PROCESSING AND ANALYSIS OF DIGITAL IMAGES: HOW TO ENSURE THE QUALITY OF DATA CAPTURED?

Juliano da Silva Ignacio[1], Sidnei Jose Buso[2], Waldemar Alfredo Monteiro[3]

[1]Materials Science and Technology Center (CCTM), IPEN - Nuclear and Energy Research Institute at USP – University of São Paulo, São Paulo, Brazil
juliano.s.ignacio@usp.br
[2]UNIP – Paulista University, Santos, Brazil
sjbuso@gmail.com
[3] Materials Science and Technology Center (CCTM), IPEN - Nuclear and Energy Research Institute at USP – University of São Paulo, São Paulo, Brazil
wamontei@ipen.br

ABSTRACT

It is a common activity for researchers in materials science, the constant use of scanned images generated by electron microscopes. While virtually all equipment that generate these images (micrographs) can use a file type most suitable for capturing image data generated (as TIFF or RAW files in case of metallography), many researchers choose to use a file format more common as JPEG, for example, perhaps the reason of the space available on portable storage devices (USB, CD or DVD) that owns, or by the lack of knowledge about the types of image files and their appropriate use. The problem with the use of certain types of image formats is mainly the loss of the original data captured by an electron microscope. As if that were not enough, the application of filters and processes in the original image must also be carefully crafted so as not to lose or change data captured or data relevant to the study. This article seeks to highlight the treatment of images in research and publications done by researchers with no knowledge of this matter, since the use of scanned images is only a resource to continue the progress of their own research. Furthermore, this article aims to promote a discussion on how to treat the problem of digital images published in scientific papers so that researches can really be replicated in full.

KEYWORDS

Image Processing, Digital Image Analysis, TIFF, JPEG, Metallography, Microscopy, Image Data

1. INTRODUCTION

The dissemination of scientific research around the world has no other aim than to share with all the discoveries and innovations achieved in several areas. However, there is at least one problem in the images required for the replication of a complete and reliable work has documented: the images published in most cases are not the original images used in the research.

An example of this problem can lighten their understanding.

Based on the study presented in [1], theoretically, following step by step all the necessary requirements to replicate the image processing to correct grain counting, the resulting image displayed in the study can be seen in Figure 1.

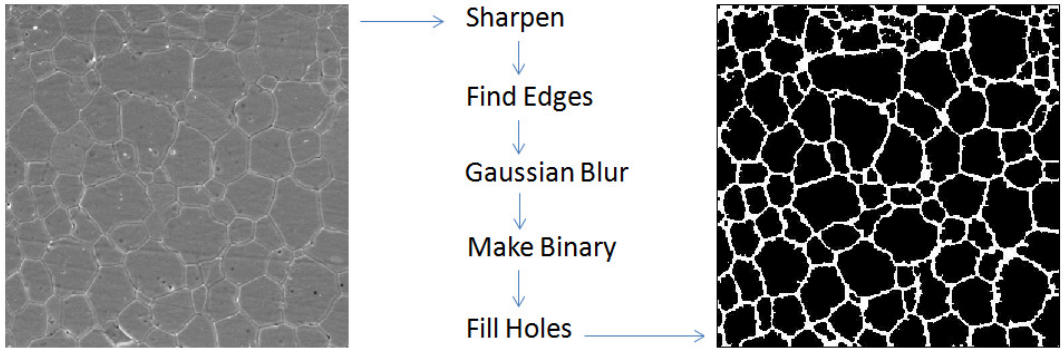

Figure 1 - Process binarization image of a microstructure of zirconia-based ceramics doped compound of yttrium and rare earths.

However, replicating this process to the images available in the published document, the result showed entirely different and undesirable, as shown in Figure 2.

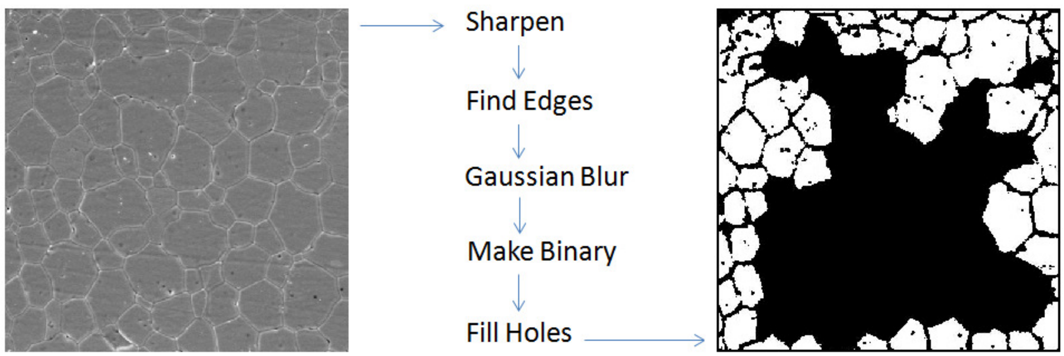

Figure 2 - Replication of binarization process shown in Figure 1 with images provided on the published document.

Despite some initiatives that seek to increase the resolution of images in scientific publications, even so, these images are not original images.

1.1. The growing use of processing and analysis of digital images

The digital imaging began in the early 20's of last century, when images were digitized and sent by a submarine cable between London and New York to be published in American newspapers. This initiative, reduced the transport time of images through the Atlantic Ocean from more than a week to less than three hours. However, only in 1964, during the early stages of space exploration that emerged automatic analysis of digital images, with methods to improve the visual information for human analysis and interpretation, rapidly and with greater accuracy. At this time, several other areas such as medicine, geography, archeology, physics, biology, etc., started the increasing use of processing and analysis of digital images (PADI) [2].

In the area of microscopy specifically, the creation of electron microscopes (optical, scanning, transmission and others), PADI has become critical since these microscopes capture sample information being analyzed and transform this information into a digital image file. Moreover, the technology for this had been embedded in the equipment and not publicly disseminated in concepts, methods and practices for the right treatment of digital images [3].

1.2. Editing digital images: a resource available to all

With the widespread use of computers and access to image editing software for any users (like Photoshop for example [4]-[5]), more and more people begin to work with digital images, including researchers from various areas that today are used to use PADI to accelerate their research processes. But, make use of this resource for specific purposes also requires something more than just notions of use some equipment and software, requires minimal knowledge about image types and files, when and where to use them.

2. DIGITAL IMAGES

2.1. Types of Images

There are two types of images: vector and bitmap. The vector image (by mathematical vectors) allows resizing image at any scale without loss of quality, however, at the moment that must be printed for example, it is converted in a bitmap suitable to the characteristics of the equipment. The process of conversion from vector to bitmap is called rasterization. The bitmap image is - as the name suggests - mapped bits, where the image is organized into a series of rows and columns formed by pixels (pixels' matrix), and each pixel (picture element) has only one color (where each color is referenced by numeric intervals of 2, 4, 8, 16, 24 or 32 bits), not having a fixed dimension since the size will be assigned to the pixel at the display time, in video or printer.

So the bitmap is the final type of the entire image displayed. Therefore, it is necessary to understand the meaning of resolution. Resolution in this case is the ability of a capture or reproduction images system has to reproduce details. The resolution is given in dots per square inch (dpi), so, how higher the resolution is, better the details reproduced by a given system (considering the "system" as all parties involved to capture image: CCD, lenses, software, etc.). It is important to make clear that the resolution of the equipment must be greater than the resolution of the bitmap file desired.

The development of equipment and software in the area of capturing and processing images made with various types of files that were created to support increasing demands of everyday life. Today it is common to anyone taking pictures or filming with a mobile phone or digital camera. Although it does not seem like something so complex, but it was, because this need was the restart of a new image revolution with regard to these files types, to ensure highest capture rate, with more fidelity of detail, with the lowest possible storage space. It was in this scenario that emerged the type of bitmap file more used nowadays, JPEG [6]. Below, there are some image file types:

- *JPEG (Joint Photographic Experts Group)*: The JPEG file type is a file type that allows the use of various compression techniques, whereas most of them, present losses. In this case the original picture will not be exactly identical to the recovered image after decompression. Often such losses are not perceived by the naked eye, making this type of file a file suitable for capturing images to be viewed in this way, for example in scientific articles like this.
- *TIFF (Tag Image File Format)*: The TIFF file type is the file type most suitable for capturing images. It is one of the file formats that digital supports more colors allowing for transparency and opacity data, and the use of layers. Although some types of possible compression does not lose any detail, but its size is pretty big.
- *PNG (Portable Network Graphics)*: The PNG file type emerged as an evolution of the GIF file type, supporting millions of colors, transparency and animation. This file type can be used in place of JPEG in some situations because, despite having compression, this does not generate losses.
- *DICOM (Digital Imaging and Communications in Medicine)*: The medical area, for example, has created a specific file type called DICOM [7]. Initially created with the purpose of standardizing the formatting of diagnostic images such as CT scans, magnetic resonances, x-rays, etc, this file type is transformed into a set of standards for treatment, storage and transmission of medical information in an electronic format, allowing medical images and associated information to be exchanged between diagnostic equipment imagers, computers and hospitals safely, quickly and with the highest quality.

Of course there are many other types of image files (like GIF, BMP, DWG, SVG, TGA, EPS, IMG, PCX, PSD, CDR, VML, among others), however, the first three presented above are sufficient to demonstrate the risks involved in the use of inappropriate images for PADI.

3. Risks to the Quality of Captured Data

3.1. Converting the file type TIFF to file type JPEG

Figure 3 shows a generic image of a fish. The original image is a file type TIFF with a resolution of 300 dpi, 24-bit color and total size of 7.17 MB. This file is too large to compose a document like this article, then, this TIFF file is converted to a file type JPEG with the same resolution of 300 dpi, same 24-bit color, and the file size resulting in almost a thousand times smaller than the original, only 856KB. The drastic reduction of the size of file confirms that many original image information were discarded and lost. Such information has great chance to be critical for automated PADI, despite the impression that gets nothing changed.

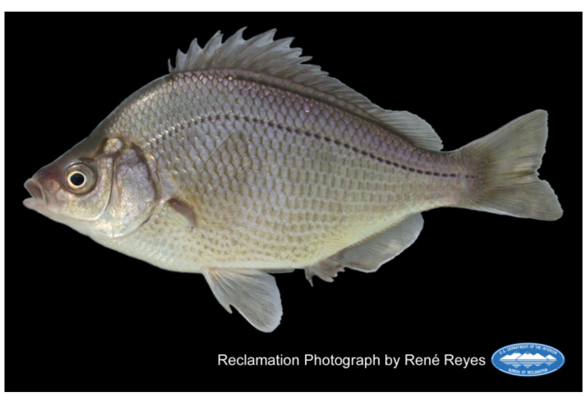

Figure 3 – Generic image of a fish (because the lack of images related to the area of materials science with public links to the original images).

Download higher resolution TIFF file (7.2 Mb) at
http://www.usbr.gov/pmts/tech_services/tracy_research/photos/fish/tule_perch_(unbarred).tif

Download higher resolution JPEG file (857 Kb) at
http://www.usbr.gov/pmts/tech_services/tracy_research/photos/fish/tule_perch_(unbarred).jpg

3.2. Image files with compression and loss

The most aggravating factor for the PADI in an automated way is the loss of original image data, because if the file information has been changed, any processes running under this new information must submit results of processing different from the original file, which would not make sense for a particular study.

By the way, there is still one more aggravating when it comes specifically to the JPEG file type (the file type most used nowadays). Each time the file is saved, the calculation of image compression is performed again. If you have edited any pixel in the image, the calculation of its compression will significantly alter the pixels around it (see Figures 4, 5 and 6). If these pixels were originally a cell, a grain or a precipitate, which will have its size measured in an automated way, the sizes resulting would be different, which could compromise a study requiring accurate and lower margins of possible error, as in the areas of precision mechanics and nuclear.

Figure 4 – JPEG file type original generic image detail. Here are the 3 (three) highlighted pixels to be edited: one in the left upper corner, two in the right upper side of image where shows a space (which could be a grain or a cell) limited by a range of pixels value.

Figure 5 – JPEG file type generic image detail after edited only 3 (three) pixels and saved 20 (twenty) times. It is clear that the areas around the edited pixels have changed because of file compression calculation. Therefore, the original image data has changed, which can result negatively in the application of PADI.

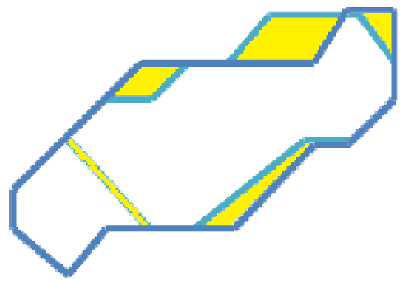

Figure 6 - Overlapping areas observed in figures 2 and 3. In the filled areas are presented the differences between them.

3.3. "Beautification"

The desire to edit an image is too big when there seems to be perfect or it does not show what the researcher expected. Therefore, once the image is edited, there is a great risk of this change have a negative impact on the final results of study being done [8], after all, "beautification" is a form of misrepresentation of the original image [9].

4. GUIDELINES FOR BEST PRACTICES IN IMAGE PROCESSING

The journal Nature published a warning note in 2006 questioning the quality of the images that some authors would post them in your articles [9] and yet, within the policies of articles published in Nature describes the importance of access to the original images of any study that uses PADI [10]. The increasing use of PADI led to universities and research centers around the world publish some basic and relevant information for the use of this type of resource [11]-[12].

Here are some topics that should be used like guidelines for best practices in image processing [11]:

- *Treating Images as Data:* An image can be translated as a multidimensional array of values. Assuming that a bitmap image is as a real picture with two-dimensional and, the three-dimensional effects are achieved through painting techniques, so each pixel has its place in the matrix (such as a spreadsheet) and, in each location is stored one or more values (a vector or sub-matrix) that defines the color or brightness, contrast and saturation of the pixel. Since all these figures, the manipulation of an image is done by mathematical functions which alter these values within this matrix sometimes uniformly and sometimes (with great care) not. Moreover, precisely because the human eye does not detect all variations of colors, working with values ensure that there will be greater control in image manipulation.
- *Save the Original:* It is very important (and safe) that manipulated images can be compared with the original image (the first captured image) to make sure that important or relevant data to research does have not been lost or, which have not been entered new data (artifacts) in the processed images. Therefore, the original image must always be maintained.
- *Making Simple Adjustments:* As photographic images, simple adjustments are generally acceptable to help visualize the information that is present in the image, such as brightness, contrast and gamma (in moderation). However, be careful with commercial tools that make automatic adjustments often aggressive to scientific images. One should always use a software tool that documents the changes made to the image.

- *Cropping is usually OK:* Crop a picture to publish a figure is usually considered acceptable since the interest is in showing a specific area of the image. A cut can be used to make the image smaller in size in pixels cutting the empty space around its edges or removing pieces of the picture that are not in the universe of research interest. Scientists must be very careful when cropping an image, not to transform the motivation to improve the image composition in a way to hide something that is not in accordance with the hypothesis adopted.
- *Comparing Images:* A precise comparison between the treated and control images will be reliable only if both are acquired under identical conditions. In this case you need to be especially careful about the configuration of image acquisition devices (electron microscopes, for example), because these settings are often closely linked to the interpretation of the image data. Latest equipment store metadata about the machine settings for each image.
- *Manipulating the entire Image:* The manipulations performed in a specific area of an image and not performed in other are questionable. The use of such filters can selectively enhance specific areas of an image. The images need to be improved uniformly and, the adopted procedures need to be formally declared.
- *Filters Degrade Data:* The commercial software tools (well known) for manipulating images were created mainly for printing and design, and not for scientific research. The filters are mathematical functions called kernels of convolution. These mathematical functions alter the magnitude of a variable quantity of pixels around the pixel of the convolution kernel. Although the mathematical function is uniform, its effect is different in different parts of the image, at the risk of creating new artifacts in the image that could be interpreted as meaningful data.
- *Cloning Degrades Data:* The case for clone (copy) parts of the image is usually to cover imperfections in it. While this feature is often not visible to the naked eye, it can be detected with image processing. The best solution in this case is to capture another image from the sample (or if a new sample prepared under the same conditions). The use of any feature image retouching through selective imaging is unsuitable for scientific images. The use of cloning techniques to specifically create or copy objects in an image that did not previously exist may lead to allegations of research misconduct (falsification or fabrication of images).
- *Making Intensity Measurements:* They should be performed on the image data processed uniformly (preferably raw data to avoid potential artifacts) and calibrated to a known standard. These measurements are very difficult to perform properly with rigorous scientific manner, because many variables must be considered and controlled before data can be considered significant (for example, things like the uniformity of the sample preparation techniques, types of microscopes, aberrations, physical limitations, electronic limitations, electronic noise, etc.).
- *Lossy Compression Degrades Data:* There are two types of compression of image files, one with and another without loss. Avoid the use of lossy data compression. For example, the TIFF image file uses a compression that does not alter the integrity of the data. The problem is that this still does not result in compression sizes comparable to the other type, and the TIFF file type is not a universal standard indeed. Already a JPEG image file, a universal standard for image storage, uses in its compression an algorithm of the discrete cosine transform, which changes the X and Y spatial resolution and image intensity value of pixels, throwing away all the image information that are not noticeable to the human eye. According to Dr. John Russ, *"The reason for recording images in scientific studies is not to keep remembrances of familiar objects and scenes, but to record the unfamiliar. If*

it is not possible to know beforehand what details may turn out to be important, it is not wise to discard them. And if measurement of features is contemplated (to measure size, shape, position or color information), then lossy compression, which alters all of those values, must be avoided". Many general-purpose software such as Microsoft Office and Adobe Acrobat automatically uses lossy compression.

- *Issues with Magnification:* Digital images of real world objects are samples of the object where each pixel of the image has a scale. The scales are generally the same for both X and Y dimensions, however, sometimes the image has the Z dimension, which may result in misinterpretation if not considered. The magnification of an image is determined by the difference between the original scale of the pixel and its final scale. It is suggested that the expansion does not exceed two to three times magnification since this assists in achieving higher contrast (essential for the correct resolution of structures in microscopy). Higher magnifications can produce greater accuracy in measurement of some characteristics of the image, however there is a large loss of contrast making it difficult to determine the edges of an object and the risk of creating new artifacts.
- *Issues with Pixels:* Use caution when changing the size (in pixels) of a digital image. Decreasing the image size in pixels, reduces the spatial resolution of their dimensions by mathematical interpolation (in most cases), creating a new resolution of the image dimensions and the intensity values of each pixel. Increasing the size of the image resolution does not increase its size, however, also using mathematical interpolation, creates more jagged edges less distinct. It is recommended to change the image size only once, it prevents the appearance of artifacts by successive image resizing.

5. Conclusion

Although there is already a real concern about the minimal care needed regarding image manipulation oriented to PADI - processing and analysis of digital images - very little has been done so that basic information (and practices) comes to the researcher who only uses PADI as a resource to power your own research. The best practices presented are the initial path: replace where possible images of type JPEG by PNG, be aware of what makes a filter before applying it to the image, keeping the original image safely. As publishers of scientific publications, could follow the example of Nature publisher, demanding that original images are available in public repositories, and his address published with the article. This attitude has a greater purpose than just publication of quality data, is to continue the scientific research because, in possession of the original data, researchers worldwide will have a greater chance of replicating the study published and give continuity to it. Today, in the PADI area, many studies suffer from a lack of access to original data having to redo all the steps of referenced research, resulting - in most cases - in rework, consumption of time and resources, change of focus, and other factors that hamper a natural evolution of scientific works. We need to take actions that can promote change this scenario.

Acknowledgement

The authors thank all researchers and professionals who contributed their experiences and opinions regarding the use and necessity of processing the digital image analysis.

References

[01] F. C. Dias, “Using ImageJ software for quantitative image analysis of microstructures of materials,” M.S. thesis, Science and Technology of Materials and Sensors, INPE, São José dos Campos, SP, 2008.

[02] R. C. Gonzales and R. E. Woods, Digital Image Processing. São Paulo, Brazil: Edgar Blücher, 2000.

[03] Handbook of Microscopy - Applications in Materials Science, Solid-state Physics and Chemistry – Method I, VCH, Weinheim, DE, 1997.
[04] Using Adobe Photoshop CS5 for Windows and Mac, Adobe, San Jose, CA, 2010.
[05] D. W. Cromey. (2012, Aug.). Potentially the most dangerous dialog box in Adobe Photoshop, University of Arizona, Tucson. [Online]. Avaliable: http://www.ecb.epm.br/~ramortara/confocal/Aulas/Photoshop_Image_Size_dialog_box.pdf
[06] O. Marques Filho and H. Vieira Neto, Digital Image Processing, Rio de Janeiro, Brazil: Brasport, 1999.
[07] DICOM - Digital Imaging and Communications in Medicine, ISO 12052, 1993.
[08] M. Rossner and K. M. Yamada, "What's in a picture? The temptation of image manipulation," The Journal of Cell Biology, vol. 166, no. 1, pp. 11-15, July 2004.
[09] "Not picture-perfect". (2012, Jun.). Journal Nature, no. 439, pp. 891-892, Feb. 2006. [Online]. Avaliable: http://www.nature.com/nature/journal/v439/n7079/full/439891b.html
[10] Guide to Publication Policies of the Nature Journals. (2012, Jun.). Nature, London, GB, 2012. [Online]. Avaliable: http://www.nature.com/authors/gta.pdf
[11] S.H. Vollmer. (2012, Aug.) Online Learning Tool for Research Integrity and Image Processing, University of Alabama, Birmingham. [Online]. Avaliable: http://www.ori.dhhs.gov/education/products/RIandImages/guidelines/list.html
[12] D. Cromey. (2013, Apr.). Microscopy and Imaging Resources, University of Arizona, Tucson. [Online]. Avaliable: http://swehsc.pharmacy.arizona.edu/micro

Authors

Juliano da Silva Ignacio: Graduated in Information Systems at Sant'Anna Academic Center (2004), undergraduate in Software Engineering at Technological Institute of Aeronautics (ITA, 2007), M.S. student in Materials Science and Technology Center at Nuclear and Energy Research Institute – University of São Paulo (IPEN-USP, 2013). Has experience in Software development, non-classical Logics (fuzzy and paraconsistent), Database Architecture and Data Analysis. Professor for graduate education. IT professional working with big data (SAP HANA).

Sidnei José Buso: Graduated in Physics (bachelor) at Catholic University of São Paulo (1994), M.S. (1999), PhD (2004) and postdoc (2010) in Nuclear Engineering at Institute of Nuclear and Energy Research - University of São Paulo. Professor for undergraduate education. Has experience in the area of Materials and Metallurgical Engineering, with an emphasis on Physical Properties of Metals and Alloys, acting on the following topics: physical metallurgy, automation and control, electron microscopy, aluminum alloys and nickel alloys.

Waldemar Alfredo Monteiro: PhD researcher in Materials Science and Technology Center at IPEN – Nuclear and Energy Research Institute – located inside the USP – University of São Paulo. Lecturer on undergraduate and graduate course at Presbyterian Mackenzie University. Have published more than 100 scientific papers, chapter and books in Material Sciences area.

INTENTIONAL BLANK