e., an original TIFF in high quality resolution, a medium resolution JPEG for Internet display, and a thumbnail for the catalogue webpage. The software - in either a local or a remote client/server context - will choose for us the resolution required.²³ Moreover, the main feature of wavelet compression is quality.²⁴ Not only in the lossless mode (ratio 1:2), but also in the standard lossy compression (1:10 for grayscale, 1:20 for color), it is very hard to detect differences with the original TIFF: with JPEG 2000 the DCT transform (discrete cosine transform) is replaced by a DWT (discrete wavelet transform), and we can no longer see the typical "pixelization" effect that we saw when using the previous JPEG format.²⁵

Compared to the previous JPEG, the JPEG 2000 format offers others several new features:

- We can store color information at 48-bit color depth; this is important because this can be done with the TIFF format while it could not be done with the previous version of JPEG. Therefore, we will not lose color depth if we compress the image in JPEG 2000, and 48-bit or 16-bit depth is relevant, especially in digitized photographs.
- We can manage large images and very large images. Multi-resolution formats, in fact, were developed to manage satellite mosaic images and for large geographic maps.
- We can also specify a ROI (region of interest²⁶) for implementing very high resolution in some parts of the image and leave the rest at a lower resolution.
- We can store metadata inside the file format, in particular geographic coordinates. (In the geo-TIFF format we needed separate files for coordinates.)

- Finally, we can manage Intellectual Property Rights (IPR) effectively, because for high quality (or "master") images, we can prevent the download of the whole image. The user will see only the portion of the image that he or she requires, and we can eventually watermark on-the-fly the visualization window.

How does JPEG 2000 work in practice? Colors are separated in three components: YCbCr. (This was possible in the previous JPEG format too, but it was only an option as progressive coding/decoding.) The image in each component is usually divided into more parts (tiles), and subdivided again in a grid (precincts) and small portions (code-blocks) that are scanned with the algorithm. All the data are then packed in a tagged file structure that we can imagine as a "matrioska" system of boxes. At the beginning of the main box is the "header box", inside the box we find other boxes for each tile with a "tile header" at the beginning, and inside these boxes there are markers and packet streams containing image data.

It is easy to guess that these headers are more relevant than the rest of the information: if they are corrupted, the entire image, or the tile or a portion of the tile, will be lost. As the extension of the headers comprises a very small percentage of the whole file, it would be easy to improve the robustness of the file, by making a copy of this information inside or outside the file:²⁷ this principle of redundancy is frequently used in information technology to protect the data.²⁸

3 File structure and robustness

As was recently shown by Judith Rog, the old concept of compression techniques as an obstacle to the digital preservation of images seem to vanish considering deeper file formats structure .

In the error control features offered by JPEG 2000, two types of errors are considered: bit errors and packet losses (during transmission). The use of error control mechanisms depends on the coder and decoder implementation. The impact of bit errors mainly affects the error location, while packet losses have a dramatic effect on the image because an error involving packets implies a synchronization loss between coder and decoder. Bit errors hence affect only the concerned code-block; for this reason code-blocks are coded independently.

In the case of packet losses, JPEG 2000 offers the opportunity to use two kinds of termination strategies of the arithmetic coder on bit-plane level, enabling the detection of bit errors and the deletion of corrupted data. In addition, using smaller precincts increases the robustness against bit errors, although the coding efficiency is decreased. In particular, maintaining synchronization plays an important role in JPEG 2000. Therefore, Start Of Packet (SOP) marker segments can be inserted into the data stream prior to each packet. Using small precincts also increases error robustness in the packet loss scenario, because the amount of lost data is reduced.

In order to maintain synchronization, the sequence numbers contained in each of the SOP marker segments have to be counted. If the sequence number does not increment by one, a packet has been lost, and it is replaced by an empty packet with a correct sequence number. In the case of transmission, the packets

belonging to lower bit-planes in the same precinct have to be discarded as well. In addition, the same technique can be applied in the case where an SOP marker segment is not found at the expected position, which means that a bit error has occurred in the previous packet header.²⁹

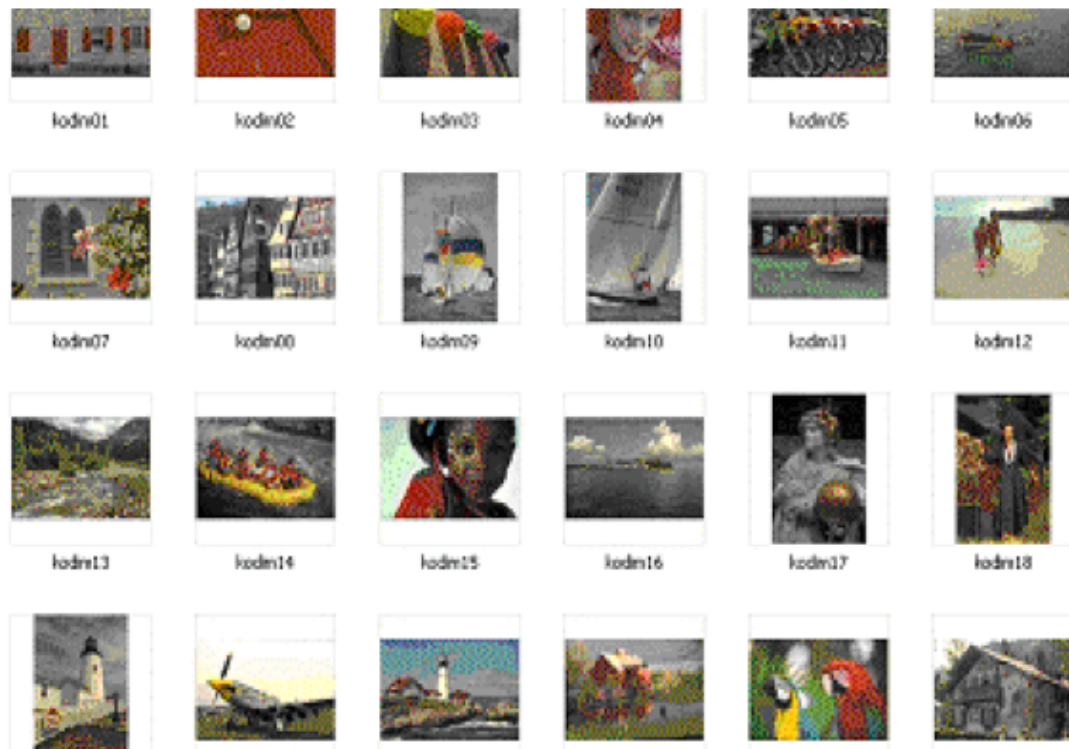


Figure 4: Kodak images test set.

4 Testing robustness

We investigated the impact of random errors on the JPEG 2000 bit stream by conducting some experiments. The image set we chose, proposed by Kodak, offers all possible chromatic and color-depth cases: the 24 images (Figure 4) are bitmap. They have 24-bit color depth and a dimension of 768 x 512 (1.153 KB for each one).

Original bitmap files were converted in TIFF (uncompressed) and then in JPEG and JPEG 2000 files. On every test image, we applied different compression ratios, which were selected considering preliminary studies. The following ratios were chosen: lossless, 1:5, 1:10, 1:15, 1:20, 1:25, and 1:30. We excluded major values because, usually, both JPEG and JPEG 2000 are used with 1:10 for grayscale images and 1:20 for true color images. We introduced errors writing a byte <0> if the pixel we changed had a value greater than or equal to <127>, and a byte of <255> if the pixel value was less than <127>. The error number was related to image dimension: 0.01% (around 10 byte); 0.1% (around 100 byte) and 1% (around 1000 byte). We didn't exceed the 1% errors because we found that after 1%, the errors percentage of any image format collapses.

The basic point is that all errors do not occur in the header (for TIFF, JPEG and JPEG 2000), where a single corrupted byte can compromise the integrity of the whole file. For this reason, we created a particular routine, generating numbers

that make the position of random errors evident, between the header and the end of file.³⁰

After this step, we tested every image by adopting a mathematical measure. The most widely used objective quality metrics is root-mean-squared error (RMSE) defined as:

$$RMSE = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [I(i,j) - I^*(i,j)]^2}$$

where I is an $M \times N$ image and I^* is the corresponding reconstructed image after compression and decompression. For each component (RGB) we estimated the RMSE and then we calculated the average. If an image could not be opened, we imposed RMSE values as 255.³¹

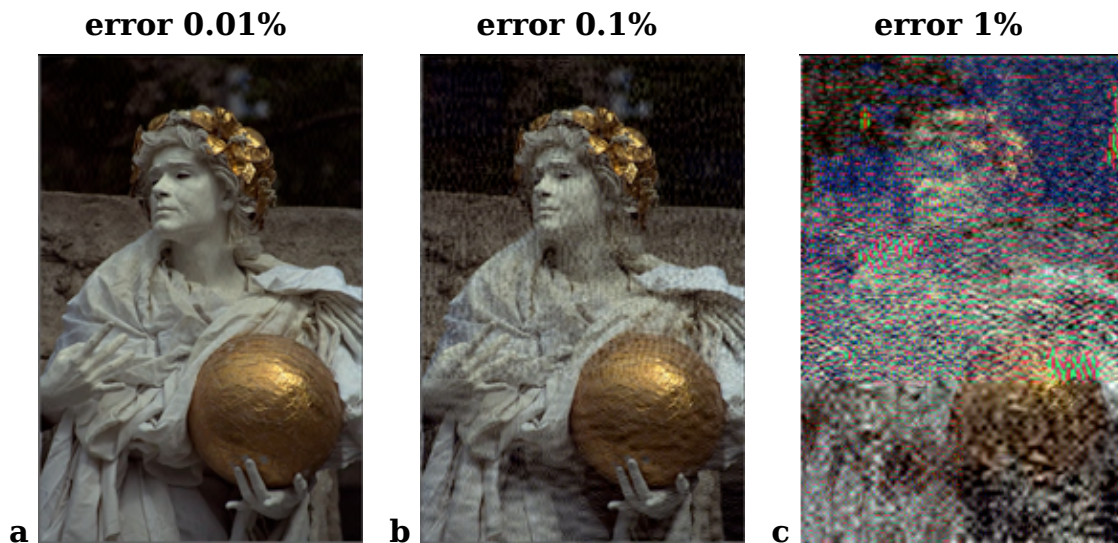


Figure 5: a corrupted JPEG 2000 image, compression ration 1:20

As shown in Figure 5a, an error rate of about 0.01% doesn't imply serious effects: the reconstructed image is very similar to the original. In Figure 5b, "noise" can be seen on a huge part of the image, but the subject of the image and its main characteristic are preserved. Figure 5c shows a bad result: in this case, we lost around 1Kb of information in different parts of the file (crucial tile, markers, etc.).

error 0.01%

error 0.1%

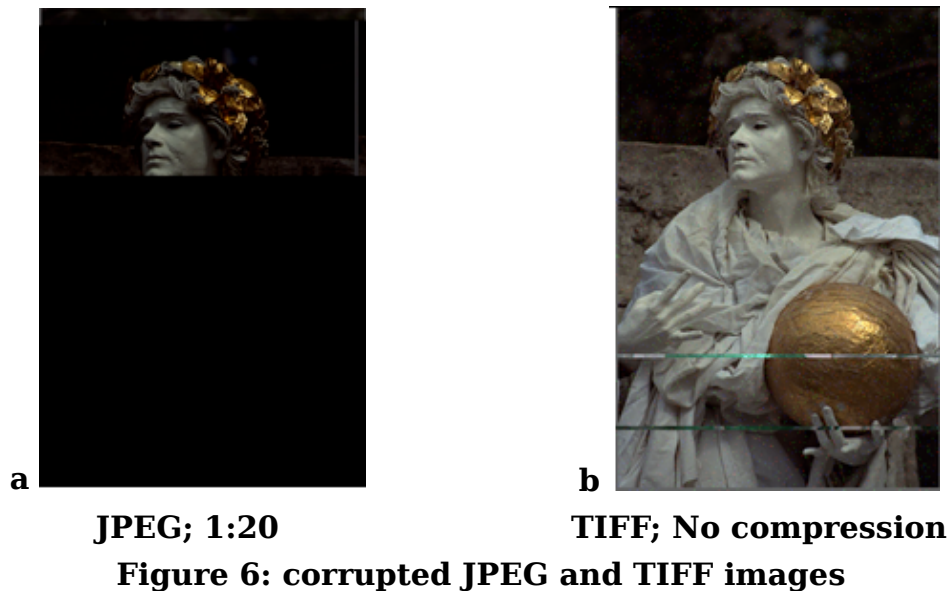


Figure 6a is a JPEG format: The image is irretrievably corrupted with only 0.01% of errors: a simple error can affect the entire bitstream and prevent correct decoding. With increasing errors, the file image cannot be opened.

In Figure 6b, we see some of the consequences of errors on a TIFF format. At 0.1% of errors the image has some lines corrupted. The TIFF format has an end-line dedicated byte; therefore, removing or corrupting this value can damage a strip. In other cases, it is not possible to open the files if 0.1% errors are introduced (7 have failed: 30% tested image, see Table 1) and 1% errors occur (24 have failed: 100% tested image) .

| Table 1 - Experimental Results: TIF error 0.1% | | | |
|---|-------------|----------------|-------------|
| IMAGE | RMSE | IMAGE | RMSE |
| Kodim01 | 5.21 | <u>Kodim13</u> | 255.00 |
| <u>Kodim02</u> | 255.00 | Kodim14. | 5.86 |
| Kodim03 | 5.00 | Kodim15 | 10.43 |
| <u>Kodim04</u> | 255.00 | Kodim16 | 5.35 |
| Kodim05 | 6.02 | Kodim17 | 7.25 |
| Kodim06 | 7.68 | <u>Kodim18</u> | 255.00 |
| Kodim07 | 5.22 | Kodim19 | 5.32 |
| <u>Kodim08</u> | 255.00 | Kodim20 | 7.00 |
| Kodim09 | 4.93 | Kodim21 | 5.19 |
| Kodim10 | 5.00 | <u>Kodim22</u> | 255.00 |
| Kodim11 | 5.48 | <u>Kodim23</u> | 255.00 |
| Kodim12 | 5.55 | Kodim24 | 5.62 |

As shown in Table 2 and Fig.7, JPEG 2000 offers good performance in most situations. TIFF offers stability and quality only if error numbers are less than 0.1% with respect to image dimension. The previous JPEG format appears to be

the worst format for preservation: in every case (high or low error rate), noise and error propagation is present on images.

| Table 2 - Experimental Results: RMSE values with different % error | | | |
|---|-------------|------------|----------|
| Error (%) | 0.01 | 0.1 | 1 |
| TIFF | 1.667 | 78.651 | 255.000 |
| JPEG | 101.982 | 115.392 | 255.000 |
| JP2 lossless | 4.755 | 38.604 | 92.297 |
| JP2 compression ratio 10 | 5.660 | 48.377 | 102.456 |
| JP2 compression ratio 20 | 11.103 | 35.459 | 91.646 |
| JP2 compression ratio 30 | 11.058 | 42.478 | 95.784 |

JPEG 2000, however, uses markers, headers and dedicated coding to prevent coding errors or transmission errors. It offers good performance at any error rate and with several compression ratios. It is significant to note that JPEG 2000 lossless files are not more robust than lossy ones. In particular, we tested compression at the ratio of 1:10, typically used for grayscale images, and 1:20, used for color image; the compression ratio 1:30 is only tested to show a possible trend in quality loss, because with JPEG 2000, there is no purpose for adopting any compression ratio more than 1:20. Otherwise, just as with TIFF and JPEG, JPEG 2000 has problems whenever errors exceed 1% (around 1Kb): some images cannot be opened, or a significant part of the information is no longer visible.

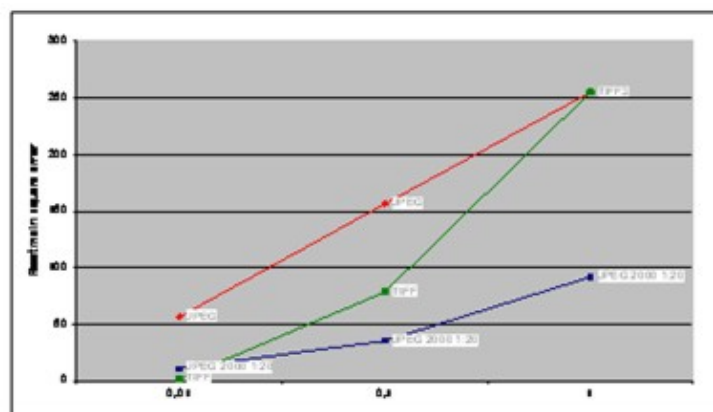


Figure 7: Experimental results: JPEG, TIFF and JPEG 2000 1:20.

5. Smart tools for image file preservation: FixIt!: for JPEG 2000

JPEG 2000 file structure is not only robust itself, but moreover it is possible to add some additional robustness enhancements to the file structure. In all formats, the header image is a crucial element: the header supplies the basic

information enabling the image to be visualized. It is a sequence of binary values, usually at the beginning of a file, where width, height, color information, and other information is held. The image may appear corrupted or not accessible even if only some bytes in the header are lost. For example, if we manipulate an original JPEG 2000 image header (see example in Figure 8a and 8b) introducing some error (only 4 bits), we cannot open the image, although the rows are integral in the remaining part. If we could duplicate this header and keep it inside or outside the image file format, we would dramatically improve its robustness.



Figure 8a - JP2000 header: highlighted values represent image height



Figure 8b - Impact on the image of corrupted header height values

FixIt! JPEG 2000 is a freeware and shareware utility implemented by the Laboratorio Digitale of the Centro di Fotoriproduzione, Legatoria e Restauro of the Italian State Archives, and can be downloaded from our website.³² It can extract the JPEG 2000 header; test and eventually fix corrupted image files; analyze a file's information and main markers, and save it in XML format. It can be also used in a recursive mode for a large number of files in a directory tree (see Figure 9a and 9b).

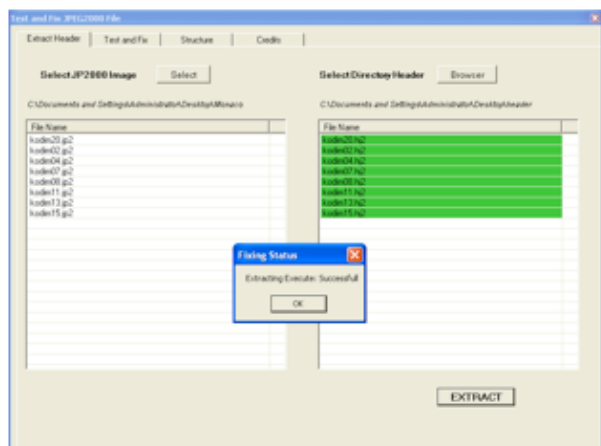


Figure 9a - Extracting and testing image headers

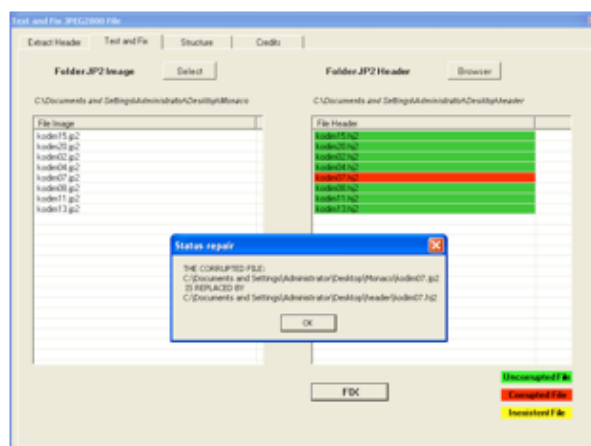


Figure 9b - Fixing corrupted files

6. Conclusion

Based on the results of the studies, we conclude that JPEG 2000 compression is

a good current solution for our digital repositories. But we do not have to provide solutions only in our cultural institutes: the digital assets created in our society in everyday life will eventually become a part of the cultural heritage. Implementing wavelet compression, and saving crucial information in extra file headers, offers everyone a flexible and inexpensive strategy for maintaining image data into the future.³³

| Comparing file format robustness | | | |
|---|-------------|------------------|-------------|
| | TIFF | JPEG 2000 | JPEG |
| Lossless Compression | ✓ | ✓ | ~ |
| Lossy Compression | ✓ | ✓ | ✓ |
| Large Images Management | ✓ | ✓ | ~ |
| Quality Compression | ✓ | ✓ | ✓ |
| Region of Interest | ~ | ✓ | ~ |
| Error Robustness | ✓ | ✓ | ~ |
| ✓ = optimal ✓ = low ~ = none | | | |

Notes and References

1. Editor's Interview: JPEG 2000. Dr. Daniel Lee, ISO SC29/WG1 (JPEG), in *RLG DigiNews*, Volume 6, Number 6, December 2002. See <<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000070519:000006287816&reqid=24150#interview>>. For general trends in digital imaging, see Steven Puglia - Erin Rhodes, *Digital Imaging - How Far Have We Come and What Still Needs to be Done?*, in *RLG DigiNews* Volume 11, Number 1, April 15, 2007. See <<http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000068890&reqid=26384>>.
2. Tim Vitale. *Digital Image File Formats - TIFF, JPEG, JPEG 2000, RAW, and DNG*. July 2007, Version 20. Available at <http://aic.stanford.edu/sg/emg/library/pdf/vitale/2007-07-vitale-digital_image_file_formats.pdf>. Vitale does not appear to be an enthusiast of JPEG 2000 "JPEG and JPEG 2000 encode the original RGB image data (to the video standard YCbCr), altering the original numerical digital data permanently... (p. 2), and further: "called Lossless, but uses YCbCr color space (50-67% color space compression); a roundtrip image will appear uncompressed, but RGB numbers are not original due to rounding and wavelet decompression errors" (p. 32).
3. Robert Buckley, *JPEG 2000 - a Practical Digital Preservation Standard?* *Technology Watch Report*, Digital Preservation Coalition February 2008. Available at <<http://www.dpconline.org/docs/reports/dpctw08-01.pdf>>.
4. Betsy Kruger, Library of the University of Illinois at Urbana-Champaign, *D-Lib Magazine*, January/February 2008, Volume 14 Number 1/2: <[doi:10.1045/january2008-featured.collection](http://dx.doi.org/10.1045/january2008-featured.collection)>.

5. Ronald Jantz, Michael J. Giarlo, Digital Preservation. Architecture and Technology for Trusted Digital Repositories, *D-Lib Magazine*, June 2005, Volume 11 Number 6, <[doi:10.1045/june2005-jantz](https://doi.org/10.1045/june2005-jantz)>.
6. Stephen Abrams, Stephen Chapman et al., Harvard's Perspective on the Archive Ingest and Handling Test, *D-Lib Magazine*, December 2005, Volume 11 Number 12, <[doi:10.1045/december2005-abrams](https://doi.org/10.1045/december2005-abrams)>. See also Justin Littman, A Technical Approach and Distributed Model for Validation of Digital Objects, *D-Lib Magazine*, May 2006, Volume 12 Number 5, <[doi:10.1045/may2006-littman](https://doi.org/10.1045/may2006-littman)>.
7. Robert Buckley, William Stumbo, and Jim Reid, Xerox Research Center Webster (USA), The Use of JPEG 2000 in the Information Packages of the OAIS Reference Model, in *Archiving 2007. Final proceedings of conference held May 21-24, 2007, Arlington, Va.* Springfield, Va.: The Society for Imaging Science and Technology.
8. "The promising - but not yet generally accepted - JPEG 2000" says Edwin Klijn, The Current State-of-art in Newspaper Digitization. A Market Perspective, *D-Lib Magazine*, January/February 2008, Volume 14 Number 1/2: <[doi:10.1045/january2008-klijn](https://doi.org/10.1045/january2008-klijn)>.
9. Oya Y. Rieger, *Preservation in the Age of Large-Scale Digitization. A White Paper*, February, 2008 (CLIR pub 141), p. 20, <http://www.clir.org/pubs/abstract/pub141abst.html>; Rieger mention after "new archival file formats, such as JPEG 2000 and PDF/A" (p. 33).
10. See again Tim Vitale. Digital Image File Formats, p. 37.
11. Both Photoshop and AcdSEE can manage JPEG 2000 files; Irfanview is free, and can convert any file format, also if they have very large dimension: <<http://www.irfanview.com>>.
12. See, for example, Lizardtech, <<http://www.lizardtech.com>>, Luratech <<http://www.luratech.com>>, and ErMapper <<http://www.ermapper.com>>.
13. See, for example, Mapping Science GeoJP2 Viewer, <http://www.lizardtech.com/download/dl_options.php?page=third>, or Lurawave, &tl;<http://www.luratech.com/download/jp2>>.
14. There are several excellent and specific expositions concerning the JPEG 2000 algorithm, including <http://en.wikipedia.org/wiki/Jpeg_2000>. We only mention here the following official documents:
 - ISO/IEC, *Information Technology-JPEG 2000 Image Coding System, Final Committee Draft Version 1.0*, ISO/IEC JTC 1/SC 29/WG 1 N1646R, March 2000.
 - ISO/IEC 15444-1, *The JPEG 2000 Image coding system-Part 1: core coding system*, - ISO/IEC/ITU-T, 2000.
 - Joint Photographic Experts Group, Website <<http://www.jpeg.org/jpeg2000/index.html>>
15. D. Salomon, *Data compression. The complete reference*, New York-Berlin, Springer, 2004 (3rd ed.); online exercise and extra

<<http://www.davidsalomon.name/DC3advertis/DComp3Ad.html>>.

16. Ivi, pp. 256-257.

17. Ivi, pp. 262-270; the image in Figure 1 is from
<http://commons.wikimedia.org/wiki/Image:Lichtenstein_bitplanes.png>.

18. Ivi, p. 279.

19. The expression is from Jan Stewart, *Nature's Numbers*, mentioned in D. Salomon, *Data compression*, ed. 1998, p. 249.

20. I. Daubechies, The Wavelet Transform, Time-Frequency Localization and Signal Analysis, *IEEE Trans. on Inform. Theory*, Volume 36, No. 5, pp. 961-1005, September 1990.

21. S. Mallat, A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Volume 11, No. 7, pp. 674-693, July 1989.

22. Some free tools allow previewing JPEG 2000 files on operative system viewers: see <http://www.optimidata.com/english/Jpeg2000sw/jp2_shell_extension.html>.

23. Daniel Lee, who worked in the ISO's JPEG group, points out, "JPEG 2000 offers a very comprehensive set of features in a single format and obviates the need to maintain two sets of files for each image. I definitely recommend that cultural heritage institutions use JPEG 2000 as a single preservation and delivery format" instead of keeping master files in a TIFF format and supporting an expensive storage system. See Editor's Interview: JPEG 2000. Dr. Daniel Lee, ISO SC29/WG1 (JPEG), in *RLG DigiNews*, Volume 6, Number 6, December 2002. See <<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000070519:000006287816&reqid=152144#interview>>.

24. See (in Italian only) F. Lotti, La qualità delle immagini nei progetti di digitalizzazione, in "*Digitalia*", Dicembre 2006, pp. 22-37, <http://digitalia.sbn.it/upload/documenti/digitalia20062_LOTTI.pdf>.

25. D. Taylor-Pearce, *Graphic file formats test page: JPEG-2000*, November 2005: <<http://www.she-philosopher.com/home/technical/graphicsformats.html>>.

26. D. Nister and C. Christopoulos, "Lossless Region of Interest with Embedded Wavelet Image Coding," *Signal Processing*, Volume 78, Number. 1, pp. 1-17, 1999.

27. The idea was proposed by Manfred Thaller to the Delos Summer School on Digital Preservation in San Miniato, in June 2006.

28. Judith Rog, Compression and Digital Preservation: Do They Go Together?, in Archiving 2007. Final proceedings of the conference held May 21-24, 2007, Arlington, Va. Springfield, Va: The Society for Imaging Science and Technology. Abstract at: <<http://www.imaging.org/conferences/archiving2007/details.cfm?pass=20>>. Rog mentions a study, in Dutch only: R. Bourgonjen, M. Holtman, E. Fleurbaay, *Digitalisering ontrafeld. Technische aspecten van*

digitale reproductie van archiefstukken (Gemeentearchief Amsterdam, April 2006,): <http://stadsarchief.amsterdam.nl/stadsarchief/over_ons/projecten_en_jaarverslagen/digitalisering_ontrafeld_web.pdf>. See also the more recent report R. Gillesse, J. Rog, A. Verheusen, *Alternative file formats for storing master images of digitization projects*, National Library of the Netherlands, March 2008, <http://www.kb.nl/hrd/dd/dd_links_en_publicaties/links_en_publicaties_intro-en.html>, and, from the same authors, the paper Life beyond uncompressed TIFF: alternative file formats for the storage of master image files, in *Archiving 2008. Final proceedings of conference held June 24-27, 2008, Bern, Switzerland*: The Society for Imaging Science and Technology, pp. 41-46.

29. P. Amon, Error resilience for JPEG 2000, in *Proc. Picture Coding Symposium*, France, April 2003, pp. 243-246.

30. Very recently, tests on image file robustness has been made by Volker Heydegger at Cologne University, introducing errors in the file without excluding the header. Results are, of course, quite different, and more simple and uncompressed files result to be more robust: "complex file formats tend to get into trouble with keeping their data against bit errors" (V. Heydegger, *Analyzing the impact of file formats on data integrity*, p. 55, in *Archiving 2008. Final proceedings of conference held June 24-27, 2008, Bern, Switzerland*: The Society for Imaging Science and Technology, pp. 50-55). We will assume that file complexity is not a problem – *if you can manage it* – i.e. if is possible to balance the risk of concentrating crucial information in the header with tools to backup the header itself, as performed by the Fixit! utility that we discuss forward.

31. Every test can be replicated (original positions of corrupted bytes are stored in an Excel file), and all documentation is available on request.

32. Centro di Fotoriproduzione Legatoria e Restauro, <<http://www.cflr.beniculturali.it/Progetti/Jpeg2000.php>>.

33. This approach was first exposed in P. Buonora, Long lasting digital charters. Storage, formats, interoperability, *International Conference "Digital Diplomats"* 28 February - 3 March 2007, Munich. <<http://www.geschichte.lmu.de/ghw/DigDipl07>>. A more technical exposition is in F. Liberati, *A Study on JPEG 2000 Error Resiliency for Long Time Preservation of Digital Images*. <<http://www.cflr.beniculturali.it/Progetti/DigitalPreservation>>.

Copyright © 2008 Paolo Buonora and Franco Liberati

[Top](#) | [Contents](#)
[Search](#) | [Author Index](#) | [Title Index](#) | [Back Issues](#)
[Previous Article](#) | [Next Article](#)
[Home](#) | [E-mail the Editor](#)

[D-Lib Magazine Access Terms and Conditions](#)

doi:10.1045/july2008-buonora