

Article

Standards for Digitization in Cases of Maps, Documents, and other Relics in the Service of Cultural Heritage

George Malaperdas

Laboratory of Archaeometry, Department of History, Archaeology and Cultural Resources Management,
University of The Peloponnese, Kalamata, Greece
Correspondence: envcart@yahoo.gr

Abstract: This paper discusses the analysis of correct digitization practices to follow for maximum performance of the technique. Although it is written for cases that fall within the broader context of culture and cultural heritage, it is ultimately about writing rules that are not limited to the above-mentioned cases, but can be used in more general situations, particularly printed materials. This paper will therefore discuss the technical characteristics of the choice of digital imaging devices and distinguish the types of quality calculation in the different cases of digitized text, digitized manuscript, digitized maps, and photographs.

Keywords: Digitization, Maps and documents, Cultural Heritage, Mapping, GIS

1. Introduction

In the process of digitization, especially when it comes to cultural relics in printed and written form- such as old maps and diagrams, old handwritten documents, rare notes, and photographs- some important specifications must be met in order to achieve the desired result. All the above-mentioned cases involve objects based on any type of illustration or writing on paper surfaces; while after digitization they can change their format and pass into cases of three-dimensional technologies, the whole digitization process is based on so-called two-dimensional digitization.

Two-dimensional digitization is the most standardized and best-documented method of digitization worldwide. It is divided into two main methods: scanning and digital photography. The equipment for these two methods is different; while the first form is mainly done through the choice of scanners, the second is done through photography and photographic techniques.

1.1. Main categories of deterioration

Discussing in general the objects that are part of the broader term of cultural heritage, they are usually objects that have stood the test of time. However, it would be extremely unlikely that we would encounter such objects without having suffered on a smaller or larger scale due to the passage of time. These damages can be summarized in three main categories:

- I) Photochemical damage: due to the aging of the construction material and the effect of environmental parameters such as relative humidity, temperature, and radiation.
- II) Mechanical damage: due to poor handling, storage, and transport practices. They include significant damage such as tearing, bulging, loss of entire parts and loss of paper.
- III) Deposits: due to mishandling and human carelessness, incorrect operations, and inappropriate storage conditions [1].

How to cite this paper: Malaperdas, G. (2021). Standards for Digitization in Cases of Maps, Documents, and other Relics in the Service of Cultural Heritage. *Open Journal of Educational Research*, 1(1), 24–31. Retrieved from <https://www.scipublications.com/journal/index.php/ojer/article/view/99>

Received: August 7, 2021

Accepted: September 22, 2021

Published: September 23, 2021



Copyright:© 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

2. Materials and Methods

The choice of method to be used must account for a number of factors that concern the technical characteristics and affect the quality of the final result. The most important factors are the resolution, the color depth, the dynamic field, and the signal to noise ratio (SNR).

• **Resolution:** The resolution is directly related to the density of information that can be captured by the scanner. This is expressed in dots per inch (DPI) or pixels per inch (PPI). It is obvious that the higher the number of DPIs or PPIs respectively, the more information will be printed. There are three main factors in choosing the right resolution of an object. These are:

A) Its dimensions. In a case where we want to digitize an object- such as a photo- of a certain scale, which as an original maximum resolution to be printed on A5 paper; when we want to enlarge the exact same photo, to print it on A4 paper, then we will have less information than in the original and therefore less analysis. On the contrary, in the initial dimensions of A5, more resolution will be needed, as more imprinted detail will be included.

B) The detail it has. For example, a document requires much less resolution than a photo on a page of the same size, as the photo has more detail.

C) The purpose for which the digitization will take place. When the purpose of a job is, for example, to capture a map and digitize it in every detail in a design software; this is completely different from needing the map to simply display it on the internet. In such a case, obviously, the scan can be done in low resolution; as, on the one hand, we are not interested in every detail and, on the other hand, we also need to reduce the file size.

• **Color depth:** The color depth is the direct function of color tones that can be captured. The tone for each pixel is represented in bits. For example, a 1-bit image is a black-and-white image, where each pixel corresponds to a bit that can be either black or white. In contrast, an 8-bit image has 256 shades ($2^8 = 256$) and a 24-bit image has millions of shades ($2^{24} = 16,777,216$).

Figure 1 shows the same image imprinted at different color depths. Capture at a greater color depth helps reduce noise and extends the range of image shades without loss of information; it is worth mentioning that the scanners on the market have a maximum color depth of 48-bit. The main rules regarding the color depth are the following:

1) For black and white documents, as in the case of photocopies, black and white digital capture is recommended due to their high contrast.

2) As for black and white originals that have images with the mapping, it is preferable to be in grayscale.

3) Concerning very valuable or very old prototypes (such as manuscripts, old publications, musical scores, maps and diagrams), it is preferable to be captured in grayscale, or even colorized, so that all the details that may be related to the shade can be distinguished and the condition of the paper, including the marks that may be on them, can be determined.



Figure 1. Color Depth 1-bit, 8-bit, 24-bit (from left to right)

• **Dynamic field:** The dynamic field measuring range between the brightest and the darkest point of an image. What triggers the dynamic field of a scanner or even a digital camera is how it affects the ability to capture shadows and highlights in an image.

As a rule, scanners with greater color depth have the ability to capture higher optical densities. The dynamic field arises as a function of the optical density logarithm. While the value pair (0,0) is the absolute white, respectively, the absolute black is observed in the value pair (4,0). If we were to specify all this, we would say that the dynamic field is the range of optical density values, between 0 and 4, that a scanner can distinguish. Flatbed scanners have a dynamic range of 2.5 to 3.5; while more expensive scanners, which contain a drum, reach up to 3.8. The latest technology scanners also have a special value for the darkest spot that can capture called dMax. The higher the dMax value a scanner has, the better it captures shadows. This is especially important for all digitization cases involving maps, slides, negatives of an old camera, and, in general, where there is a lot of detail that we want to render as it is. The following table (Table 1) shows the values for the dynamic field of different scan product sources.

Table 1. Dynamic field of different scan product sources

| Categories | Dynamic Field Values |
|----------------------------|----------------------|
| Newspaper | 0.9 |
| Other printed material | 1.5 |
| Photography (normal print) | 1.6-2.0 |
| Photography (high quality) | 2.0-2.3 |
| Negatives from old films | 2.4-2.8 |
| Slides | 2.9-3.2 |
| Maps and Diagrams | 3.3-3.8 |

• **The ratio of the signal to the noise:** Any unwanted component within the image signal is considered noise. It is caused by failures in the design of the recording device and, in general, you consider that the presence of noise is inevitable in all electronic devices. However, the size of the noise is variable and always depends on the quality characteristics of each electronic device.

In the case of digital images, the presence of noise appears as small dots at different shading points and in instances where the signal is low.

In order to make this visible, there should be bold printing in areas that are brighter in color or in areas where there is an increasing range of image contrast.

Noise is measured based on the ratio of the signal to the noise (SNR). A typical ratio is 60 dB for a 24-bit color depth and at least 75 dB for a 36-bit color depth. As a general principle, the higher the ratio, the better the quality of the digital image; this is very important in terms of digital processing, especially of color maps and diagrams.

3. Analysis

3.1. Limitations of Analysis

In the process of scanning any product that comes from it, it makes sense that as the resolution and essentially the quality of the document increases, so does the volume occupied by the digital file. Nowadays, despite the fact that the storage space of computers has increased significantly, a large amount of storage space is occupied by this type of archival material, especially when we refer to digital scans of old and historical maps and diagrams that may need a lot of detail. There is also a limit to the speed of the design software we will use, as in all software the larger the file we want to open, the more difficult it is to both import and edit (due to the fact that it loads later). Finally, there is the

limitation of over-recording information. Increasing the resolution often does not have the desired effect from the user, as it does not add any new information from the already known ones that can be attributed to a lower resolution scan. In fact, in many cases, the result obtained due to the excessive recording of information is the exact opposite, as often there is the recording of details due to noise which are not desirable (small dots, line melanoma, etc.). A typical example for this is all the scanning effort of old postcard-postcards. So, with these, the paper is of very poor quality and if someone tried to scan them in very high resolution, the result would be the imprint outside the image and texture of the paper resulting in the alteration of the desired scan result. In general, there is a point of equilibrium, in which there is complete harmonization between the resolution and the color depth of the digital capture, with the information of the original. In order for the result from the digital copy of the scanning to be ideal, this equilibrium point must be found as the extra resolution has nothing more to offer. The general recommendation is that the digital capture should be done in the maximum possible resolution that is allowed, by both the cost and the available resources, and that it is considered satisfactory for the specific object to be scanned. In this way, it is possible to extract from the digital copy a file that may have a lower resolution and therefore be much smaller in size. On the other hand, in no case will the opposite be true, i.e. to have a low quality image and export it to a higher quality image.

3.2. Methods of quality calculation (Quality Index - QI).

The great majority of large format document scanning is done at resolutions of 200 to 500 dpi (dots per inch). While higher resolutions produce better images, they also increase file size, often significantly; the file size would increase by 125 percent when the resolution is increased from 200 to 300 dpi, from 40,000 to 90,000 pixels per square inch. A grayscale scan requires more patience (Figure 2).

The scanner head must scan the same image for three different colors- namely red, green, and blue- when scanning a color image. In early color scanners, this was accomplished by scanning the same area three times for the three separate colors. Three-pass scanners are the name for this type of scanner.

Moreover, most color scanners now scan in one pass, using color filters to scan all three colors in one pass. In theory, a color CCD works in the same way as a monochrome CCD. Each color, on the other hand, is made by blending red, green, and blue. Each pixel in a 24-bit RGB CCD, for example, has 24 bits of information. A scanner that uses these three colors (in all 24 RGB modes) may output up to 16.8 million different colors.

Various researchers have developed, overtime, the issue of quality calculation for the digitization of mainly printed material. Here, we present the calculation of quality (Quality Index - QI) developed by the University of Cornell.

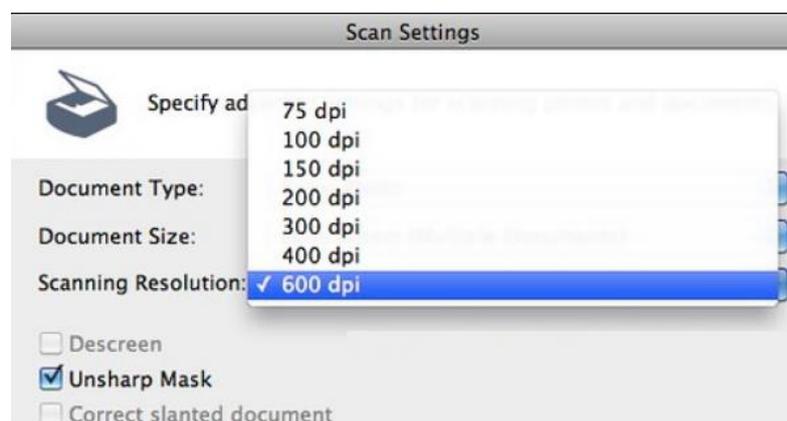


Figure 2. Scan settings

A) Texts

In the case of digitized texts, there is an initial clear distinction between black and white scanning and scanning in color or grayscale (Figure 3). It is now commonly accepted that scanning in black and white shows information loss (even infinitesimal).



Figure 3. Scanned text comparison 300 dpi (left) vs. 600 dpi (right)

There are two main factors that affect the quality of the text scanning process. These are the function of the height of characters (h in mm), and the resolution (in dpi). In small text (height 1 mm), the scan must be done at a resolution of 400 dpi in grayscale, so that the resulting digital copy is of excellent quality. On the contrary, for a text in which the characters have a height of 2 mm, it is enough to scan at half of the 200 dpi resolution, in grayscale (Table 2).

Table 2. Types of quality calculation (for digitized text)

| Black and white text | Text in shades of gray or color |
|--|--|
| $QI = (dpi \cdot 0.039 \cdot h) / 3$ $h = (3 \cdot QI) / (0.039 \cdot dpi)$ $dpi = (3 \cdot QI) / (0.039 \cdot h)$ | $QI = (dpi \cdot 0.039 \cdot h) / 2$ $h = (2 \cdot QI) / (0.039 \cdot dpi)$ $dpi = (2 \cdot QI) / (0.039 \cdot h)$ |
| Levels of Image Quality (8.0) Excellent (5.0) Good (3.6) Poor (3.0) Just Readable | |

B) Prototypes including graphics

If we have prototypes that include graphic maps, sketches, engravings, and even manuscripts, there are other types that take into consideration the width of the thinnest line or point of the drawing (w in mm). In the case of the map, where the thinnest line thickness is 0.2 mm, the resolution that should be scanned to have the desired result is at least 400 dpi in shades of gray or color.

Table 3. Types of quality calculation (for prototypes with graphics)

| Black and white text | Text in shades of gray or color |
|---|--|
| $dpi = QI / 0.039 \cdot w$ | $dpi = (1.5 \cdot QI) / (0.039 \cdot w)$ |
| Levels of Image Quality (2.0) Excellent (1.5) Good (1.0) In dispute (<1) Insufficient – not acceptable | |

C) Photographic material

With regard to cases of photographs and photographic material, it should be emphasized that there is no specific type of calculation of its scan of the two previous cases. This is mainly because the measure of detail is subjective to each photograph and there is no single measure that could determine the smallest unit of detail above the line and point on a map. The desired resolution is therefore determined by the dimensions of the original. Halftones images¹ require very high resolutions as the way they were printed in the past was done through repetitive patterns of dots and lines. Thus, even when scanning, it is easier to deform due to the phenomena of the wavy lines (moiré effect).

The general rule for such cases is to scan in shades of gray and with a resolution four times the scale of the sinusoidal image. In cases of aerial photo scanning, the minimum resolution requirements are the same as in art, the minimum, ie 800 dpi, while for the rest of the images; the minimum resolution is set at 400 dpi (Figure 4).



Figure 4. Scanning a halftone image in 150 dpi (left) and 400 dpi (right).

3.3. Scanning on digital cameras

The second scanning mode is made with the help of the cameras. The resolution in this case is measured in Megapixels, which result from the number of pixels that the digital camera can print. The following (Table 4) is a table of minimum resolution and color depth requirements for scanning with a digital camera. What will determine the final specifications for the best result depends mainly on the nature of the original object, the objectives of the work, the degree of specialization of the users involved, and, finally, the budget of the work.

¹Halftones are methods of graphic reproduction processes using dots (discrete locations) of different sizes or even distances from each other to simulate the smooth tonal continuity of a natural image. This technique is also known as dithering. Non-dot image printing techniques are called contone prints, e.g. laser printing on photochemical papers or dye sublimation printing.

Table 4. Minimum scanning requirements through digital photography

| Original Item | Minimum Resolution | Color Depth |
|---|-------------------------|-------------|
| Printed material | 3264 * 2448 (8Mpixel) | 24-bit |
| Photos | 4064 * 2 704 (11Mpixel) | 24-bit |
| Maps (always depends on their scale) | 4064 * 2 704 (11Mpixel) | 24-bit |
| Works of art and fabrics | 4064 * 2 704 (11Mpixel) | 24-bit |

3.4. The accuracy of scanned images in GIS

Scanned images have now become the main source of input for GIS; the increased use of scanners in the GIS environment has prompted us to consider the scanners' limitations in terms of scanned image accuracy. Since most GIS software has very strict accuracy criteria, this accuracy of the input data must be quantified before the user uses it. In practice, input data must be accurate to 0.4572 mm in order to be used in a GIS database. This means that, at the scale of the map, an input data position must be within 0.4572 mm of its actual geographic location. As a result, a scanner cannot generate more positional accuracy error than the GIS's maximum error limit. Standard accuracy problems, including media continuity, source accessibility, and gaps in data collection methods which can be easily quantified; as such, the user can determine if the resulting data is suitable for their GIS before integrating it. Presently, with the sudden wave of scanned data, there is a new problem to address: the input scanner's accuracy. The ability of a scanner to generate an image with output dimensions that are proportional to the input document is known as accuracy. Since scanners are still relatively costly, scanning large volumes of data that do not meet the GIS's accuracy criteria can be disastrous. Within the defined tolerances, the scanned image can be dimensionally accurate, but nothing can be said about the data within the image's body. Even if the scanner is working within specified accuracy requirements, features inside the image can be as far as 7-10 mm from their correct position at the scale of the map, even if the image has exactly the same amount of pixels. Depending on the size of the source map, 7-10 mm will equate to hundreds of meters of error on the field. This is unacceptably bad for any GIS. As a result, it is important to know how accurate the scanned image is so that corrective steps can be easily integrated into the study.

Although scanning and table digitizing can handle the majority of conversion needs- from textual data to graphics and even image data and video images- special techniques for entering material from other sources have been created. This includes everything from basic programs that make entering survey coordinates on the keyboard easier to technologies that reconcile aerial photographs with geographic data. Additional possible input sources include photogrammetric, remotely sensed, and CAD-generated data [2-5].

4. Conclusions

Digitization and the whole process allowed to a wide audience to access both objects and information they probably would not even know existed. The use of the internet nowadays helps in this direction where even the so-called non-experts have access to such cases.

Digitization is synonymous with extroversion as the most important thing it provides is that through digitization, the object or information is available to such a wide range of audiences, while the use of authentic, original heirlooms is protected "by many" [6].

That is the reason of why digitization is one of the most contradictory methods regarding extroversion; as it is at the same time, an extroverted process that gives knowledge and information to such a large audience, but also with such great introver-

sion as the user who will digitize the object is often not even part of a working group but is an individual user who as a modern Apostle aims to transmit this information around the world.

Digital information depends mainly on machines for decoding and reproducing data on digital screens. There are two critical factors that favor this process: appropriate equipment and human intervention [7].

Nowadays, due to the increasing speed of the internet, but also due to the many free (unpaid) software and websites that exist in the vast internet universe, there are websites that can automatically calculate the best needs for the best quality for both printing, as well as for scanning documents, texts, images, and maps [8].

However, before using our data, particularly with regard to cartographic approaches, we should be certain of its source (knowing the scan quality and estimating any errors) or we have performed the scan ourselves according to the requirements of each task assigned to us.

Acknowledgments: This project was implemented within the scope of the “Exceptional Laboratory Practices in Cultural Heritage: Upgrading Infrastructure and Extending Research Perspectives of the Laboratory of Archaeometry”, a co-financed by Greece and the European Union project under the auspices of the program “Competitiveness, Entrepreneurship and Innovation” NSRF 2014-2020.

References

- [1] Kyriazi, E., Tataris, G. and Soulakellis, N. "Preservation and Digitization of Maps of the early 20th century of the city of Mytilene", 11th National Conference of Cartography, The Cartography of the Greek State, 2010.
- [2] Geospatialworld.net. Available online: <https://www.geospatialworld.net/article/gis-and-scanning-technology/> (accessed on 7/8/2021).
- [3] Chiang, Yao-Yi &Leyk, Stefan &Knoblock, Craig. "A Survey of Digital Map Processing Techniques" ACM Computing Surveys 2014, Vol.47 (1), pp. 1-44
- [4] Malaperdas, G; Zacharias, N. "A Geospatial Analysis of Mycenaean Habitation Sites Using a Geocumulative versus Habitation Approach" Journal of Geoscience and Environment Protection, 2018; 6, pp. 111-131. <https://doi.org/10.4236/gep.2018.61008>
- [5] Anagnwstou, E. "Study and Methodology of Good Development Practices for Digitization of Culture Content" University of Patras, 2005.
- [6] Malaperdas, G. "Digitization in Archival Material Conservation Processes", European Journal of Engineering and Technology Research, 2021; 6(4), pp. 30-32. doi: 10.24018/ejers.2021.6.4.2444
- [7] Smith, A. (1999) "Why Digitize?" Council on library and information resources, commission on preservation and access.
- [8] Fulton, W. "A Few Scanning Tips" Available online: <https://www.scantips.com/calc.htm> (accessed on 7/8/2021).