

UNIVERSIDADE FEDERAL DE PERNAMBUCO CENTRO DE INFORMÁTICA PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

DANIELA RAPOSO NUNES DE MELLO

Which is the most suitable scanner resolution for documents?

Recife

DANIELA RAPOSO NUNES DE MELLO

Which is the most suitable scanner resolution for documents?

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Processamento e Síntese de Mídias e Interação

Orientador: Prof. Dr. Rafael Dueire Lins

Coorientador: Prof. Dr. Raimundo Oliveira

Recife

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Mello, Daniela Raposo Nunes de.

Which is the most suitable scanner resolution for documents? / Daniela Raposo Nunes de Mello. - Recife, 2025. 80f.: il.

Dissertação (Mestrado) - Universidade Federal de Pernambuco, Centro de Informática, Programa de Pós-Graduação em Ciência da Computação, 2025.

Orientação: Rafael Dueire Lins.

Coorientação: Raimundo Corrêa de Oliveira.

Inclui referências e apêndices.

1. Image resolution; 2. Document digitization; 3. Nyquist-Shannon Theorem. I. Lins, Rafael Dueire. II. Oliveira, Raimundo Corrêa de. III. Título.

UFPE-Biblioteca Central

Daniela Raposo Nunes de Mello

"Which is the most suitable scanner resolution for documents?

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação. Área de Concentração: Mídia e Interação.

Aprovado em: 27/01/2025.

BANCA EXAMINADORA

Prof. Dr. Carlos Alexandre Barros de Mello Centro de Informática / UFPE

Prof. Dr. Gabriel de França Pereira e Silva Unidade Acadêmica do Cabo de Santo Agostinho/UFRPE

> Prof. Dr. Rafael Dueire Lins Centro de Informática / UFPE (orientador)

RESUMO

Definir a resolução correta de imagem para digitalização de documentos é essencial para garantir a preservação precisa de todas as informações necessárias e, ao mesmo tempo, otimizar a eficiência de aquisição, o tempo de processamento, a demanda de armazenamento e a largura de banda de transmissão. É fundamental que a resolução de digitalização garanta que o documento digitalizado, quando impresso em alta resolução nas dimensões originais, mantenha integralmente as informações do documento original, de forma a propiciar a preservação e manutenção da informação. Apesar de sua importância para a engenharia e preservação de documentos, os atuais padrões e recomendações de digitalização baseiam-se predominantemente em evidências experimentais e conhecimento empírico. Considerando esses parâmetros, a questão crucial "Qual é a resolução de digitalização mais adequada para documentos?" foi recentemente levantada pelo Professor George Nagy, um dos pesquisadores pioneiros na Engenharia de Documentos, destacando uma preocupação de longa data que não foi devidamente esclarecida pela comunidade de pesquisa. Este trabalho propõe uma resposta fundamentada nos princípios do teorema de Nyquist-Shannon da Teoria da Informação, esta abordagem tem por objetivo fornecer uma fundamentação teórica para a definição adequada da resolução na digitalização de documentos, de forma a assegurar a preservação das informações.

Palavras-chaves: Resolução de imagem. Digitalização de documentos. Teorema de Nyquist-Shannon

ABSTRACT

Defining the correct image resolution for document digitization is essential to ensure the accurate preservation of all necessary information while optimizing acquisition efficiency, processing time, storage demand, and the transmission bandwidth. It is crucial that the digitization resolution ensures that the digitized document, when printed in a high resolution at its original dimensions, fully retains the information from the original document, thus enabling the preservation and maintenance of the information. Despite its importance for document engineering and preservation, current digitization standards and recommendations are predominantly based on experimental evidence and empirical knowledge. Considering these parameters, the crucial question, "Which is the most suitable resolution for document digitization?" was recently raised by Professor George Nagy, one of the pioneers in Document Engineering, highlighting a long-standing concern that has not been adequately addressed by the research community. This work proposes an answer based on the principles of the Nyquist-Shannon theorem from Information Theory, this approach aims to provide a theoretical foundation for determining the appropriate resolution for document digitization, ensuring the preservation of information.

Keywords: Image resolution. Document digitization. Nyquist-Shannon Theorem.

LIST OF FIGURES

| Figure 1 - | Representation of an digital image acquisition | 18 |
|-------------|---|----|
| Figure 2 – | Representation of different dimensions of pairs of lines | 19 |
| Figure 3 - | Representation of different dimensions of pairs of lines | 20 |
| Figure 4 - | Representation of different sizes of objects formed by points (pixels) | 20 |
| Figure 5 - | Variation of spatial resolution in an image | 21 |
| Figure 6 – | (a) Left figure: Continuous function $f(t)$ sampled ΔT units apart. | |
| | (b) Right figure: Samples in the x -domain, where x is discrete | 24 |
| Figure 7 – | Steps in the Nyquist-Shannon sampling theorem | 26 |
| Figure 8 - | Samples of images from the Text-only Dataset | 30 |
| Figure 9 – | Samples of images from the General Dataset | 31 |
| Figure 10 – | Datasets Workflow | 31 |
| Figure 11 – | Example of the different resolutions of Image 01 in the original scan size | |
| | and after the cropping process. | 35 |
| Figure 12 – | Example of the different resolutions of Image 01 in the original scan color | |
| | and after the binarization process | 37 |
| Figure 13 – | Image 01, the frequency map and the Spectral Density of 75 dpi resolution. | 40 |
| Figure 14 – | Image 01, the frequency map and the Spectral Density of 150 dpi resolution. | 40 |
| Figure 15 – | Image 01, the frequency map and the Spectral Density of 200 dpi resolution. | 40 |
| Figure 16 – | Image 01, the frequency map and the Spectral Density of 300 dpi resolution. | 41 |
| Figure 17 – | Image 01, the frequency map and the Spectral Density of 500 dpi resolution. | 41 |
| Figure 18 - | Image 01, the frequency map and the Spectral Density of 700 dpi resolution. | 41 |
| Figure 19 – | Image 01, the frequency map and the Spectral Density of 1200 dpi resolution. | 42 |
| Figure 20 – | Small image, measuring 4.71×4.47 centimeters, in the resolution of 75 dpi. | 50 |
| Figure 21 - | Quantitative of image resolutions with minimal Levenshtein distance - 18 | |
| | images (μ : mean, σ^2 : variance) | 55 |
| Figure 22 – | Image illustrating the same Levenshtein Distance results to different reso- | |
| | lutions | 55 |
| Figure 23 – | Quantitative of image resolutions with minimal Levenshtein distance - 10 | |
| | images (μ : mean, σ^2 : variance) | 56 |

LIST OF CODES

| , | 34 |
|--|----|
| Source Code 1 — Cropping Algorithm applied in the dataset | 67 |
| Source Code 2 — Otsu binarization Algorithm applied in the dataset | 70 |
| Source Code 3 - Fourier Transform Algorithm applied in the dataset | 72 |
| Source Code 4 - Spectral Density Plot Algorithm applied in the dataset | 75 |
| Source Code 5 - OCR and the Levenshtein Distance Algorithm applied in the dataset. | 78 |

LIST OF BOARDS

| Board 1 | - | Example of | Levenshtein | distance metric | | | | | | | | | | | | | 29 |) |
|---------|---|------------|-------------|-----------------|--|--|--|--|--|--|--|--|--|--|--|--|----|---|
|---------|---|------------|-------------|-----------------|--|--|--|--|--|--|--|--|--|--|--|--|----|---|

LIST OF TABLES

| Table 1 – | Documents Category | 22 |
|-----------|---|----|
| Table 2 – | Documents and their spatial resolutions. (S.R.1 - Spatial Resolution provi- | |
| | ded by Gonzalez and Woods (2018), S.R.2 - Spatial Resolution provided by | |
| | Kenney and Chapman (2001)) | 23 |
| Table 3 – | Scanners specifications | 33 |
| Table 4 – | Text Dataset - Original images results (Image size in centimeters (cm), | |
| | Font size in points (pt), dpi - image resolution in dots per inch, png/bmp | |
| | - file size of the png/bmp image in megabytes (MB), ${\bf HF}$ - Highest frequency, | |
| | LD - Levenshtein distance.) | 45 |
| Table 5 – | Text Dataset - Binarized images results (Image size in centimeters (cm), | |
| | Font size in points (pt), dpi - image resolution in dots per inch, png/bmp/- | |
| | tiff - file size of the png/bmp/tiff image in megabytes (MB), HF - Highest | |
| | frequency, LD - Levenshtein distance) | 47 |
| Table 6 – | Proportion of highest frequency density above the maximum frequency of | |
| | 200 and 300 dpi in the General dataset - Original images | 51 |
| Table 7 – | General Dataset - Original images results. (Image size in centimeters (cm), | |
| | Font size in points (pt), dpi - image resolution in dots per inch, png/bmp | |
| | - file size of the png/bmp image in megabytes (MB), \boldsymbol{HF} - Highest frequency). | 57 |
| Table 8 – | General Dataset - Binarized images results (Image size in centimeters (cm), | |
| | Font size in points (pt), dpi - image resolution in dots per inch, png/bmp/- | |
| | tiff - file size of the png/bmp/tiff image in megabytes (MB), HF - Highest | |
| | frequency) | 58 |
| Table 9 – | Proportion of highest frequency density above the maximum frequency of | |
| | 200 and 300 dpi in the Text dataset - Original images | 60 |

LIST OF ACRONYMS

BMP Windows Bitmap

bpp bits per pixel

cm centimeters

DocEng'24 ACM Symposium on Document Engineering 2024

dpi dots per inch

HF Highest frequency

ICDAR International Conference on Document Analysis and Recognition

in inches

LD Levenshtein Distance

lp/mm line pairs per millimeter

MB megabytes

mm millimeters

OCR Optical Character Recognition

PNG Portable Networks Graphics

PSM Page Segmentation Mode

pt points

RGB Red Green Blue

TIFF Tagged Image Format

LIST OF SYMBOLS

 Δ Delta

 σ Sigma

 μ Mi

 $\pi \qquad \qquad \mathsf{Pi}$

CONTENTS

| 1 | INTRODUCTION | 3 |
|-------|---|---|
| 1.1 | OBJECTIVES | 6 |
| 2 | THEORETICAL BACKGROUND | 7 |
| 2.1 | IMAGE RESOLUTION | 9 |
| 2.1.1 | Spatial resolution | 9 |
| 2.2 | HISTORICAL DOCUMENTS IMAGE QUALITY | 1 |
| 2.2.1 | Historical documents resolution | 2 |
| 2.3 | CONTINUOUS-TIME AND DISCRETE-TIME SIGNALS | 3 |
| 2.4 | THE NYQUIST-SHANNON THEOREM | 4 |
| 2.4.1 | The Nyquist-Shannon Theorem applied to Images | 7 |
| 2.5 | EVALUATION MEASURE | 8 |
| 2.5.1 | Levenshtein Distance | 8 |
| 3 | MATERIALS AND METHODS | 0 |
| 3.1 | SCANNING PROCESS | 2 |
| 3.2 | CROPPING PROCESS | 3 |
| 3.3 | BINARIZATION PROCESS | 5 |
| 3.4 | SPECTRAL ANALYSIS | 7 |
| 3.5 | OPTICAL CHARACTER RECOGNITION (OCR) 4 | 3 |
| 4 | RESULTS | 5 |
| 4.1 | TEXT DATASET | 5 |
| 4.2 | GENERAL DATASET | 6 |
| 5 | CONCLUSIONS | 2 |
| | REFERENCES | 3 |
| | | |
| | APPENDIX 60 | 6 |
| | Appendix A – CROPPING ALGORITHM 6 | 7 |
| | Appendix B – OTSU BINARIZATION ALGORITHM 7 | 0 |
| | Appendix C – SPECTRAL ANALYSIS ALGORITHMS 7 | 2 |
| | Appendix D – OCR AND LEVENSHTEIN DISTANCE ALGORITHM 7 | 8 |

1 INTRODUCTION

The correct determination of scanning resolution is essential to ensure that all information present in a document, especially in historical documents, is accurately captured. According to The Library of Congress (2014), the choice of spatial and tonal resolution directly impacts the quality of document preservation in digital images. If the scanning resolution is set too low, critical elements of the document may be lost, conversely, excessively high resolution drastically increases scanner acquisition time yielding too large images that are slow to process and claim unnecessary storage space and computer bandwidth for network transmission.

Back in 1991, Rafael Dueire Lins was challenged by Graziela Peregrino, head of the Documentation Department of the Joaquim Nabuco Foundation in Recife, Brazil, to find a way to preserve all the information on the rich file of letters of Joaquim Nabuco¹ to future generations Lins (2011). Using the only scanner available at the Universidade Federal de Pernambuco in Brazil then, Lins digitized several letters of Joaquim Nabuco in 75, 150, 200, 300, 500, and 700 dpi in true-color and printed the images using a high-quality color laser printer. The printed documents were handed to the documentation experts of the Joaquim Nabuco Foundation, who concluded that the 150 dpi images had good enough quality to preserve all the information of the original document. As there was a huge international interest in the letters of the Nabuco bequest by social science researchers from all over the world, the 200 dpi scanner resolution was adopted by then, as it was compatible with FAX-machine standards, the technology available to send documents via computer/phone networks.

Over the years, prominent institutions dedicated to document digitization and preservation, including The Digital Library Federation Benchmark Working Group (2002), The Library of Congress (2014) and Forschungsgemeinschaft (2016), have conducted empirical studies to identify and recommend optimal scanning resolutions for different document types. The technical guidelines provided by those institutions are primarily based on visual inspection of the images generated in which documents were scanned at different resolutions, with the optimal resolution selected based on the visual "quality" of the resulting images. Among the findings, it is possible to observe a consensus of approximately 300 dpi as the ideal minimum resolution for grayscale and RGB-mode documents. According to Puglia et al. (2004), this resolution can also serve as a baseline for other document types, such as illustrations, maps, plans, ma-

Brazilian statesman, writer, and diplomat, the first Brazilian ambassador to the U.S.A, one of the key figures in the campaign for freeing black slaves in Brazil (b.1861-d.1910).

nuscripts, and large-format documents. Despite the agreement of using 300 dpi among those important institutions, there is no technical justification beyond empirical data analysis to support such recommendations.

During the welcome reception of the International Conference on Document Analysis and Recognition (ICDAR) 2023, Professor George Nagy asked Daniel Lopesti and Rafael Dueire Lins, the fundamental question: "Which is the most suitable scanner resolution for documents?". Rafael Dueire Lins promptly answered that the proper way to determine the resolution, considering the maintenance and preservation of information, would be through the Nyquist-Shannon theorem. Such answer was immediately acknowledged as providing the correct solution to the problem.

Whenever converting an analog signal into a digital representation two phases are involved. The first one is determining the number of samples one must take. As explained by Lathi (2005), the Nyquist-Shannon Theorem (1938) teaches that any (continuous) signal sampled with twice its maximum frequency may be re-constructed without losses. The second step in the digitization process is the quantization error, the maximum error tolerated in the representation of each sample.

Although the response given was immediately recognized as providing the solution to the question, as the theorem provides a theoretical basis for defining the minimum resolution required to capture all document details without information loss, ensuring that all the details of the document are retained whenever viewed or printed on the original size the validity of the solution had still to be proved. The suggested solution needed to be detailed and proved correct, which motivated the research developed in this dissertation.

The initial results of this dissertation Lins et al. (2024) was originally published in a short paper in the proceedings of the ACM Symposium on Document Engineering 2024 (DocEng'24), which was sent to Professors George Nagy, Daniel Lopesti, and Elisa Barney-Smith. Professor Nagy main feedback was:

"... your experiments provide reassurance about the adequacy of generally accepted digitization parameters. Testing both bi-level and color scans, and using the spectral density for estimating the range of image frequencies, seem sound."

Professor Elisa Barney-Smith informed us that:

"It is an interesting paper. It was the topic of a WG at the DAS 2002 workshop, but no detailed experiments were run, just a discussion and WG report paper (published in IJDAR in 2004)."

Besides that, Professor Barney-Smith listed some important papers of hers on image scanning Barney-Smith (1998), Barney-Smith and Andersen (2005) showing that the quality of scanner output has evolved drastically:

"Some information about the scanner itself can be measured through the use of various test targets, including a simple knife edge (to return the step response of the optics). See several of my early papers where I spent a lot of time exploring this. Scanners today are much higher quality than scanners 30 years ago, we have the disk space and the processing power to process over-sampled images, so it isn't as critical today, but people still explore the scanner optics."

This point raised by Professor Barney-Smith is very important because the Shannon-Nyquist theorem does not address *quantization errors*. But it is also relevant to notice that such errors are device-dependent and that their effect would be more noticeable in higher-resolution images. Another concern raised by Professor Barney-Smith was:

"Another point you should mention in future papers would be a discussion of WHAT information you want to extract from a scan of a document. You touched on this when you mentioned historical documents vs maps... If it is just the text content, it has a much lower frequency content than if one were looking at the paper fibers of the same document."

It is important to stress here that this paper aims to determine the "correct" scanner resolution to keep all the document information making a high-definition same-size print of the scanned document indistinguishable from the original. However, on two occasions scanning documents in 300 dpi was enough to distinguish the age of documents as described by Barboza, Lins and Jesus (2013) and even to detect frauds with latter-added strokes in hand-written documents as detailed by Barbosa et al. (2014).

This dissertation also widens the scope of Lins et al. (2024) in two ways. The first one was the inclusion of more complex and fine-detailed images that employ tiny fonts that would generate very high-frequency components. The second path uses Optical Character Recognition (OCR) to analyze how the resolution variation affects the recognition rate.

1.1 OBJECTIVES

The general objective in this dissertation is to bring a theoretical basis and show evidences that the most suitable resolution for document digitization are based on the principles of the Nyquist-Shannon theorem from Information Theory and ensure the accurate preservation of all necessary document information. Thus, this dissertation expands the scope of the study published in Lins et al. (2024), incorporating new experiments, additional images and detailed analyses.

Such target is reached by:

- Enlarging and evaluating the datasets of document used considering a diversity of detailed images, including those with tiny fonts and high-frequency components, at multiple resolutions to represent a much wider range of document types in the experiments performed.
- Investigating the effects of the binarization process in enhancing the recognition of text and fine details at different resolutions.
- Using OCR tools to evaluate the impact of resolution on recognition rates, using the Levenshtein Distance to analyze the effects of scanner resolution in the accuracy of the transcription of the texts.
- Experimentally validating the Nyquist-Shannon theorem as a basis for determining optimal scanner resolution, considering several practical implications.

2 THEORETICAL BACKGROUND

According to Gonzalez and Woods (2018), in the context of digital image processing, an image can be described as a two-dimensional function, f(x,y), where x and y represent spatial coordinates of a plane and the amplitude of f is referred to as the intensity or gray level of that specific point. If x, y and the intensity values of f are all finite and discrete, the image is considered a digital image. The value of f at spatial coordinates f(x) is regarded as a scalar quantity physically determined by the image source, with values being proportional to the energy of the image. Most images are acquired through the interaction between the lighting source falling on the scene and the reflection or absorption of energy carried out by the elements within the photographed scene. Thus, the function f(x,y) can be characterized as the outcome of the combination of the illumination and reflectance elements, defined respectively by f(x,y) and f(x,y):

$$f(x,y) = i(x,y)r(x,y)$$
(2.1)

where

$$0 \le i(x, y) < \infty \tag{2.2}$$

and

$$0 \le r(x, y) \le 1 \tag{2.3}$$

Gonzalez and Woods (2018) asserts that reflectance is determined by the properties of the imaged objects and constrained between 0 (indicating total absorption) and 1 (indicating total reflectance). In instances of images formed through transmission, like a chest X-ray, the function would rely on the transmissivity function rather than the reflectance function. However, the boundaries remain consistent, and the image function remains represented as the product in equation 2.1.

Reference Sonka, Hlavac and Boyle (2015) states that computerized image processing relies on digital image functions. Therefore, an image must be represented using suitable discrete data, such as a matrix. In Figure 1, there is a representation of a digital image acquisition process using array sensors. In such an example, the imaging system collects the incoming energy and focuses it onto an image plane. Since the array sensor coincides with the focal plane, the resulting outputs will be proportional to each sensor's integral of the light received.

These outputs are then converted to an analog signal by digital and analog circuitry, which is then digitized, resulting in a digital image.

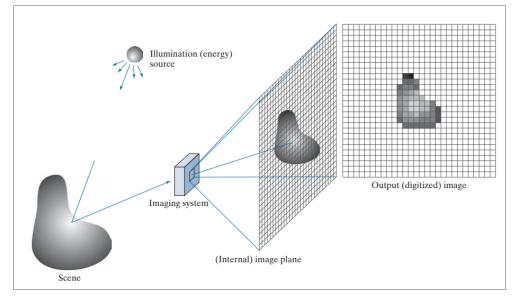


Figure 1 – Representation of an digital image acquisition.

Source: GONZALEZ; WOODS (2018)

Considering the diversity of potential applications for digital image processing, Gonzalez and Woods (2018) highlights that images can be categorized by source, such as X-ray, visual, and infrared. The most prevalent energy source utilized today is the electromagnetic spectrum. This spectrum is grounded in the revelation that the resultant light beam is not white but comprises a continuous spectrum of colors extending from violet to red. Bushong (2013) describes that when sunlight passes through a prism, it becomes clear that white light consists of photons spanning a spectrum of wavelengths, and the prism functions to disperse and organize the emitted light into colors by refracting various wavelengths through distinct angles. Visible light comprises only a portion of the electromagnetic spectrum, extending over 25 orders of magnitude. Figure 2 illustrates the values of energy, frequency, and wavelength and identifies the three imaging windows of X-ray, visual, and MR imaging according to their associated values.

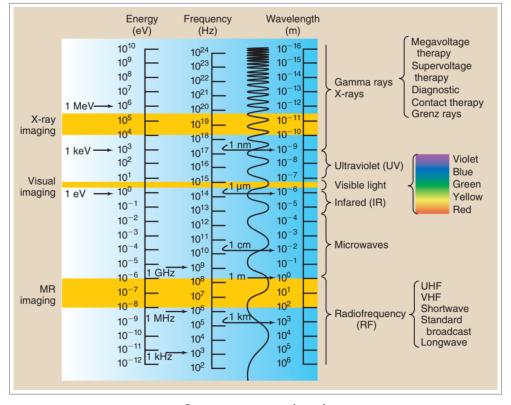


Figure 2 - Representation of different dimensions of pairs of lines.

Source: BUSHONG (2013)

2.1 IMAGE RESOLUTION

Regarding the quality of a digital image, Sonka, Hlavac and Boyle (2015) emphasizes that it is directly proportional to its resolution. The author defines that the resolution of an image can be categorized into four main types: spatial resolution, spectral resolution, radiometric resolution and temporal resolution. This work will focus more specifically on spatial resolution and its impact on document scanning.

2.1.1 Spatial resolution

Bushberg (2012) states that spatial resolution refers to the ability of an imaging system to identify small details, that is, to distinguish small nearby elements that have different contrasts in an image. Sonka, Hlavac and Boyle (2015), highlights that the distance of the samples in the image plane determines this aspect. Therefore, the higher the spatial resolution, the better the visualization of details in an image.

Gonzalez and Woods (2018) add that, when considering the measurement of the smallest

detail discernible in the image, the resolution in space is usually quantified in terms of line pairs per unit distance or dots (pixels) per unit distance. According to Bushong (2013), a line pair is defined by a high-contrast line separated by a space of width equal to that of the line. For radiological images, for example, it is common to use the unit line pairs per millimeter (lp/mm), referring to the lp/mm present in the image composition. According to Gonzalez and Woods (2018), one point per unit of distance refers to the size of the pixel present in the image composition. This measurement is generally used for printing files and is usually represented by the number of pixels in an inch, using the unit of dots per inch (dpi).

In Figures 3 and 4, respectively, representations of dimensions of pairs of lines and points (pixels) that can be applied concerning distance to determine spatial resolution are illustrated.

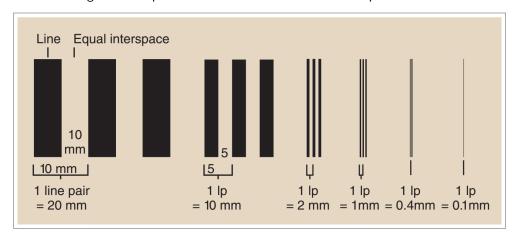


Figure 3 – Representation of different dimensions of pairs of lines.

Source: BUSHONG (2013)

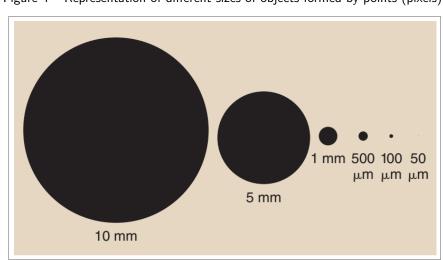


Figure 4 - Representation of different sizes of objects formed by points (pixels).

Source: BUSHONG (2013)

In figure 5 it is possible to see an example of the impact generated in an image at different spatial resolutions, considering the image at resolutions of 930 dpi, 300 dpi, 150 dpi, and 72 dpi. Gonzalez and Woods (2018), highlights that the difference in spatial resolution also affects the size of the image. However, in terms of comparing the impact of spatial resolution, the images were zoomed in for better visualization.

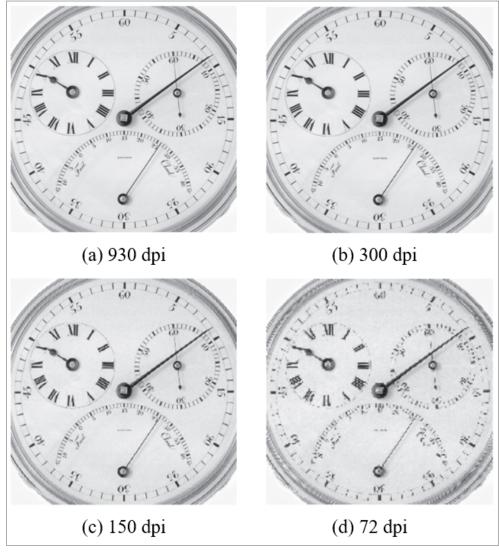


Figure 5 – Variation of spatial resolution in an image.

Source: GONZALEZ; WOODS (2018)

2.2 HISTORICAL DOCUMENTS IMAGE QUALITY

According to Kenney and Chapman (2001), determining image quality requirements depends on the characteristics of the type of document to be scanned. Capturing a finely detailed

historical document written in pen and ink, for example, requires a higher resolution to correctly identify its elements. The author defines that, in general, documents can be classified into four identification categories as illustrated in table 1.

Table 1 – Documents Category

| Document Category | Details |
|-------------------|---|
| TEXTS/DASH | Can be produced manually, mechanographed or by machine. Usually in black and white. Includes books, manuscripts, newspapers, reports, typed or laser-printed documents, architectural plans, maps, line drawings, etchings, lithographs, and musical scores. |
| HALF TONE | In color or black and white. Reproductions, usually created from a photograph, made up of small dots or squares or lines, which are used to represent continuous tones. Most photographs in publications are halftones. |
| CONTINUOUS TONE | Color or black and white. Includes graphic illustrations in which all gray and color values can be reproduced: photographs, crayons, chalk and some pencil drawings, acrylics, watercolors and photographically reproduced facsimiles. |
| MIXED | Color or black and white. Refers to items containing half- tone text and images or continuous tone text and images, such as newspapers, magazines, picture books, programs, and sheet music covers. Does not include text and line drawings together. |

Source: Prepared by the author based on (KENNEY; CHAPMAN, 2001), 2024

Kenney and Chapman (2001) emphasizes, that in addition to document categories, other important factors to be considered in the scanning process are: level of detail, type of paper, level of contrast between ink and paper, sharpness, production process (machine-made or handmade) and quality of the document, considering if it is damaged, stained or incomplete. The quality of the digital image is inherently dependent on the quality of the source. For example, scanning a blurred image will result in a blurred digital copy, regardless of the resolution. In addition, the choice of hardware and software used in the digitization process can also influence the final output.

2.2.1 Historical documents resolution

According to Nagy and Seth (1984), the resolution applied by the scanner in the document scanning process is defined by the level of detail required in the image. When considering a

two-level image (black and white), for example, 400 lines per inch would be considered an adequate value for spatial resolution.

The resolution for scanning documents for printing, according to Gonzalez and Woods (2018), is defined in terms of dpi. Table 2 contains different categories of documents followed by their spatial resolution information for printing provided by Gonzalez and Woods (2018) and Kenney and Chapman (2001).

Table 2 – Documents and their spatial resolutions. (S.R.1 - Spatial Resolution provided by Gonzalez and Woods (2018), S.R.2 - Spatial Resolution provided by Kenney and Chapman (2001))

| Document | S. R. 1 | S. R. 2 |
|-----------------------|----------------|-------------------------------|
| Newspapers | ≈ 75 | $\approx 120 \text{ or } 300$ |
| Magazines | ≈ 133 | _ |
| Flyers | ≈ 175 | ≈ 600 |
| Books | ≈ 2400 | ≈ 120 or 300 |
| Monographs | _ | ≈ 600 |
| Periodicals | _ | ≈ 600 |
| Office documents | _ | ≈ 300 |
| Tecnical draws | _ | ≈ 600 |
| Maps | _ | ≈ 600 |
| Handwritten materials | _ | ≈ 300 |

Source: Prepared by the author based on Gonzalez and Woods (2018) and Kenney and Chapman (2001), 2024

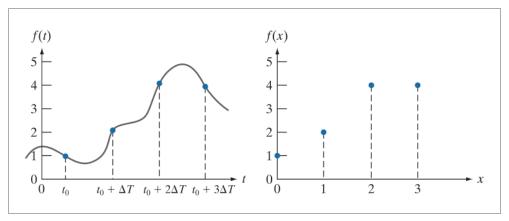
2.3 CONTINUOUS-TIME AND DISCRETE-TIME SIGNALS

Gonzalez and Woods (2018) explains that a continuous signal must be converted into a sequence of discrete values to be processed by a computer, a process that involves both sampling and quantization. Lathi (2005) emphasizes that continuous-time signals can be effectively analyzed by processing their discrete samples using discrete-time systems. To ensure accurate reconstruction of the original signal from its samples, the sampling rate must be sufficiently high, minimizing reconstruction errors. The sampling theorem provides the theoretical foundation for this process, defining the conditions under which continuous-time signals can be reliably sampled and reconstructed from their discrete sampled values.

Lathi (2005) further describes sampling theory as the connection between continuous-time and discrete-time representations. A continuous-time signal is expressed as a sequence of pulses and is specified in a continuum of time values t, while a discrete-time signal represents the same

information numerically, as discrete values of t. Thus, both formats represent the same data, ensuring that each sample's information is equivalent, regardless of the representation. Some examples of continuous-time signals include outputs from telephones and analog cameras, while examples of discrete-time signals include monthly sales figures of a corporation and daily stock market averages. Figure 6(a) illustrates a continuous function f(t) sampled at regular intervals ΔT , where each sample corresponds to a specific point on the continuous function. In contrast, Figure 6(b) represents these samples in the discrete x-domain, where the samples are indexed sequentially by integers (0, 1, 2, and 3) instead of their exact positions in time. This highlights that the sampled values at these points remain the same, ensuring information equivalence between the two domains.

Figure 6 – (a) Left figure: Continuous function f(t) sampled ΔT units apart. (b) Right figure: Samples in the x-domain, where x is discrete.



Source: GONZALEZ; WOODS (2018)

2.4 THE NYQUIST-SHANNON THEOREM

According to Lathi (2005), the basis of the Nyquist-Shannon Theorem is the Sampling Theorem, originally known as the "Interpolation Formula". Siebert (1986) describes that the general formulation of the "Sampling Theorem" was credited by H. S. Black to Cauchy in 1841 and around the 1920s, Nyquist, Carson, and Hartley revisited and refined the established concepts, laying the foundation for modern communication theory.

In 1948, the work of C. E. Shannon further formalized and popularized the theorem, Shannon (1949) even points out in his work that the theorem had already been portrayed previously by mathematicians in other forms, but had not appeared explicitly in the communication theory literature until Nyquist returned to work on the essential idea of the theorem and stated

that, considering a f(t) limited in bandwidth to a maximum frequency W and limited to the time interval T, with samples taken at intervals of $\frac{1}{2W}$ seconds apart, a total of 2TW samples will fully capture the information within the interval. Therefore, approximately 2TW numbers are enough to specify the function, based on a Fourier series expansion over the time interval T.

A band-limited function is defined as a function f(t) whose Fourier transform is zero for frequency values outside a finite interval $[-\mu_{max}, \mu_{max}]$, as described by Gonzalez and Woods (2018). This restriction implies that, for a function to be confined within a specific time interval T, all frequency values outside this range must be zero, as stated by Shannon (1949).

Shannon articulated the theorem as follows:

If a function of time f(t) is limited to the band from 0 to W cycles per second, it is completely determined by giving its ordinates at a series of discrete points spaced $\frac{1}{2W}$ seconds apart (SHANNON, 1948, p. 34).

According to Lathi (2005), the samples can be taken arbitrarily at any instants, resulting in different intervals, as long as the sampling instants are recorded and on average of 2B samples per second. Gonzalez and Woods (2018) states that the term *Nyquist rate* defines a sampling rate equal to twice the highest frequency.

Based on Lee and Varaiya (2011), figure 7 illustrates an example of a signal sampling and reconstruction process based on The Nyquist-Shannon Theory in terms of its basis. Each line of pair graphs in the figure represents a step of the process as follows:

- Step 1: The left image represents the continuous-time original signal x(t), and the right image shows its continuous Fourier transform X(w), where the data is in terms of the frequency and band-limited to $[-\pi/T, \pi/T]$.
- Step 2: The signal x(t) is sampled at intervals of T, resulting in the discrete signal y(n). On the right, the discrete-time Fourier transform (DTFT) Y(w) is shown, where the spectrum is periodically replicated with a period of $2\pi/T$.
- Step 3: The discrete-time signal y(n) is converted into a train of impulses w(t) by the Impulse Generator. In the frequency domain, W(w) represents the scaled and periodic spectrum 1/T.
- Step 4: The impulse train w(t) is filtered by the $Sinc_T$ function, which is a scaled version of the sinus cardinalis function, defined as:

$$\operatorname{sinc}_{T}(t) = \frac{\sin(\pi t/T)}{\pi t/T} \tag{2.4}$$

where:

- -t = time
- -T =the sampling period

This function acts as an interpolation kernel, ensuring that each sampled point contributes correctly to the reconstruction of the original signal. In frequency domain terms, $Sinc_T$, performs low-pass filtering, removing spectral replicas and preserving the original spectrum. Then, the inverse Fourier transform reconstructs the continuous-time signal z(t), shown on the left, which matches the original signal x(t).

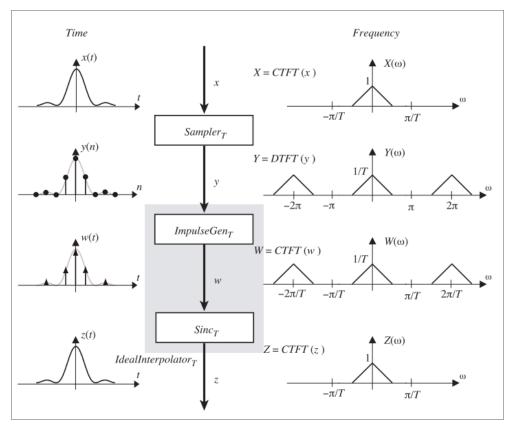


Figure 7 – Steps in the Nyquist-Shannon sampling theorem.

Source: LEE; VARAIYA (2011)

2.4.1 The Nyquist-Shannon Theorem applied to Images

In terms of a digital image, Sonka, Hlavac and Boyle (2015) describes that a continuous function f(x,y) can be sampled by applying a discrete grid of sampling points in the plane, or the image function can also be expanded by applying an orthonormal function as a basis, such as the Fourier transform, resulting in coefficients of the expansion that represent the digitized image.

Colarusso et al. (1999) points out that the Nyquist-Shannon theorem is essential to direct imaging applications, as the signal is sampled through discrete pixel elements in an array. Sonka, Hlavac and Boyle (2015) also emphasizes that, there is a connection between the digital sampling density and the details that an image will contain. The Nyquist-Shannon principle implies that the sampling rate should be at least twice the highest spatial frequency to ensure accuracy. For example, Colarusso et al. (1999) describes that considering a diffraction-limited arrangement, the product of the lateral spatial resolution and the total magnification must be twice the pixel size of the image, as defined by Equation 2.5:

$$M_{tot}R_L \ge 2p \tag{2.5}$$

where:

- M_{tot} :total magnification
- R_L : lateral spatial resolution
- p: pixel size

Adhering to the Nyquist rate helps maintain the spatial fidelity of images and prevents sampling artifacts. In this context, oversampling does not provide additional information, resulting in a scenario commonly referred to as "empty magnification." Conversely, undersampling below the Nyquist-Shannon limit leads to distortions or artifacts, such as Moiré patterns, a phenomenon known as spatial aliasing Gonzalez and Woods (2018). Sonka, Hlavac and Boyle (2015) describes that aliasing happens when there is an overlapping of the periodically repeated results of the Fourier transform f(u,v) of an image with band-limited spectrum, a result that can be prevented by choosing a sampling interval that it is less than half of the smallest interesting detail in the image.

Colarusso et al. (1999) explains that the resolution of an imaging device is primarily determined by its maximum capability. In terms of document scanning, Barney-Smith (1998) points out that the bitmap resulting from a two-level scan depends greatly on the parameters of the scanner used. The most common and simplest way to digitize a document in companies, or even at home, is through a flatbed scanner. Federal Agencies Digital Guidelines Initiative (2016) states that these models are easy to use, versatile and can achieve different resolution ranges, with values from 75 to 1200 dpi depending on the model used. Therefore, selecting the appropriate scanner resolution is essential to capture all document details effectively while avoiding the drawbacks of prolonged image acquisition and processing times, excessive storage requirements, and unnecessary network bandwidth usage.

2.5 EVALUATION MEASURE

2.5.1 Levenshtein Distance

Defined by Levenshtein (1966) in "Binary codes capable of correcting deletions, insertions and reversals", the Levenshtein Distance (LD) refers to the minimal number of insertions, deletions and replacements needed to to make two strings identical. The metric was proposed in the context of the code theory to solve problems in detecting and correcting transmission data. According to Navarro (2001) the Levenshtein distance is also known as *Edit distance* and, in a simplified definition, all the operations involved to guarantee the string matching are assigned as a cost of 1. The LD is widely used in applications such as evaluating spell-checkers, OCR systems, and DNA sequence analysis.

Shuwandy et al. (2020) defines that mathematically, the Levenshtein distance between two strings a, b with respectively length of |a| and |b| is given by $lev_{a,b}$ and detailed in Equation 2.6. The $lev_{a,b}(i,j)$ is the distance between the first i characters of a and the first j characters of b.

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) \text{ if } \min(i,j) = 0, \\ lev_{a,b}(i-1,j) + 1, & \text{(deletion)}, \\ lev_{a,b}(i,j-1) + 1, & \text{(insertion)}, \\ lev_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)}, & \text{(match or mismatch)}. \end{cases}$$
 (2.6)

According to Shuwandy et al. (2020), the first element in the minimum refers to deletion, the second to insertion and the third refers to the match or mismatch of symbols, depending on their similarity. The $1_{(a_i \neq b_j)}$ is the indicator function in the match or mismatch elements, where:

$$\begin{cases} 1_{(a_i \neq b_j)} = 0 \text{ if } a_i = b_j \\ 1_{(a_i \neq b_j)} = 1 \text{ if } a_i \neq b_j \end{cases}$$
 (2.7)

Shuwandy et al. (2020) reports an LD example between the words 'GEEI' and 'GEELY' to evaluate their similarity, considering 'GEELY' as the password of a system. Board 1 shows that the LD is 2, because of the two last characters.

Board 1 - Example of Levenshtein distance metric

| G | Е | Е | | |
|---|---|---|---|---|
| G | Ε | Ε | L | Υ |

Source: SHUWANDY et al. (2020)

3 MATERIALS AND METHODS

A distinct set of diversified files, representing the scope of document engineering, was selected to establish the optimal scanning resolution that balances image quality, processing time and storage space in document digitization. This selection was referred to as the "Test dataset" and is broadly divided into two main categories: text-only documents ("Text") and documents containing both text and images ("General"). The test dataset includes twenty-eight distinct documents scanned at resolutions: 75, 150, 200, 300, 500, 700, and 1200 dpi. These documents are book pages and single paper files in different papers, font styles, sizes, formats and colors. The content includes:

- Typed Texts, in English, Portuguese and Spanish;
- Handwritten Texts in Portuguese;
- Maps;
- Architectural Plan,
- Images and Illustrations.

In figure 8 it is possible to observe samples of images from the text-only dataset followed by samples of images from the general dataset in figure 9.

5
Canonical transformations; Poisson brackets

**The Cano

Figure 8 – Samples of images from the Text-only Dataset

Source: The author (2024)

A CONSTTUCÃO CIDADA

(Upper Cimerative

and a printer beigning of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the State of Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer beigning of the Carents of Army or account performance

and a printer be

Figure 9 – Samples of images from the General Dataset.

Source: The author (2024)

The eighteen images of English texts were classified into the "Text Dataset", while the remaining ten images were organized into a separate group referred to as the "General Dataset". The flowchart in Figure 10 presents the process carried out to prepare each data set for the Fourier Transform application.

Cropped Binarize Binarized Images Text and Images Apply Fourrier
Transform **Dataset** Crop Image: (10 images) Test Dataset (28 images) Cropped Images Apply Fourrier Binarize Binarized Image **Images** Transform Text-only Datase Crop Image Labels **Dataset Process**

Figure 10 - Datasets Workflow

Source: The author (2024)

As mentioned in the previous section, the Nyquist-Shannon theorem applies to continuous signals that are bandwidth-limited. It is a principle in Fourier analysis that sharper signals (those with higher derivatives in the time or spatial domain) exhibit higher frequency components. When signals display "discontinuities" in time or space, a phenomenon known as "frequency overshooting" occurs, resulting in intense oscillations in the frequency domain. By closely examining a color document image, one observes that the pixel colors at boundaries transition smoothly, whereas binary images have sharp edges, which can introduce perceived

signal discontinuities. The selected documents were examined in true-color and binary formats, to facilitate a more thorough frequency domain analysis.

To determine the optimal scanning resolution, the approach here involved analyzing the spectral density of various resolutions of the same image to locate the "maximum frequency". The smallest resolution to yield such a frequency is considered the most efficient choice, balancing scanning time, storage requirements, network transmission bandwidth, and processing time, as processing time typically scales with file size. This study also evaluates the effect of different scanning resolutions on the transcription quality of text-only documents by comparing the Levenshtein distances between transcribed texts and their ground-truth counterparts.

3.1 SCANNING PROCESS

In this study, the first step in determining the optimal resolution for scanned documents was to digitize each document at 75, 150, 200, 300, 500, 700, and 1200 dpi. The files were scanned in color mode and saved in Portable Networks Graphics (PNG) format, a widely used raster format known for its efficient lossless compression and high bit-depth compatibility, as described by Puglia et al. (2004).

The scanning process was performed using three scanners: Epson EcoTank L4160, Epson EcoTank L8180, and HP LaserJet Pro Color M479FDW. The technical specifications of each device, as provided in their respective manuals, are summarized in Table 3. Based on these specifications, the first two scanners were capable of digitizing documents at all seven predefined resolutions. However, the HP LaserJet Pro Color M479FDW scanner was limited to the following resolutions: 75, 200, 300, and 1200 dpi.

Table 3 – Scanners specifications.

| Specification | Epson L4160 | Epson L8180 | HP M479FDW | | | |
|-----------------------|--|---|---|--|--|--|
| Scanner type | Flatbed | Flatbed | Flatbed | | | |
| Photoelectric device | CIS | CIS | CIS | | | |
| Effective pixels | 10200 × 14040 pixels (1200 dpi) | - | - | | | |
| Maximum document size | US letter or A4 (8.5 \times 11.7 inches (in) or 216 \times 297 millimeters (mm)) | Legal (8.5 \times 14 in or 216 \times 356 mm) | Legal (8.5 \times 14 in or 216 \times 356 mm) | | | |
| Optical resolution | 1200 × 2.400 dpi | 1.200 × 4.800 dpi | 1200 × 1200 dpi | | | |
| Output resolution | 50 to 9600 dpi in 1 dpi increments | - | - | | | |
| Image data | Color: 48 bits per pixel (bpp) (internal), 24 bpp (external). Grayscale: 16 bpp (int.), 8 bpp (ext.). Black and White: 16 bpp (int.), 1 bpp (ext.) | Color: 48 bpp (internal), 24 bpp (external). Grayscale: 16 bpp (int.), 8 bpp (ext.). Black and White: 16 bpp (int.), 1 bpp (ext.) | - | | | |
| Light source | LED | LED | LED | | | |

Source: Prepared by the author based on EPSON (2019), EPSON (2021) and HP (2019), 2025.

3.2 CROPPING PROCESS

The Test Dataset was processed using a cropping algorithm, presented in Appendix A, to remove visible scanner edges and ensure a clear focus on the document page.

The Source Code 1 begins by selecting the image to be processed and defining its path. After verifying the existence of the image according to the specified path, the function

"cv2.imread(path)" loads the found images and displays an error message if the image is not located. The cropping parameters are manually and individually adjusted for each image, considering its specific characteristics. The parameters 'top', 'bottom', 'left', and 'right' define the cropping percentages for the top, bottom, left, and right edges of the image, respectively. These values are stored in the dictionary "cut_values", ensuring a clear mapping between cropping directions and their respective values.

The cropping function cut_and_save calculates the number of pixels to be cropped based on the defined percentages and performs the cropping operation using image[top:bottom, left:right]. This function takes the following parameters:

- image: the image to be cropped;
- output_path: the path to save the cropped image;
- resolution: the resolution of the image being processed;
- cut_percent: the cropping percentages.

It is important to highlight that the cropping percentages were applied across all different image resolutions, in line 43 of the Source Code 1, it is established as follows:

```
top, bottom, left, right = [int(p * max(image.shape) / 100) for p in
    cut_percent]
```

The line defines that, the cropping values are calculated as a percentage of the maximum image dimension (max(image.shape)). Thus, the cropping process is proportional to the image size, adapting to different resolutions. This approach helps maintain relative consistency in the visual content regardless of varying resolutions.

At the end of the process, the new cropped images are saved in PNG format and the applied cropping parameters are saved in an Excel file. An example of the cropping result for Image 01 across different resolutions is shown in Figure 11.

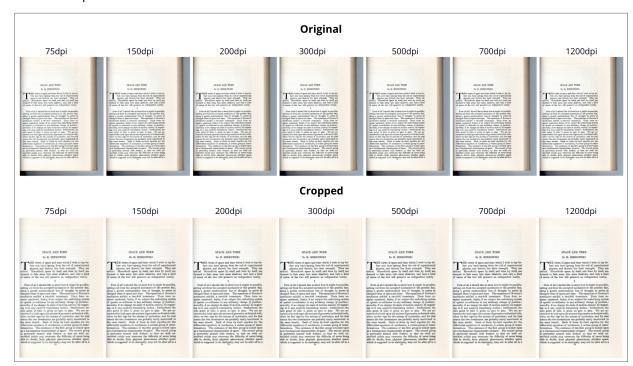


Figure 11 – Example of the different resolutions of Image 01 in the original scan size and after the cropping process.

Source: The author (2025)

3.3 BINARIZATION PROCESS

After the cropping process, the images were subjected to binarization. According to Lins et al. (2021), the binarization of document images remains a significant research area, particularly due to its importance in processing stages of document analysis and optical character recognition (OCR). In this study, the Otsu binarization algorithm Otsu (1979) was applied to the test images due to its accessibility, ease of implementation, and consistent image quality. The implementation of this method can be found in Appendix B.

Initially, the Source Code 2 defines the specific image to be processed and sets its corresponding path. Subsequently, a validation step is performed to verify if the image exists. If the image is successfully found, it is loaded using the function cv2.imread(path) from the OpenCV library. In the event of an invalid path or missing image, an error message is displayed.

The main function responsible for the binarization process is binarize_and_save. This function begins by converting the original image to grayscale using the OpenCV function cv2.cvtColor(image, cv2.CoLoR_BGR2GRAY). This conversion is essential because Otsu's thresholding technique operates more effectively on grayscale images, where pixel intensities range from 0 (black) to 255 (white). According to the OpenCV documentation Doxygen

(2024), the use of the conversion code cv2.COLOR_BGR2GRAY ensures a reduction in computational complexity, facilitating the accurate application of the binarization technique.

As defines by Doxygen (2024), the subsequent step in line 30 of the Source Code 2, applies the Otsu thresholding method using the following OpenCV flags:

- cv2.THRESH_BINARY: Performs binary thresholding, where pixel values below the threshold are set to 0 (black) and values equal to or above the threshold are set to 255 (white);
- cv2.THRESH_OTSU: Automatically calculates the optimal global threshold for the image, minimizing the intra-class variance between the foreground and background pixel intensities.

Otsu's method is effective with images that exhibit a bimodal histogram, that presents two distinct peaks corresponding to the foreground and background, by automatically selecting the ideal threshold. It eliminates the subjectivity associated with manually determining a suitable threshold value, thus reducing the likelihood of misclassification.

Finally, the binarized image is saved to the specified path using the cv2.imwrite(output_path, binarized_image) function. This approach ensures that the resulting binarized images can be utilized for further analytical steps. Figure 12 presents an example of the binarized result for Image 01 in its different resolutions.

Original 75dpi 150dpi 200dpi 300dpi 500dpi 700dpi 1200dpi **Binarized** 75dpi 150dpi 200dpi 300dpi 500dpi 700dpi 1200dpi

Figure 12 – Example of the different resolutions of Image 01 in the original scan color and after the binarization process.

3.4 SPECTRAL ANALYSIS

The images were prepared as described previously and then the Fourier Transform was applied to each one using the NumPy library in Python, a tool widely used for scientific computing. The source code of the Fourier transform algorithm applied in the dataset is in Appendix C

The Source Code 3 starts locating the defined images, followed by checking for their existence. The images are loaded in grayscale mode using "cv2.IMREAD_GRAYSCALE", considering that Red Green Blue (RGB) images require separate transforms for each channel, the use of grayscale images reduces complexity, improving computational efficiency while still preserving meaningful frequency data based on the different intensities.

For each image, the Fourier Transform was applied considering the following code, located in lines 75-77 of the Source Code 3

```
transform_image = np.fft.fft2(image)
center_transform_image = np.fft.fftshift(transform_image)
magnitude_spectrum = np.log(np.abs(center_transform_image) + 1)
```

where:

- np.fft.fft2(image): computes the 2D Fourier Transform, performing the calculation
 of the two-dimensional discrete Fourier transform across specified axes within an Mdimensional matrix;
- np.fft.fftshift(): moves to center the zero-frequency component within the spectrum, providing better visualization;
- np.log(np.abs(center_transform_image) + 1): converts the frequency domain to a log scale for improved contrast, calculating the logarithm of the magnitude of the spectrum.

The function calculate_max_frequency finds the maximum frequency component in the transformed image, corresponding to the dominant structure in the image. The highest frequency was identified with the numpy.argmax, which finds the indices of maximum values and was used here to pinpoint the peak frequency within the magnitude spectrum as defined by Numpy (2024). With the magnitude spectrum results, it was possible to plot a frequency map and frequency range for each image resolution.

Following the same Fourier-based spectral analysis on the Source Code 3, the Source Code 4 shows the results in a Spectral Density Plot. The Spectral Density function S(f) represents the distribution of energy across different frequency components and, as described by Oppenheim, Willsky and Nawab (1994), is defined as:

$$S(f) = |X(f)|^2 (3.1)$$

where

• X(f) = Fourier transform of the signal x(t)

Considering that the signal has finite energy:

$$E = \int_{-\infty}^{\infty} |x(t)|^2 dt < \infty$$
 (3.2)

To estimate the Spectral Density, the magnitude spectrum was calculated using the Fourier Transform, and then a normalized histogram of the frequency magnitudes was created. This histogram represents the relative occurrence of different frequencies in the image and the spectral analysis functions were defined as:

```
def calculate_spectral_energy(magnitude_spectrum):
    return np.sum(magnitude_spectrum**2)
```

 calculate_spectral_energy: calculates the total spectral energy by summing the squares of the magnitude of the spectrum. This value represents the total energy of the signal in the analyzed frequency components;

```
def calculate_max_frequency(magnitude_spectrum):
    row_center, col_center = np.unravel_index(np.argmax(magnitude_spectrum),
        magnitude_spectrum.shape)
    return magnitude_spectrum[row_center, col_center]
```

 calculate_max_frequency: finds the pixel with the highest frequency magnitude in the spectrum;

```
def calculate_frequency_distribution(magnitude_spectrum):
    hist, bins = np.histogram(magnitude_spectrum.flatten(), bins=50, range
        =(0, magnitude_spectrum.max()))
    hist = hist / hist.sum()
    predominant_frequencies = bins[:-1][hist > 0.1]
    return hist, bins, predominant_frequencies
```

• calculate_frequency_distribution: creates a histogram of frequency magnitudes to identify the frequency distribution of the image. The relative frequency threshold above 0.1 was used, as defined by "bins[:-1][hist > 0.1]", which facilitated the generation of a frequency recurrence graph for the spectrum, normalizing the values so that the sum is 1 and considering values above 10% of occurrence according to Numpy (2024);

The frequencies in the resulting spectrum are based on the image resolutions and correspond to cycles per pixel. The Fourier Transform allowed the extraction of a frequency map, the identification of the frequency range, the determination of the highest frequency and the creation of a Spectral Density plot for each image, all using the Numpy library based on NumPy (2024) and subsequently visualized with the Matplotlib library as stated in the Matplotlib (2024) documentation.

Figures 13 to 19 show the frequency maps and ranges for Image 01 across different resolutions, accompanied by the Spectral Density graph. The maps are displayed using Matplotlib's jet colormap model described in Matplotlib (2024).

Figure 13 – Image 01, the frequency map and the Spectral Density of 75 dpi resolution.

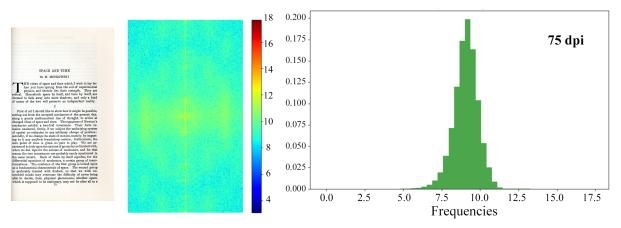
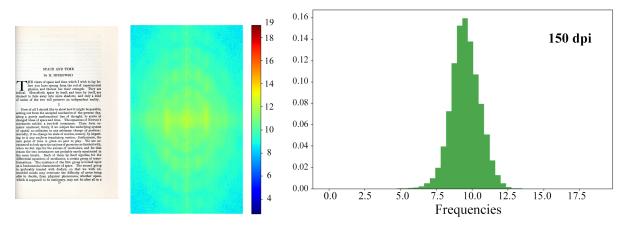


Figure 14 – Image 01, the frequency map and the Spectral Density of 150 dpi resolution.



Source: The author (2024)

Figure 15 – Image 01, the frequency map and the Spectral Density of 200 dpi resolution.

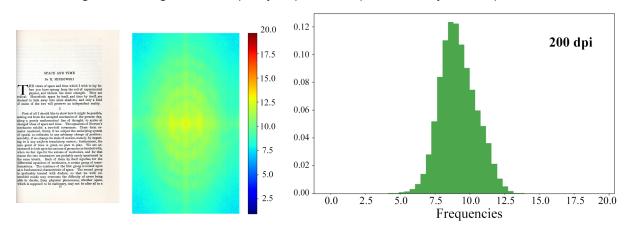


Figure 16 – Image 01, the frequency map and the Spectral Density of 300 dpi resolution.

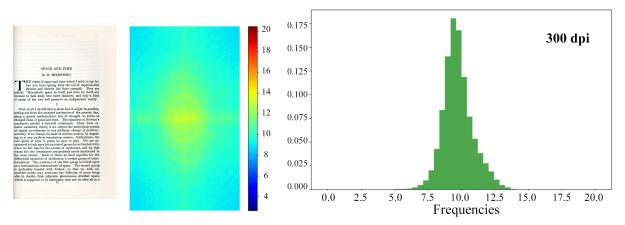
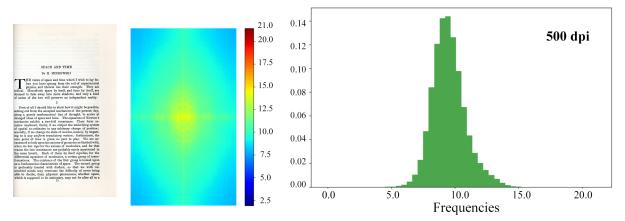


Figure 17 - Image 01, the frequency map and the Spectral Density of 500 dpi resolution.



Source: The author (2024)

Figure 18 – Image 01, the frequency map and the Spectral Density of 700 dpi resolution.

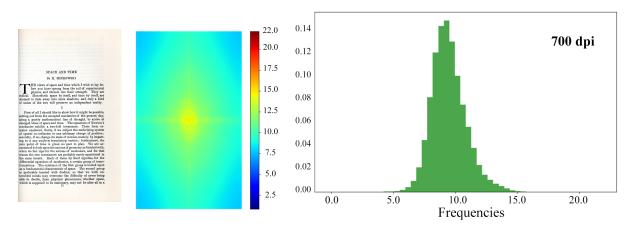
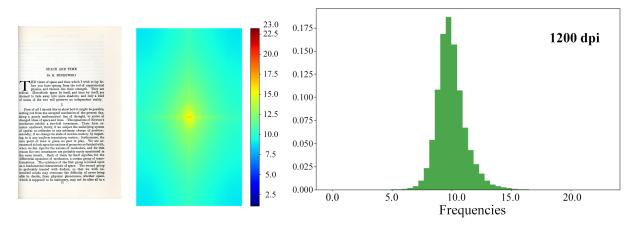


Figure 19 – Image 01, the frequency map and the Spectral Density of 1200 dpi resolution.



3.5 OPTICAL CHARACTER RECOGNITION (OCR)

According to Puglia et al. (2004), digital images must meet quality standards to ensure that OCR conversion achieves a certain level of accuracy. Barney-Smith and Andersen (2005) highlights that specific combinations of parameters governing scanning degradations can cause characters to appear either similar or significantly different when degraded. Beyond the influence of the optical resolution of the scanner on the document quality, factors such as contamination, manual underlining, or annotations can also pose challenges and interfere with the OCR process, as detailed by Forschungsgemeinschaft (2016).

Given the impact of document resolution on transcription accuracy, text-based documents were transcribed using optical character recognition (OCR), and the Levenshtein Distance was calculated for each resulting file. The implemented algorithm is presented in Appendix D.

The Source Code 5 begins by selecting the image to be processed and defining its path. After verifying the existence of the image according to the specified path, the function "cv2.imread(path)" loads the images and displays an error message if the image is not located. For the OCR process, Python-tesseract, a Python wrapper for the Google Tesseract-OCR Engine, was used, as referenced in PyPi (2024). Specifically, Tesseract OCR version 4.1.1 was applied and accessed by the pytesseract library (version 0.3.13) within the Python environment. The algorithm was executed without using the language dictionary or loading frequent words, this approach eliminates the influence of predefined words on OCRbased word recognition, prioritizing character-level identification for evaluating text element recognition. Based on doxygen (2015) and tesseract-ocr (2024), these settings were defined in line 60 of the Source Code 5, using the commands:

```
tesseract_config = --psm 6 load_system_dawg=0 load_freq_dawg=0]
```

where:

- -psm 6: sets the Page Segmentation Mode (PSM) to assume a single uniform block of text;
- load_system_dawg0 load_freq_dawg0: disables the dictionary-based correction to allow a raw text extraction without auto-correction.

The OCR transcription of each document was saved in the correspondent TXT file format and the Python-Levenshtein library was employed by applying the code "distance = levenshtein(reference_text, ocr_text)" to calculate the Levenshtein distance between each transcribed text and the corresponding ground-truth text as described in PyPi (2024), the LD results were then saved in an Excel file and analyzed considering the OCR accuracy at the different image resolutions, based on measuring how many character changes (insertions, deletions, substitutions) were needed.

4 RESULTS

The data collected using the scanner Epson L4160 for the color and binary versions of the documents are summarized in four tables. Tables 4 and 5 contain data from the Text Dataset, while Tables 7 and 8 contain information from the General Dataset.

4.1 TEXT DATASET

Table 4 contains detailed information for the 18 images of the text dataset considering the colored images scanned with the Epson L4160 scanner. The information includes image size in centimeters (cm), text size in points (pt), resolution in dpi, file size for both PNG and Windows Bitmap (BMP) formats in megabytes (MB), the Highest frequency (HF) and Levenshtein distance measurements for each document across the seven previously defined resolution levels.

Table 4 – Text Dataset - Original images results (Image size in centimeters (cm), Font size in points (pt), dpi - image resolution in dots per inch, png/bmp - file size of the png/bmp image in megabytes (MB), HF - Highest frequency, LD - Levenshtein distance.)

| Image | dpi | png | bmp | HF | LD | Image | png | bmp | HF | LD |
|---|------|-------|--------|----|-----|---|-------|--------|----|-----|
| CHITCHE STATE (CHITCHE STATE | 75 | 0.35 | 0.59 | 18 | 97 | ingenerating a fermionic characterist, Where the notion of sameness conten in it has the fermion you end up with his the time reason and charge as the long-top was started with. This | 0.25 | 0.6 | 18 | 101 |
| SPACE AND THE BW IL MINISTOWELL | 150 | 1.28 | 2.37 | 19 | 12 | brings to an one of the insent distinctive predictions of super- pressments; in simplests that for every bosons, there's a femines with the same must and slorge, and view verse. One thing we know the reast which there would had particully superpressments: If there were a boson with the same reast pressments in the same way to know which has made reast things, it would stoodly after their structure of the same. What | 0.75 | 2.4 | 19 | 11 |
| THE trime of space and disse which it wish to buy he. Every to keep regard from the self of expressions in Express, and thereof in the last strategit. They are decreased to the expression attention that they are decreased to the expression attents tablecome, and other to little and of earlier of the tree will preserve to independent reality. They are all if handles to have well the independent reality. | 200 | 2.28 | 4.21 | 20 | 22 | regist be bapeous in then oper-memorary in Two-loan," or visitually, you were mechanism or induce to subsyme conferen- tion. If the idea of a range over operatory that such valley a recovery stem reading you feel the french vanishe and account somewhere. I show blasse you like reads of range account somewhere. I show blasse you like reads of range account somewhere. I show blasse you like reads of range account somewhere is the continued belong with the first recover with correlational feel with | 1.21 | 4.28 | 20 | 11 |
| First of all Lord like to show here is algob be qualified, using and four the assemble conduction of the greath day, along a person temberanteel line of though, it is strive at alongs allows of apara and low. The apparatum of Severita's production exhibit a two-field investmen. Their facts are consistent of the contract of the contract of the con- ervation of the contract of the contract of the con- ervation of the contract of the contract of the con- tract of the contract of the contract of the contract of the contract of the contract of the contract of the con- tract of the contract of the contract of the contract of the product of the contract of the contract of the contract of the product of the contract | 300 | 5.12 | 9.47 | 20 | 21 | If the weited below of superviscentry and furnissing classification between using histories and to I-IOC, it will be a triamph of pace reason beyond appelling that happened in our weitherness. A low of people ready have their happened in our weitherness. A low of people ready have their happened in our weitherness. A low of people ready have their happened as it is the weiting seen which is no. Supervisoration or down, it is not in the people of their low | 2.39 | 9.61 | 21 | 9 |
| manament to look upon the actions of geometry as furtherd with, when we feel rep for the seriment of sectionaries, and for that manner the two formationes are probably manly mentioned in the numb tends. Duch of them by their seguidant, for the probable of the section of the section of the feel fermalism. The existence of the first going in those of an a fundamental characteristic of open. The second group as a fundamental characteristic of open. The second group | 500 | 13.74 | 26.27 | 21 | 22 | The theory of encepthing—maple Here is a sprought of the cascustical lebes about how sating thousy describes the soil would. Evilage theory states can though the soil of the cascustical lebes about how sating the soil of the cascustical lebes are about the cascustical streng development, so there are now a deliberated described. | 5.74 | 26.69 | 22 | 12 |
| in problemally formats with contain, as that or write the transport of the contained of the contained of the contained of the table to decide, from plypoint diplementaries, whether speed, which is expressed to be stationed, easy, easy and he after all fit is which is expressed to be stationed, easy, and he after all fit is the contained of the contained of | 700 | 25.82 | 51.49 | 22 | 24 | dissertion, but left put them saids for a moment. Six of SUPERSYMMETRY AND THE LHC | 9.91 | 52.27 | 22 | 11 |
| 11.89 x 18.95 Font size: 10 | 1200 | 69.81 | 151.27 | 23 | 28 | 11.45 x 21.00 Font size: 10 | 25.18 | 153.59 | 23 | 12 |
| and of the common operation. But, have the course are admitted out at any are of these teams for agreeming agrees in the beliefs. This photocomes, the the proceed after of the course of process or agreement in the course of the finished- ter. The course of the cours | 75 | 0.39 | 0.75 | 18 | 490 | behave light to be the words west to sun. The woods west | 0.32 | 0.48 | 17 | 945 |
| We have accounted the capation meditations of contrasour extrastive is assumption on the product of the extrastive process of the capation of | 150 | 1.29 | 2.98 | 19 | 164 | dark arresoft them. All the uptraced forces were filled with awe and horror. They had never seen such a beaut. The White Knight—their heev, their champion, the convolvence approposit to fight for them against all | 1.18 | 1.92 | 19 | 912 |
| We have been a second and with the property of the control of the | 200 | 2.25 | 5.29 | 20 | 164 | hereres—paped at the terrifying creature, disapped the Vorpal Sword, and Hed into the trues. It said only a moment for panie to seep through the crowd. If the Kinglin was too afroid to hight, then there was no one to defend them all from the | 1.99 | 3.41 | 19 | 909 |
| Patte 16.6 Sponsoon-Summer Studies," Anne Studen sement Sullines as bedan | 300 | 4.51 | 11.9 | 21 | 170 | Jubberweckly. Scenam hroke out as receptor ran for his life, pushing and transpling arrowne in the way. Fire streaming from the Jubberwoody's needsh blased over hater heads. The Queen's house rearred, and the Queen lost | 4.76 | 7.66 | 20 | 912 |
| Grant Swadmind Person Personaged Sociality Spirity of S | 500 | 12.12 | 33.03 | 22 | 172 | her crown. The Hatter run forward to grab the retus. Issuing his hat in the process. He led her to safety, but for one reservent he leaked bath. | 12.2 | 21.26 | 21 | 914 |
| 14.19 x 21.00 | 700 | 21.71 | 64.75 | 22 | 178 | 11.62 x 16.56 | 23.02 | 41.67 | 22 | 916 |
| Font size: 10 | 1200 | 54.9 | 190.23 | 23 | 180 | Font size: 12 | 59.57 | 122.44 | 23 | 913 |

| Image | dpi | png | bmp | HF | LD | Image | png | bmp | HF | LD |
|--|------------|-------|--------|----|------|---|-------|--------|----|------|
| Fits bell (for only resum to play a grow with a system, of consen. • Passine Passine is a very removed freign ab or held general fooder have been for all resulted files and also say been after death arms of the state of the s | 75 | 0.42 | 0.87 | 18 | 28 | 5 Canonical transformations; Poisson brackets | 0.27 | 0.68 | 18 | 196 |
| chillinguigi table, over and over 3 to very fair or yet, the context are more server genes as "Videliners procession and heave' that mid-the over context." * Indiatation, Although mediatation to pump it validate in the sciences, flower or common of works with the result in various that was made that works in the context in the result is various allow are sent in the various of the various and vario | 150 | 1.44 | 3.49 | 19 | 14 | hab ann a ba'r | 0.87 | 2.71 | 19 | 6 |
| Many States, Section of the Control of Controls, a piper cost, allows controls, and are never for of owns after the Cost, a piper piper might be controlled to the Cost of th | 200 | 2.54 | 6.2 | 20 | 14 | an environment vergen Per (initializing (initivated in such invasion). Date data between the contract of the | 1.44 | 4.82 | 20 | 6 |
| skallinging like fictions of the Thirty proof encoper, sends has a familiar well- lowed blood or filling where where a opportune of the real enter. There is working everage with using a pance or a gave review for where proposed. But many other things would also were. So there appeals there of me appeals we wak make passes usings on all special. 3 young has preconciling appeal cannot small actions. Cannot small shoot bessary or | 300 | 5.47 | 13.94 | 21 | 18 | It is married to the contraction of the Lagrange requirement and the contract of contraction of the Lagrange requirement and the contract of contraction of the Lagrange requirement and the contraction of | 2.84 | 10.85 | 21 | 9 |
| yili, a simerikini (incredely vishedo. Troi i short harring it is omer when there is no present from consequence, and that is why game matter. | 500 | 14.39 | 38.69 | 22 | 19 | for equations of abouts, its agreement is all supported in the contract of the requirement of the contract of | 6.94 | 30.1 | 22 | 8 |
| 15.68 x 22.28 | 700 | 26.06 | 75.8 | 23 | 20 | 12.94 x 21.00 | 11.95 | 58.98 | 22 | 9 |
| Font size: 12 | 1200 | 66.77 | 222.75 | 24 | 19 | Font size: 11 | 30.5 | 173.31 | 23 | 12 |
| N. BRIEF AND QUARTETY STATES 55 The regulation by pulling the cigarest smoke time the installing state of a tage Death store in one you. Me has been a few to be | 75 | 0.35 | 0.58 | 18 | 36 | mented dis older, The "plant fast travir our severally, blace distributions are not early sent in a position attention and the second plant distribution was showned as the second plant distribution and the second plant distribution and the second plant distribution are not extra travition and the second plant distribution and the second plant distribution are neveral travition and the second plant distribution are never as the second plant distribution are never as the second plant distribution are not as | 0.56 | 0.84 | 18 | 3071 |
| states. Contact Conv. the Scientific Discount of Hillips Radio Conveyory, but an apprehensive with him to the afternoon. Before the meeting Can (a discounting of the afternoon. Before the meeting Can (a discounting of burboultage to have his laster out and discounting of burboultage to have his laster out and discounting contact on his clear the barbor was making lattice into the blood hast first hard to report for two bown before the blood hast first hard to report for two bown before the | 150 | 1.24 | 2.31 | 19 | 7 | the form of garderical earse. But when are: In the control of garderical earse. But when are: In the control of garderical earse. But when are: In the control of garderical earse. But we can be control of garderical earse. In the control of garderical earse. But we control of garderical earse. In the control of garderical earse. But we control of garderical earse. In the control of garderical earse. But we control of garderical earse. In the control of garderical earse. But we control of garderical earse. In the control of garderical earse. But we control of garderical earse. In the control of garderical earse. But we control of garderical earse. In the control of garderical earse. But we control of garderical earse. In the contro | 1.96 | 3.35 | 19 | 3110 |
| the med of the bottom distipant, And, of course, no- body would see so well Thermithe all to blind Bob in the course of the course of the course of the course to be a more relay of before all among crientine, the Balpalaneer/Fame (or in sepreluber of length or spec- tation in the Approach or this book (). Hamplers placed from it the Approach or this book (). Hamplers placed by the design and a value of Collection (in the arterior). | 200 | 2.14 | 4.11 | 20 | 7 | beginn, frange, an own from dash for filled to the property of the filled control of th | 3.29 | 5.97 | 20 | 3104 |
| The personality of Nails fields and the plantaries of fife and work in the finite one set of terms in an enter- tion of the contract of the contract of the three personal associator or no still give associated the flower of that merit menerals the same. After posting the comprehensive examination are the analysis of the contract of the contract of the contract of the analysis of the contract of the contract of the contract of the analysis of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the analysis of the contract of | 300 | 4.49 | 9.26 | 20 | 7 | control, we accounted the better of pictures and the control and c | 6.92 | 13.4 | 21 | 3205 |
| speed for commoths attending assumes subset as the Uni- versity of Uniform, of the filt much be the Conf. "Bubble of the Uniform of the Uniform of Uniform of Uniform to be bordle to cash other, that our cleanings all is device Resoluted for the professor of principle and the part of the Uniform of Uniform of principle and the part of Resolute mobile and the Uniform of Uniform Open Commondations when we would | 500 | 11.81 | 25.7 | 21 | 7 | attent on the state of the control o | 17.68 | 37.24 | 22 | 3211 |
| 11.04 x 21.00 | 700 | 21.68 | 50.35 | 22 | 7 | 15.10 x 22.28 | 31.94 | 72.95 | 22 | 3201 |
| Font size: 12 | 1200 | 56.13 | 147.98 | 23 | 12 | Font size: 10 | 81.99 | 214.32 | 24 | 3205 |
| Valoritate the reserve | 75 | 0.34 | 0.95 | 18 | 66 | THE VERSE OF TASK! THE VE | 0.42 | 0.59 | 18 | 59 |
| "Multitude grants the Commons and "who before were on, mining assuming institute and the world drug or how the Demonsor desired point of the set the world drug or how the Demonsor desired unbulge, being the salese, "Be for dark mining the listent," are none more brand and however." But which click the Demonsors any?" one of the juny since. That I can't memoritee and the Maltine. That I can't memoritee and the Maltine. | 150 | 1.06 | 3.81 | 20 | 23 | light of the new rold within some have one them to securously, be the other few years graph read and finish of Heroret so- conducted in distincting the similar change which counts when light and the second of the second second the second second second for the second second second second second second second to first plan second second second second second second to find the second sec | 1.43 | 2.37 | 19 | 1 |
| The materians between when the first, but The way our statement of the materials between the regions in the stress and section attention and section according to the stress and section according to the Third process, the stress of the stress and the stress according to the stress of the stress according to the stress | 200 | 1.86 | 6.77 | 20 | 23 | effort, Pound and Rebba were able to measure the frastional charges in the process, which to the theoretic affiliate in the issue and than the office opposed in light contrapt grows the Value. The contract of the contract grows the Value of the Contract the basical good light from a distinct user as it passes colour to the basic of viewment and the contract can be a long to passes colour to the San. In its possible to drives an effort of this sort on the basic of viewments arthur, but the transling alongs in the | 2.46 | 4.21 | 20 | 1 |
| I will gave regulate to you how it was done. They had a large careers hap, which find up on the means when showers passed that theiry allipped the gainst-pick hand from early from your ha). "I've did I've seen of the early of high. How I've we askess read in also assumpposes, at the early of right, I have I've we some strong at applicars, which was careful might proposed by the officers of the | 300 | 3.9 | 15.23 | 21 | 25 | appared ashedood or the data is bested in prince of infiniterior data of the control of the con | 5.36 | 9.47 | 20 | 1 |
| opprises, deside was internability (applicated by the doctors or the opportunity of the form of | 500 | 10.07 | 42.31 | 22 | 24 | for present to compare their pre-leave from which their production single. The first own arrange to our Enterthic prediction was raised by thirds expedition the left of N other A findings in our the left production of the left of the left of the left of the left of the strength of the left of the left of the left of the left of the extending the left of the left of the left of the left of the left of the left of the left of the left of the left of the left of the left of the left of the left of the left of the left of left of the left of left of the left of the left of the left of the le | 14.11 | 26.27 | 21 | 1 |
| 16.29 x 23.47 | 700 | 18.12 | 82.89 | 23 | 23 | at later edges august the Eddington underredinated the | 25.97 | 51.49 | 22 | 2 |
| Font size: 12 | 1200 | 46.59 | 243.54 | 24 | 27 | 11.89 x 19.95 Font size: 12 | 68.84 | 151.27 | 23 | 4 |
| SANTO EZEQUIEL MORENO Santo Forquid Memor assert car Alliar - Forqui Espenible, (to-the 9 des shell de 1008, No. 100. | 75 | 0.15 | 0.25 | 17 | 1638 | 1-12 Press te pliry or presse. Presione gars reproduct o pausar. | 0.05 | 0.18 | 17 | 442 |
| Agentinus-Boolean, et Montaquid-Observit, promissional palera, natable e desblortin pas sod e volt, Cronia even contine nu et anche e desblortin pas sod e volt, Cronia even contine nu et anche e volta e participato de la contine e contine nu et en estimatio, par 11 mars. En 1162, A manus acception, como enticipato, par 13 mars. En 1162, A mars acception de Montaquis-Deep not Coloniale en 1808, del montare o Debres due de mars de la contine de la contine de la contine de la contine de la contine de la contine de la puezo e plan midiation. E algir se plus este desposibles, la prop- grance e la contine de la conti | 150 | 0.53 | 1.01 | 18 | 55 | President para oprasdatir op pouse. Double-press to skip forward. President dost vector para process. President dost vector para process. | 0.18 | 0.71 | 18 | 48 |
| State a l'opinité soite insurer une complétion du sur sincer en marie, un combination de la comme de la comme de la comme aux Superson combination collemant admirent de la comme del la comme de la comme del la comme de la comme del comme del la comme del la comme del la comme del la comme de | 200 | 0.87 | 1.8 | 19 | 46 | ** * * The areas to skip back. ** Problem areas to skip back. ** Problem area veces para retroceter. ** Prosider this water. ** Prosider this water. | 0.33 | 1.26 | 19 | 47 |
| Mercen 1. St bandlende pit de Bright Ball V beide C All Services Despuid (2005). A little of the Mark Ball Services Despuid Ball V beide C All Securities of 120 ft de Julius C 1000, Marian de Jones Votte de Dies, de Brience Des 10 ft de Julius C 1000, Marian de Jones Votte de Dies, de Brience de Dieses L'acqueil de Proposition de La companyation de la companyation de Dieses L'acqueil de Proposition de La companyation | 300 | 1.87 | 4.05 | 20 | 44 | Side Stay Thing Shift to exclivate Side. Of York Shift para exclive Skid. Organ for all Shift para eviders. Skid. | 0.74 | 2.83 | 19 | 39 |
| Dangskaupte de Antifrix. | 500 | 4.62 | 11.25 | 21 | 45 | | 2 | 7.85 | 20 | 42 |
| 8.57 x 11.89 | 700 | 8.41 | 22.04 | 21 | 43 | 7.96 x 8.94 | 3.85 | 15.39 | 21 | 39 |
| Font size: 7 | 1200 | 22.18 | 64.78 | 22 | 45 | Font size: 7 | 10.33 | 45.2 | 22 | 44 |
| Song controls Tap an album cover and swipe left for Repeat, Genius, and Shuffle song controls. Double-click | 75 | 0.03 | 0.05 | 15 | 210 | Controles de música Toque na capa de um álbum e passe o dedo para a esquerda para que apareçam os controles musicasis Repetit, Genius e Headroto. Clique dusa vezes no botão | 0.03 | 0.07 | 15 | 279 |
| the Sleep/Wake button to play the next track. | 150 | 0.11 | 0.21 | 16 | 56 | Repetir, Genius e Aleatório. Clique duas vezes no botalo Repouso/Despertar para reproduzir a próxima faixa. | 0.12 | 0.27 | 17 | 71 |
| Scorge if A More (10h Ave. | 200 | 0.18 | 0.37 | 17 | 52 | Some of Administration | 0.22 | 0.47 | 17 | 62 |
| GH. | 300 | 0.43 | 0.84 | 18 | 39 | | 0.5 | 1.07 | 18 | 32 |
| | 500 | 1.12 | 2.31 | 19 | 20 | | 1.3 | 2.95 | 19 | 25 |
| 4.71 x 4.47 Font size: 5 | 700 | 2.13 | 4.53 | 20 | 29 | 5.01 x 5.35 Font size: 5 | 2.5 | 5.79 | 20 | 37 |
| | 1200 | 5.69 | 13.32 | 21 | 49 | rom size. 3 | 6.66 | 17 | 21 | 54 |

| Image | dpi | png | bmp | HF | LD | Image | png | bmp | HF | LD |
|--|------|-------|-------|----|------|--|-------|--------|----|------|
| Controles de canción Pulse la portada de un álbum y deslicese hacia la izquierda para que aparezcan los controles la izquierda para que aparezcan la descripción. Haga | 75 | 0.03 | 0.06 | 15 | 279 | | 0.03 | 0.07 | 16 | 786 |
| la izquierda para que apaiezcar de la carción Repetición, Genius y Aleatorio. Haga doble clic en el botón de reposo/activación para reproducir la pista siguiente. | 150 | 0.12 | 0.25 | 17 | 71 | Informações Importantes Sobre o Uso do IPod para Consumidores Brasileiros O aplicativo Tieras e Utilidad posa transferir másicas, video e fotos para o Pod, e para atualizar e restanzar o IPod. O Illunes não está incluido meste produta, nas voço pode fazar o | 0.12 | 0.27 | 17 | 282 |
| Serge A More (In Are) | 200 | 0.21 | 0.45 | 17 | 62 | download no site wawappic.com/bultaness. A Apple fornace atawalkappics on fine ne recomenda que sook masterha o seu software filmes sempre atawalkappic. Functionamento a 4-mazenamento do iPod O iPod cordetin componentes sendivisi, tala como uma tela CCD e uma membrio filast, que na podeme state systémica se | 0.2 | 0.47 | 17 | 41 |
| | 300 | 0.5 | 1 | 18 | 32 | quedis ou a pressão excession. Tão circunstância podem alterar ou reduzir delempenho pear de producto. Utilize o iPod em lugares onde a temperatura estaja entre 0° e 35° C. Fora de tais limites o desempenho do iPod pode ser afestado. | 0.44 | 1.06 | 18 | 24 |
| | 500 | 1.3 | 2.78 | 19 | 25 | | 1.11 | 2.94 | 19 | 20 |
| 4.71 x 5.38 | 700 | 2.45 | 5.44 | 20 | 37 | 5.59 x 4.78 | 2.04 | 5.76 | 20 | 21 |
| Font size: 5 | 1200 | 6.51 | 15.99 | 21 | 54 | Font size: 4 | 5.33 | 16.91 | 21 | 22 |
| Bully and handling special arrays copies on the original barys copies or the original barys copies or the original barys copies and the original barys copies are the original barys copies and the original barys copie | 75 | 0.1 | 0.18 | 17 | 2586 | Bernalest | 0.87 | 1.14 | 18 | 4049 |
| Interior with an and are self-clief. The leading benefit and a self-clief. It is to exceed with a self-clief. It is controlled to the leading benefit and controlled to the leading benefit and controlled to the leading benefit and controlled to the leading benefit and properties of the leading benefit and controlled to th | 150 | 0.4 | 0.74 | 18 | 1906 | mellines in the product of the second of the | 3.04 | 4.54 | 20 | 4152 |
| Marketine (marketine) in support of Architecture and Architecture (marketine) in support of Arch | 200 | 0.66 | 1.31 | 19 | 1960 | the could be comediated by the companion of the companion | 5.12 | 8.09 | 20 | 4158 |
| Willing as on the extension table (2000) and settlement that (2000) and the extension table (2000) and the extension of the e | 300 | 1.44 | 2.94 | 19 | 1978 | prime verbitter og dettere sprande villed de desiliarde de silve de verbitter (verbitter de silver de desiliarde de silver de verbitter de verbitter (verbitter de desiliarde de silver de verbitter de verbitter de de silver de verbitter de verbitter de verbitter de de silver de verbitter de verbitter de verbitter de verbitter de de verbitter de verbitte | 10.94 | 18.18 | 21 | 4153 |
| Display, Gloves swile a written sharper, well when you have swile a written sharper, and | 500 | 3.57 | 8.16 | 20 | 1982 | A contract of the contract of | 27.47 | 50.5 | 22 | 4155 |
| 8.57 x 8.64 Font size: 5 | 700 | 6.43 | 15.99 | 21 | 1976 | 18.05 x 25.26 | 50.1 | 98.93 | 23 | 4158 |
| FUIL SIZE. 3 | 1200 | 16.94 | 46.97 | 22 | 1978 | Font size: 8 | 130.4 | 290.75 | 24 | 4169 |

Table 5 shows data for the 18 images in the Text Dataset considering the binarized images scanned with the Epson L4160 scanner. The information includes image size in cm, text size in pt, resolution in dpi, file size for PNG, BMP and Tagged Image Format (TIFF) formats in MB, the highest frequency and Levenshtein distance measurements for each document across the seven previously defined resolution levels.

Table 5 – Text Dataset - Binarized images results (Image size in centimeters (cm), Font size in points (pt), dpi - image resolution in dots per inch, png/bmp/tiff - file size of the png/bmp/tiff image in megabytes (MB), HF - Highest frequency, LD - Levenshtein distance).

| Image | dpi | png | bmp | tiff | HF | LD | Image | png | bmp | tiff | HF | LD |
|---|------|------|-------|------|----|-----|--|------|-------|------|----|-----|
| | 75 | 0.01 | 0.2 | 0.01 | 18 | 229 | to be presenting a fermionic dimensionic. Where the motion of successes causes in it is had the fermion you and up with har | 0.01 | 0.2 | 0.01 | 18 | 264 |
| SPACE AND TIME 30 IL MEMOROWSKI | 150 | 0.04 | 0.79 | 0.02 | 19 | 16 | the curse was and change in the boose you cannel with. This being as in one of the new distinctive prelations on deper- preenersy; it implies for the every boson, their is firmine with the varies was and allowing. As with very ress. One change whose for early their the world in particul- query quarties. If these were a financy with the varies was and change is an observat, well samely know about a few one about it would standly about the resource of the same. What | 0.04 | 8.0 | 0.02 | 19 | 12 |
| THE view of spore and time which I with to by the flow you have going from the sol of everywherein adjustment of the everywherein adjustment and therein has taken strongers. They are not assume that the property of the solid country of the solid country in the document to fall as well you have above, on the only a birst of union of the two will passers an independent tradity. But of all Exheld Risk to gain by the 'purity being the presenting. | 200 | 0.06 | 1.4 | 0.02 | 20 | 12 | might be hopposing in that supersymmetry in "beslow." or violated, by some mediations insulate as taskym conferen- tion. If the left of a tonoge new pyrametry that into really a ryammary warm analogy good for like them's contin and metrous conceived. I don't bleam you. Like mode of arrang and the content of the content of the content of the first content with a postlement devices. | 0.06 | 1.43 | 0.02 | 20 | 14 |
| meting out book and a manipul annihilation of the present day, shour a purely unabsensable live of themptic, as arrive at the contract of the contract of the contract of the methods excluded a two-field invariantee. Their form re- mains continued, they of the subjects the methoding param- ter of the contract of the contract of the contract of the secondary of the contract of the contract of the contract annually, if we change that state of produce, annually by impor- ting to a tray may form resolutionsy under a formation of the contract of the contract of the contract of the contract of the produce of the contract of the contract of the contract of the secondary of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract | 300 | 0.12 | 3.16 | 0.03 | 20 | 10 | If the varied their of separaparametry and fermionic di- mensions are been control polarization at the IEE, and the a triangle of pare means beyond sophising that happened near this terminan, he of people right have that began an interest the second south the second source of the second soor approximant from, or it bink. Plankly, I'll be surpticed sides may. | 0.11 | 3.21 | 0.03 | 21 | 13 |
| maximum to toke upon the actions of geometry as finished web, when me first type for the actions of authorities, also for that it when me first type for the content of authorities, also for that the action broads. This of them by forest significant, for the differential equations of authorities, a varieties paper in trans- cript and action of the content of the content of the as a fundamental characteristic of space. The record group is preferrably broaders with Adults, in that in our with the | 500 | 0.25 | 8.77 | 0.06 | 21 | 12 | The timing of overgining—maple. Here is a symposis of the canonical ideas about how storing sheery describes the real vesseld. Being there years not out with raw distances of treatment, for taking about super-sizing below, when, to be the extense followed fermionic and the contract of | 0.24 | 8.9 | 0.05 | 22 | 12 |
| which is deather, from physical potentiares in, whether spoon, which is supposed to be statinger, may not be after all it is 111.89×18.95 | 700 | 0.38 | 17.16 | 0.08 | 22 | 14 | 11.45 x 21.00 | 0.36 | 17.42 | 0.08 | 22 | 11 |
| Font size: 10 | 1200 | 0.73 | 50.42 | 0.15 | 23 | 15 | Font size: 10 | 0.68 | 51.22 | 0.14 | 23 | 12 |

| Image | dpi | png | bmp | tiff | HF | LD | Image | png | bmp | tiff | HF | LD |
|--|------|-------|-------|-------|----|------|--|-------|-------|-------|----|------|
| ONIOS, RIGIDAM under the convery specials. This can do going as already state only one of laws seen, the converse year is to believe. The phinocensis, fact, and the converse is to believe to the phinocensis. See | 75 | 0.02 | 0.25 | 0.01 | 18 | 839 | bhipes light | 0.01 | 0.16 | 0.01 | 17 | 983 |
| the pound clade of the openas does not present the removely of the Hamiltonian is under "postational groundsys buildings". In the "postational groundsys buildings" is provided by the postation of the postation | 150 | 0.05 | 0.99 | 0.03 | 19 | 172 | leathery wings blocked out the sun. The woods went dark accused these. All the spanned faces were filled with uses and hereur. They had sower seen such a beast. The White Keight-their here, their champion. | 0.03 | 0.64 | 0.01 | 19 | 912 |
| shalled greats, interesting consequences them; in this case, where it is mer- countable beliefung or ground stars price mercany stars or two more infragence to contract the start of great stars are the star of the star of the department, then energing a type of motion dues to which the load ground stars that the start of the star of the star of the star of the star of the star weetength, that is sear in relating the star of the star of the star of the weetength and the star of the star of the star of the star of the star of several stars are the star of the star of the star of the star of the weetength and the star of the star of the star of the star of the weetength of the star of weetength of the star of "Outdoorness these." | 200 | 0.09 | 1.77 | 0.03 | 20 | 164 | the one who was supposed to fight for them against all hererse-speed at the stretching creature, despeed the Vorpal Sword, and fled into the trees. It took only a moment for panic to seep through the crewd. If the Knight was too alraid to fight, | 0.04 | 1.14 | 0.01 | 19 | 914 |
| Tradic companies of Coldenne contration include phonous in a crystal hairs, which opening the best and contration of the contration of the contration of his contration of the contration of the contration of the contration of his prompanies, which reduced not be the contration of the contration of his prompanies, which contration of the contration of the contration of makes the department of the contration of the contration of Table 4.8 I formation of the contration of formation of the contration of the contration of formation of the contration of the contration of the contration of formation of the contration of the contration of the contration of formation of the contration of | 300 | 0.17 | 3.97 | 0.05 | 21 | 166 | then there was no one to defend there all from the jabbreworky. Screams broke out as everyone the for his life, pushing and transpiling arrows in the way. The streaming from the jabbreworky's mouth bland over their heads. The Queen's harme rearred, and the Queen lost | 0.08 | 2.55 | 0.02 | 20 | 911 |
| Cycled Traditional Priese Framespare Sendend Upda to use Superfield Chief gang Priese Superfield Chief gang Priese Superfield Chief gang Chief gang Chief Superfield Chief gang Chief Superfield Chief Andreas Superfield Chief Andreas Superfield Chief Superfield | 500 | 0.36 | 11.01 | 0.08 | 22 | 169 | The Quesa's harse rearred, and the Queen last her crows. The Hatter run forward to guide the retire, lessing his has in the process. He do he to adopt, but for our rooment he locked hack. 700 | 0.16 | 7.09 | 0.04 | 21 | 912 |
| 14.19 x 21.00 | 700 | 0.54 | 21.6 | 0.12 | 22 | 179 | 11.62 x 16.56 | 0.25 | 13.89 | 0.05 | 22 | 917 |
| Font size: 10 | 1200 | 1.02 | 63.41 | 0.22 | 23 | 178 | Font size: 12 | 0.48 | 40.81 | 0.1 | 23 | 915 |
| For soil the only reseat to give a game with a system, of creatur. Presides Preside its very company distingts also with games Studies. | 75 | 0.02 | 0.29 | 0.01 | 18 | 66 | 5 | 0.01 | 0.23 | 0.01 | 18 | 395 |
| President President in very common design in the subspaces Statistics have shown that assumed for all cast seals may be received of their his treat "deliberate president," where the general practicing fells hierard for challenging tasks, over on the "Dis very had not be a considerated for challenging tasks, over on the "Dis very had not when it was more more more on" deliberate practice mediated "that make the vert seals. Indicates a challenging tentation in populy regard the developes, there are common decreases their resure in revieway aprecision amount due would asset gain participation present deliberated states; prices depict all present plants of interest and the action of the control between the control tentance and the control | 150 | 0.05 | 1.17 | 0.02 | 19 | 14 | Canonical transformations; Poisson brackets | 0.04 | 0.9 | 0.02 | 19 | 10 |
| Support Ballage from games control the morths and of coverns a plant rate. Support Ballage from games control the morths and of coverns a plant rate over you like some assessment in these southers than 10 to game agreem which also counterprofess. We can desire that have been than 10 to game for corn it some drawn about the southers than 10 to game for corn it some drawn about the southers than 10 to game that the southers than 10 to game the sout | 200 | 0.08 | 2.07 | 0.03 | 20 | 14 | In this chapter we high by considering contented has glovestern. Then are vanishmenton that present the form of Fandwick spatiests. This is full- lowed by a single of Persiden Pathons, in algorates that for subship contented transferenties. Pathol we recently administrated stream formation and, is an exception, we had at suppler momentum in terms of Persiden Studies. | 0.06 | 1.61 | 0.02 | 20 | 9 |
| There is nothing wrong with using a game or a game system for these purposed. But many other things would also serve. So these qualities all not speak to what makes games unique and speaks. To come the reconcilient many favours sently stories. Games sently about housets or | 300 | 0.16 | 4.65 | 0.04 | 21 | 21 | 8.1 Integrating the equations of metions in our other) of marjoral mediance was how made for weather grine right better to effective a set of conjustion of metion. The first an contain of the language systems and the months of means of the material variances and to appear our procession of the months of means of the complete and the species and disturbance or quantities are as at of the complete first water differential sequence. The officines good rise of phenoide delayed to be their a general metion for the | 0.12 | 3.62 | 0.03 | 21 | 7 |
| delight Games aren't show) observing the season abuses. Here should not meet engle, as essentialing to assist instancials from a since beneating an accounter bases there is no present these consequence, and that its "vdg games matter." | 500 | 0.31 | 12.9 | 0.07 | 22 | 19 | the equations of motions. In Languagean Cyanatics this equipose integrating the equations of minors with This is desire, and fifthen between the Languagea (so the born the equations of closellar deposits between the colors and the production of the colors deposits but only not so the constraint of the colors for the colors for the colors and the colors of the col | 0.24 | 10.03 | 0.06 | 22 | 9 |
| 15.68 x 22.28 | 700 | 0.47 | 25.28 | 0.1 | 23 | 20 | 12.94 x 21.00 | 0.37 | 19.68 | 0.09 | 22 | 10 |
| Font size: 12 | 1200 | 0.89 | 74.25 | 0.19 | 24 | 20 | Font size: 11 | 0.71 | 57.77 | 0.16 | 23 | 14 |
| E again and oppositive almosts 22 bit modelles for suffer the reference scales bits the | 75 | 0.02 | 0.19 | 0.01 | 18 | 107 | estation of the files for fact instance of the control of the cont | 0.03 | 0.28 | 0.02 | 18 | 3090 |
| looding pass of a large Dotts force in one proce. He described any self-scale before the of their data for data for the described any self-scale before the of their data for data for Redic Corespory. Indeed an apprehension with this to the structures. Notice the contenting Case is alternative of structures. The contenting Case is alternative of structures of the content of the content of the structures of the content of the content of the hadronise on have this hade one and albeit weekes until belief hade. The content of the content of the belief hade. The large of the few locates to belief the belief hade. The large of the few locates to belief the | 150 | 0.05 | 0.77 | 0.02 | 19 | 12 | The second of th | 0.08 | 1.12 | 0.04 | 19 | 3187 |
| appointment with Eleveralite welfaling the alexant to let have been described as the Eleverality of the Eleverality of have been described as the Eleverality of the Child Selb in bottom for weverst than Elysfull after the Eleverality of the Eleverality of the Eleverality of Elegendation of Facility of the Selb Selb Selbson described leading to the depression to the books, Eleverality affects described to the Selbson of Eleverality of Eleverality of the Selbson of Eleverality of by showing him as when of Clarithmen (the association). | 200 | 0.08 | 1.38 | 0.03 | 20 | 7 | stand other, well for some defined in the below with present the control of the control of the below the below in the control of the control of the below in the bel | 0.13 | 1.99 | 0.05 | 20 | 3163 |
| Be and worth his fundame are all same in any remo- eries of the years from 12% to his death, and I hope and the same of the same of the same of the factor of that some mental the same of the factor of that some mental the same After peacing the membranels constitution on the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the same of the space of the same of the same of the same of the same space of the same of the same of the same of the same space of the same of the same of the same of the same space of the same of the same of the same of the same space of the same of the same of the same of the same space of the same of the same of the same of the same space of the same of the same of the same of the same space of the same of the same of the same of the same space of the same of the same of the same of the same space of the same of the same of the same of the same space of the same of the same of the same of the same space of the same of the | 300 | 0.15 | 3.09 | 0.04 | 20 | 7 | Note his would have a variously to find that middly for the control of the contr | 0.24 | 4.47 | 0.07 | 21 | 3204 |
| seden' not 'repérable' s'solone, which an upposed to be borden to est facilité and activate du feieur parties de la principal de la companie de la com- panie de garier persision to exchange so many frontes subte loss en equidant assume of Camesa de la companie de la companie de la companie de la University perfesers, I manugel to obtait a midus | 500 | 0.3 | 8.57 | 0.07 | 21 | 7 | other benimm delated, at the proper forecess, the made person sent the first of the first of an extra section of the first of the contract of the first of the contract of the first of the contract of the co | 0.52 | 12.41 | 0.12 | 22 | 3208 |
| 11.04 x 21.00 | 700 | 0.45 | 16.78 | 0.1 | 22 | 7 | 15.10 x 22.28 | 0.78 | 24.32 | 0.18 | 22 | 3205 |
| Font size: 12 | 1200 | 0.85 | 49.33 | 0.18 | 23 | 12 | Font size: 10 | 1.47 | 71.44 | 0.32 | 24 | 3210 |
| WARR STREET SIZE CARREST | 75 | 0.02 | 0.32 | 0.01 | 18 | 79 | THE YEARS OF FAME motion of light in a growtheriand field. When light haven a massive body is loss energy in distillage out of the body's growinstand field. This loss of energy shows up as a change | 0.02 | 0.2 | 0.01 | 18 | 159 |
| He deview at 2 and 6 for \$100 ferror on the part of "ACLE are ground to Common of the 4 ferror was in adulting attitude proof in one of the action of the 4 ferror was in adulting attitude proof in one of the world along in one for the for Dominion department of the action of the action of the action of the Common Part of the action of the States, "Cod onto a contributed and home- "But with actid did the Dominione Reyff" and who appropriate The action of the | 150 | 0.05 | 1.27 | 0.02 | 20 | 29 | in release the light becomes reddered, and so we speak of the Timutes and helf. Alternate to observe them of his in the light of the seas and section uses here are them to a successful, but show the years ago Portune and Robbin of Hirvent suc- cessful in durinting the volver change which occurs when light fulls from a nevertry-disord-look were to the flow; it this case the light gains ourget by falling and becomes blook, but the olover change is functatefully must for such a short path- | 0.05 | 0.79 | 0.02 | 19 | 2 |
| The mean annual of metabolish fielding in UTILine type assumed." The metabolish fielding object the resump and two-derel between and were down on not know. Our power man, your Helperty has larges. "You'll'en a very power perspective," which is family, Here cone of the gainer-pips already, and was limited mily appropriately by the offices of the court; (of the interfret header word, | 200 | 0.07 | 2.26 | 0.02 | 20 | 20 | And the part of th | 0.09 | 1.4 | 0.03 | 20 | 2 |
| I will just explain a syn have I was from 1997 to be infrom executing which and by a ret to membra where just make having being the problems of the problems o | 300 | 0.12 | 5.09 | 0.03 | 21 | 25 | appeared derotion of the stee is review as gone in Einstein's theory as it is in Norman's. An observational stee of Einstein's prediction in our possible in the ordinary way, boscume when the Sin is in the day his light is to relight any way, boscume when the Sin is in the day his light is to relight any connect some any status at all. However, Zimenic primate one that during a state oldges of the Sin the artists become which, and it should be possible to compare their positions then with their positions some months lates when the same stars would be to while at | 0.17 | 3.16 | 0.04 | 20 | 1 |
| contribution or understand what is measured in lamin. "Theshold you know about it, you may much describ confirmed the Plany W = 100 + + | 500 | 0.25 | 14.11 | 0.05 | 22 | 26 | rapid. The hold alleady to bed Landence y questions was unably be Billeid repositions bed by the Arther Likelingers just at the end of the Great Wine. At that time at was considered very sticking that a purps of Delivial antenomene should attempt to seen a theory proposed by a German. The results of the test were claimed as a temple of Execute, and the sew-delivial facts class from the execut. Institutely comply, attempts made at hear colleges suggest that Attemptor underventured the | 0.33 | 8.77 | 0.07 | 21 | 1 |
| 16.29 x 23.47 | 700 | 0.38 | 27.65 | 0.07 | 23 | 20 | " 11.89 x 19.95 | 0.5 | 17.16 | 0.11 | 22 | 2 |
| Font size: 12 | 1200 | 0.71 | 81.21 | 0.14 | 24 | 22 | Font size: 12 | 0.94 | 50.42 | 0.19 | 23 | 3 |
| SANTO EZEQUIEL MORENO Same Empid berner samer en Aller- Boje- Equato. In die de | 75 | 0.007 | 0.09 | 0.006 | 17 | 1602 | > 11 Press to play or payer. | 0.002 | 0.06 | 0.002 | 17 | 442 |
| Same Derright Memorates and early a liquid for the fine of the collection of the col | 150 | 0.02 | 0.34 | 0.01 | 18 | 107 | Per Persis bright or centure. Persiste para reproduct or passar Persisten para reproduct or opassar opassar. Persisten para reproduct or opassar. Persisten para reproduct or opassar. Persisten des verar para annular. Persisten data severa para annular. Persisten data severa para annular. Persisten data severa. | 0.006 | 0.24 | 0.003 | 18 | 49 |
| sold a Septekt sold near or or security of the photolocis. In 2006, to the Septekt sold near or or security of the photolocis. In 2006, crown on Memorgalic, not be grown of a protest some stopping combination and mandated intermediates for the controllers. So combined to an administration of the controllers of th | 200 | 0.03 | 0.6 | 0.02 | 19 | 63 | 4 Tigle-press to also back. Prolone tres words are responded. President tres words para responded. President vid-words para responded. | 0.009 | 0.42 | 0.004 | 19 | 46 |
| the contract of the conflict o | 300 | 0.06 | 1.36 | 0.02 | 20 | 47 | Birli Bay They Side in activate Side. Gir Yoye Side ones activate Side. Girge "E al Side" para activat a Elis. | 0.02 | 0.94 | 0.005 | 19 | 41 |
| | 500 | 0.14 | 3.76 | 0.04 | 21 | 46 | | 0.04 | 2.62 | 0.009 | 20 | 43 |
| 8.57 x 11.89 Font size: 7 | 700 | 0.23 | 7.36 | 0.05 | 21 | 43 | 7.96 x 8.94 Font size: 7 | 0.06 | 5.13 | 0.01 | 21 | 45 |
| | 1200 | 0.45 | 21.6 | 0.1 | 22 | 47 | | 0.14 | 15.08 | 0.04 | 22 | 58 |

| Image | dpi | png | bmp | tiff | HF | LD | Image | png | bmp | tiff | HF | LD |
|--|------------|--------|-------|--------|----|------|--|--------|-------|--------|----|------|
| Song controls Tap an album cover and swipe left for Repeat. | 75 | 0.001 | 0.02 | 0.0004 | 15 | 210 | Controles de música Toque na capa de um álbum e passe o dedo para a · | 0.0008 | 0.02 | 0.0004 | 15 | 279 |
| Genius, and Shuffle song controls. Double-click the Sleep/Wake button to play the next track. | 150 | 0.003 | 0.07 | 0.001 | 16 | 85 | esquerda para que apareçam os controles musicais Repetir. Genius e Aleatório. Clique duas vezes no bocão Repouso Despertar para reproduzir a próxima faixa. | 0.004 | 0.09 | 0.002 | 17 | 104 |
| Palation Palation Surregion of November (no | 200 | 0.005 | 0.13 | 0.002 | 17 | 53 | Special State of Stat | 0.005 | 0.16 | 0.002 | 17 | 65 |
| | 300 | 0.009 | 0.28 | 0.004 | 18 | 43 | | 0.01 | 0.36 | 0.004 | 18 | 53 |
| () | 500 | 0.02 | 0.77 | 0.008 | 19 | 25 | | 0.02 | 0.99 | 0.009 | 19 | 39 |
| 4.71 x 4.47 Font size: 5 | 700 | 0.04 | 1.51 | 0.01 | 19 | 24 | 5.01 x 5.35 Font size: 5 | 0.04 | 1.93 | 0.01 | 20 | 45 |
| 1 0111 5120. 5 | 1200 | 0.08 | 4.44 | 0.02 | 21 | 49 | TOHE SIZE. 3 | 0.09 | 5.67 | 0.03 | 21 | 48 |
| Controles de canción Puise la portada de un álbum y deslicese hacia | 75 | 0.0008 | 0.02 | 0.0005 | 15 | 294 | | 0.002 | 0.02 | 0.002 | 15 | 783 |
| Pulse la portada de un album y desenso. la siquierda para que apareccan los controles de canción Repetición, Genius y Aleatorio. Haga doble clic en el botón de reposoractivación para reproducir la pista siguiente. | 150 | 0.004 | 0.09 | 0.002 | 17 | 116 | Informações Importantes Sobre o Uso do iPod para Consumidores Brasileiros O aplicativo iTunes é utilizado para transferir másicas, video e fotos para o iPod, e para atualizar e retinatar o iPod O iTunes | 0.005 | 0.09 | 0.004 | 17 | 360 |
| South Collection | 200 | 0.006 | 0.15 | 0.002 | 17 | 69 | O aplication Disnote e dell'application para beneficier musiciare, video priess para se l'indice a poi anche dell'application per mecazione di Bod Chillines nales cella includio monte produtor, mais socio pode fazzo o diversidate o soti se mississippio del production per contrare di que vocite maniferation so mai encorrenta que vocite maniferation so mai encorrenta que vocite maniferation so mai production dell'application della contrare | 0.008 | 0.16 | 0.005 | 17 | 65 |
| | 300 | 0.01 | 0.34 | 0.005 | 18 | 47 | quedas ou a presalo excessivas. Tais circurotálecias podere alteser ou reduzir o dissempenho pesi do produso. Utilizer o 1904 em lugares corde a rempestatas estejas entre (°C ± 5°C C. Fora de tais limitas o desempenho do Pod pode ser alexado. | 0.02 | 0.35 | 0.007 | 18 | 32 |
| | 500 | 0.03 | 0.93 | 0.009 | 19 | 52 | 5.59 x 4.78 | 0.04 | 0.98 | 0.01 | 19 | 22 |
| 4.71 x 5.38 Font size: 5 | 700 | 0.04 | 1.82 | 0.01 | 20 | 48 | Font size: 4 | 0.06 | 1.92 | 0.02 | 20 | 25 |
| 1 0110 3120. 3 | 1200 | 0.09 | 5.33 | 0.03 | 21 | 42 | | 0.13 | 5.64 | 0.03 | 21 | 20 |
| Manage and sending several services as the a shape set sending, transport services. Asserted services or services. | 75 | 0.006 | 0.06 | 0.006 | 16 | 2586 | Because the second of the seco | 0.04 | 0.38 | 0.03 | 18 | 4066 |
| insight of white and date and men. They was a second or a control of the second of the | 150 | 0.02 | 0.25 | 0.01 | 18 | 2325 | makes an antique framework of the antique of the an | 0.11 | 1.52 | 0.05 | 20 | 4152 |
| Extrinsic State of the Control of th | 200 | 0.03 | 0.44 | 0.02 | 18 | 1976 | The residual content of the galacteristic residual content of the gala | 0.18 | 2.7 | 0.07 | 20 | 4153 |
| Might and it as a labor with an empirical way and another than the companion was another than the companion of the companion of the companion of the companion of the companion of the companion of the companion of the companion of the companion of the companion of the companion of the companion of the companion of the companion of the companion of the companion of the companion of the com | 300 | 0.06 | 0.98 | 0.03 | 19 | 1981 | Improve to legal price of the control of the contro | 0.35 | 6.06 | 0.11 | 21 | 4148 |
| concerned deligible control of the c | 500 | 0.13 | 2.72 | 0.04 | 20 | 1983 | common de miner serbero. Serbero se serbero mente a conservar de la conservar | 0.75 | 16.83 | 0.18 | 22 | 4156 |
| 8.57 x 8.64 Font size: 5 | 700 | 0.21 | 5.34 | 0.06 | 21 | 1979 | 18.05 x 25.26 | 1.13 | 32.98 | 0.25 | 23 | 4150 |
| | 1200 | 0.45 | 15.66 | 0.1 | 22 | 1982 | Font size: 8 | 2.15 | 96.93 | 0.46 | 24 | 4162 |

The analysis of the images in the Text Dataset shows that applying a low scan resolution, such as 75 dpi, leads to very poor OCR performance, with minimal document details preserved. In these cases, the Levenshtein distances were significantly higher than those at 150 dpi, highlighting that low resolutions hinder accurate recognition of document features, including finer character details. For binarized images, low resolution further exacerbates this issue, especially in images with small details. For instance, a small image from the dataset is illustrated in Figure 20, measuring 4.71 x 4.47 centimeters and with a font size of 5 pt. The figure shows that during binarization all text content was lost, resulting in an illegible image.

At higher resolutions, when analyzing the density of frequencies exceeding the maximum frequency obtained at 300 dpi, the difference in spectral densities observed approaches the level of scanning noise, with values typically below 0.001. This suggests that increasing the resolution beyond

Original Binarized

Song controls
Top an album cover and swipe left for Repail,
Genius, and Shuffle song controls. Double-click
the Skeep-Wake button to play the next track.

Figure 20 – Small image, measuring 4.71 x 4.47 centimeters, in the resolution of 75 dpi.

300 dpi adds minimal useful detail, as the spectral density for such high frequencies does not significantly contribute to document clarity.

The maximum frequency observed in the 200 to 300 dpi range tends to be consistent, generally falling between 19 and 21. This stability in maximum frequency across these resolutions indicates that further increases in resolution yield diminishing returns in terms of detail capture. Consequently, the maximum frequency difference between 300 dpi and even higher resolutions is minor, with a maximum discrepancy of about three units. Such findings imply that 300 dpi may be an optimal balance point, as higher resolutions lead to larger file sizes without substantial improvements in detail discernment or spectral information.

Table 6 presents the percentage of frequency density exceeding the maximum frequencies obtained at 200 dpi and 300 dpi for the original text images. The results show that the use of different scanners did not result in significant changes in the frequency results, and the observed difference in density values can be considered a scanning error of the device.

Table 6 – Proportion of highest frequency density above the maximum frequency of 200 and 300 dpi in the General dataset - Original images

| Image | dpi | Epson | L4160 | Epson | L8180 | HP M4 | 79FDW |
|---|------|---------|---------|---------|---------|---------|---------|
| | | 200dpi | 300dpi | 200dpi | 300dpi | 200dpi | 300dpi |
| MILITORIO STATEL | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| SPACE AND TIME by M. MINKOPSKI | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| THE views of space and time which I wish to lay he has been an operand from the order experimental to be you have not operand from the order experimental and the state of the | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| atting on from the scropted combasies of the preside day, above a purply combanated like of the chipt, to retrieve at the combanate of the combanate of the combanate of the mendation schilds a twelful invasion. Their from re- mains exalized, forty, if we solyted the analoging apona of updat de-collisions in any othersy closery of president of updat de-collisions in any othersy closery of president in the collision of the collisions of the collisions of the light is a say sufficient resultants president, inflamment, the | 300 | 0.1761% | 0.0000% | 0.1832% | 0.0000% | 0.3225% | 0.0000% |
| ment panis of taxon in gires to just to just, who are the size of the contract | 500 | 0.5880% | 0.1843% | 0.5502% | 0.1455% | - | - |
| terables (risks may resonance the Efficially of terror being able to decide, foreign optional phenomen, whether spows, which is supposed to be intelligent, may not be other all in a supposed to be intelligent, may not be other all in a | 700 | 0.9939% | 0.4210% | 0.8421% | 0.3093% | - | - |
| 11.89 x 18.95 Font size: 10 | 1200 | 1.2763% | 0.7350% | 0.9576% | 0.3971% | 1.7536% | 0.8336% |
| Ingeneering a fermionic discussion). When the serious of sameness course in a fine the fermionic year end-up with his the tase cream and indeed as the house year extend with. This belongs to so out of the most districtive preferring relations of super-presently, it includes that for every house, there's dismiss. | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| an engineer on the view of the control production in suppli- tion in the lease are sent and observation of production of the which is to see are sent of the sent of the control productly observations. One of the control production of the sent control productly operatements: Even over a boose with the same mus- sul-design as an electron, over descript him or down it. For ene- thing, a vessal track, but he or secretars of the same. Only might be happening in that operatements in "Industra" or related, by overal trackly, but the control production of the related, by overal trackly, but the control of the control production. | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| a property starts prairies you find like there's starks and | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| mirror surveivant. Gody blear you, like much of string dever, suppressence you have glob and of recenting whether face content with supersecuring legislate. Mention of the content of the content of the content of the mentions to before out the development in the LECC, as the low averaged by the services become all rapiding days between neutralization. As for all principles and has the proposed neutralization. As for all principles and have the supersection in the self-string, and that the content of the content of the content of the content of the content of the content of the content of the self-string content of the content of the content of the content of the self-string content of the content of the content of the content of the self-string content of the content of the content of the content of the self-string content of the content of the content of the content of the self-string content of the conten | 300 | 0.2400% | 0.0000% | 0.1377% | 0.0000% | 0.1707% | 0.0000% |
| The theory of corpoling—maple. Here is a sproposis of the canonical show shows how string theory formers are red words. Sering theory man our consistency formers are red words, and the consistency of the consistency of the consistency of the consistency formers and the consistency formers are the consistency formers. | 500 | 1.2415% | 0.4823% | 0.6000% | 0.2044% | - | - |
| SUPERSYMMETRY AND THE LHC | 700 | 1.7568% | 0.8456% | 0.8859% | 0.3397% | - | - |
| 11.45 x 21.00 Font size: 10 | 1200 | 1.8204% | 1.0146% | 1.2866% | 0.6932% | 1.7131% | 0.8698% |
| OFFCA_HECKNIN and/or for cymentry centrals. Mr., there the cymen sea abuilty seed in only neer of them to states, the yearned to be broken. This photometer, the life possed sold self of the openion was a price in the present of the first the life possed sold self of the cyment does any prices in present or years. The first the life possed sold self of the cyment does any prices in present or years. The company of the first block The company of the first block The company of the first block The company of the cymen | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| We have accustomed the adoption translations of operations comments as assumed to the contract of the contract | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| uses another. The greats must chose one of them, As a consequence of the diagnosts, their regimes to project of choice into a finish the level ground ease the contract of the contract of the contract of the contract of the workings, their is sent is collequent to any one growed man, which as mergi- age contract, the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract of the contract "Ordalment metals." "Ordalment metals." | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| hancemann, which realizes the last condensal receivable. Because the control of | 300 | 0.0000% | 0.0000% | 0.1362% | 0.0000% | 0.4593% | 0.0000% |
| Type Twenderman Present Femous | 500 | 0.4851% | 0.1886% | 0.5081% | 0.1905% | - | - |
| 14.19 x 21.00 | 700 | 0.8319% | 0.3965% | 0.7758% | 0.3259% | - | - |
| Font size: 10 | 1200 | 1.0326% | 0.7835% | 1.1232% | 0.6761% | 1.8980% | 0.4013% |
| blayide light- leathery wings blocked out the san. The voods west dark around them. All the upstrated faces were filled | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| with ave and horror. They had never seen such a beast. The Whise Knight—their here, their thumpion, the one who was supposed to fight for them against all | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| hereness upped at the territying creature, dispped the Viryal Stown, and Bed into the time. It sook only a moment for paste to seep through the crowd. If the Knight was too straid to hight, then there was no one to defend them all from the | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| Jahhrenoely, Scenario hooks can as everyone no- for his life, pushing and renepiling asynone in the way. First attenuing form the Jahrenovsky's meanth historic over their hoods. The Queen's harvest rearred, and the Queen loss | 300 | 0.1707% | 0.0000% | 0.1447% | 0.0000% | 0.1925% | 0.0000% |
| bee rown. The Haster run forward to grals the reins. Issuing his has to hap process. He led there to askey, but for one manners he looked back. ### 277 | 500 | 0.7210% | 0.1802% | 0.4661% | 0.1687% | - | - |
| 11.62 x 16.56 | 700 | 0.9952% | 0.4215% | 0.7909% | 0.3424% | - | - |
| Font size: 12 | 1200 | 1.1225% | 0.6127% | 0.9789% | 0.5623% | 1.0115% | 0.5206% |
| The half life only receive to play a game with a spriver, of answer. Production Production is very common fining to do with games finded to be able to the contract fining to the contract for the work when their densitiesting field that the story beam of the data to travel "Mallouse paradics," when the privince producing distribution for a contract fining to the contract fining the c | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| Meditation. Although meditation is purely stated in the vatores, there are common elements that even be various positions around the result, using a force object of some nore, undo so a matrix as in resulting pattern, and | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| Suppositing from games come with ments, and of corners a figure cost. Suppositing from games come with ments, and of corners a figure cost and present content of the cont | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| local bakes of this date when a optionized over and over. There is withing remain for his lang gaver or age precision of when proposed for more offered than several date over the date of good in what and and prove the date of the dat | 300 | 0.1576% | 0.0000% | 0.1611% | 0.0000% | 0.1615% | 0.0000% |
| See a | 500 | 0.6866% | 0.2405% | 0.7174% | 0.1990% | - | - |
| 15.68 x 22.28 | 700 | 0.8531% | | 0.9929% | 0.4385% | - | - |
| Font size: 12 | 1200 | 1.0343% | 0.6024% | 1.2005% | 0.7363% | 1.3622% | 0.7875% |

| Image dp | i Epson | L4160 | Epson | L8180 | HP M4 | 79FDW |
|--|------------------|---------|---------|---------|---------|----------|
| | 200dpi | 300dpi | 200dpi | 300dpi | 200dpi | 300dpi |
| Casonical transformations; Peisson brackets | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| In this chapter with the foundation of the chapter | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| 5.3 has provided for equations of mactions to see roll of antiplical collection is volve seen to the collection piles eight indice some different some of equations of about 100 feets or controls of the larger appointment of the collection of control of Tribusher's control equations. Larger's typicates are as and a supplication and south of the collection of the collection of the collection of the collection of the different equations. | 0.0000% | 0.0000% | 0.4237% | 0.0000% | 0.1810% | 0.0000% |
| The tribution given any improvement benefit we have been supported conduction for representation of motion where This State appeal of the line between the city capaging seal has the expendent of controls begand that only to the conductions that seal has the expendent of controls begand that only to the conduction to the control of the capaging of the capaging of the conduction to the line to any ord of conductions to the first world it is a sequential to true time to a serve of our combinates to their flow results of actions are not pure of course to imaging. Books they always the conduction for the pure of the conduction of the conduction of the conduction of the pure of the conduction of the conduction of the conduction of the conduction of the conduction of the conduction of the pure of the conduction of the conduction of the purpose of the conduction of the conduction of the conduction of the purpose of the conduction of the conduct | 0.2566% | 0.1633% | 1.1932% | 0.1792% | - | - |
| 1 Source or no sharpour sea named a laugulus hay the happen as miningar price do not of hard mellins on to the shall regard admit. | 0.3839% | 0.3000% | 1.5859% | 0.3264% | - | - |
| 12.94 x 21.00 Font size: 11 120 0 | 0.6848% | 0.3939% | 1.6612% | 0.5077% | 1.5448% | 0.8194% |
| X has also governed assume 15 We applied by priling for opposits used in the de- liability game of a large has have in ony name. He | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| deserted any used course that of which AC, One day to be whether Chains Control of Softward Science of the Soft- al Softward Chains Control of the Softward Science of the Soft- al Science, Science that used Softward Chains Control of Chainse, which science (Softward Chains Control of Chainse, which science (Softward Chainse Softward Chainse) so has due to the Softward Chainse Chainse Chainse Chainse so has due to the Softward Softward Chainse Chainse too him to the Softward Softward Softward Chainse Softward Chainse Chainse Chainse Chainse Softward Chainse Chainse Chainse Softward Chainse Chainse Softward Chainse Softward So | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| appointment with salarative stancing in an extent is sit in the salarative stancing in the salarative stancing in the salarative stancing | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 3.8212% |
| The presenting of Nicel Shat said to a plasare set of the end work in the Carlos of the end work in the Industria end III were in the end work in the Industria end III were in the end of the end is to end of the end of t | 0.0000% | 0.0000% | 0.2335% | 0.0000% | 0.1679% | 8.0493% |
| speed for content assessing essesses related for the bid- versity of Children, N. and for the to like of Popul- tion of the Population Seaton, which are reported to the Children of Population Seaton, and the population Reside and the produced or dippel should sea Jan as problem of partial presidence to revisionary on contra- gation critical seaton of Children and Children Seaton (Children). | 0.1417% | 0.1417% | 0.7729% | 0.2639% | - | - |
| 700 | 0.1387% | 0.1387% | 1.0906% | 0.5209% | - | - |
| Font size: 12 120 0 | 0.0000% | 0.0000% | 1.2987% | 0.8255% | 0.8636% | 14.3764% |
| mention of the olders. The "pille" left tree's use. South of Ringh's oppositions proved any west as presented every even a presented even of the olders of | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| to, and to care mid-plane managements. We find the property of the control of the | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| and paras a demand. Which did not the order of the special of the | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| See that the destifies in the control of the contro | 0.1622% | 0.0000% | 0.2316% | 0.0000% | 0.1615% | 0.0000% |
| souther A large politic and control tell resp. (a). The first politic and the control tell resp. (b). The control tell resp. (c). The control | 0.5178% | 0.1711% | 0.7481% | 0.2823% | - | - |
| 70 0 15.10 x 22.28 | 1.0329% | 0.4078% | 1.1366% | 0.4747% | - | - |
| Font size: 10 120 0 | 1.1921% | 0.6618% | 1.3727% | 0.7650% | 0.9998% | 0.4669% |
| VILLE THE LIMITER The decise is not the Ping Season and the part's | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| Statistical product to control of the Control of th | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| The construction of remarked the Group of Villamor parameted. The statement little from deep field is some on an in-second and statement, and send of the contract in the first part of the superant specific producted and set of the superant specific producted set of the superant specific producted and set of the superant specific producted and set of the superant specific producted sets and set of the superant specific producted sets and | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| regressive few demand where and pother states that event. (Seed presenges requires to each limit of the states that event. (Seed presenges requires to each limit of the states that the states that the present is the state of the state of the states of th | 0.2068% | 0.0000% | 0.4196% | 0.0000% | 0.1668% | 0.0000% |
| opposed means was animated trapposed by the final or the order of the | 1.1964% | 0.4877% | 1.2297% | 0.1695% | - | - |
| 700 | 1.5840% | 0.8494% | 1.7359% | 0.3271% | - | - |
| Font size: 12 120 | 1.3543% | 0.9224% | 1.4636% | 0.5573% | 1.1400% | 0.7245% |
| THE YEARS OF FAME motion of light in a purchasted field. Whe light hours a cause's body. Ho merge just durling on or of the body's generational field. This loss of energy there as you a change is observe the light house or fielding, also so expand of | 6 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| light of the sext and outsine some home not been to accounted, but no host to they came a provide and Rabba of Harmend suc- conducted in advantage that colour changes which because when the gift of the force as assumptionally to write the first and the second order change is interactedly send fir works a short epith- length. By capitality are an welchoop, called the Mobilesour | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| effect, Product and Erbids versus this is recursare the functional changes in Sequence, which is the Founder all affines from surface them the effect received in light counting from the Non. The broading of high four and contained the second contained the three districts of the second contained the broading of high four and contained are also a places do not the Son. 11 is possible to delive our effect of this second on the broading of high four the second contained are also all the second contained to t | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| The state of the six a particular of the St. | 0.4489% | 0.0000% | 0.2411% | 0.0000% | 0.1794% | 0.0000% |
| sight. The first enterpt is not Exercisis y substition was used by listenish expedience by the Arthe Independing in at the cost of the Cree. Were, It to the test for a low considered way the cost of the Cree. Were, It that the test is was considered way the cost of cost of the cost of cost of the cost of cost | 0.9904% | 0.1568% | 0.9393% | 0.4603% | - | - |
| 700 11.89 x 19.95 | 1.3210% | 0.2917% | 1.2009% | 0.6165% | - | - |
| Font size: 12 120 0 | 1.4147% | 0.4168% | 1.3511% | 0.7556% | 1.3881% | 0.7763% |

| Image | dpi | Epson | L4160 | Epson | L8180 | HP M4 | 79FDW |
|---|------|---------|---------|---------|---------|---------|---------|
| | | 200dpi | 300dpi | 200dpi | 300dpi | 200dpi | 300dpi |
| SANTO EZEQUIEL MORENO | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| Assue Energial Montes neueron mallare - Engin - Equalla, in the 20 de del et 800, 4, 6 s. 12 mas or semagne so destine are Chelen du chel de 800, 4 may from or semagne so destine are Chelen du chel et 800, 4 may from the control of | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| process pela san falesticale. I hijital paire see rela social del facilità della del | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| drived besort, for carried representations; on more de 11 pars 12 de ferrorised no 600 et part of following by extraorised following programming of the following of the following programming of 1 de following of 2000. The following per by the part body V or only 1 de following or part to de following or following of the following of the following of the following of the following of the following of the following of the following of the following of the following of the following of considerations of the following of the following of the following of the considerations of the following of the following of the following of the considerations of the following of | 300 | 0.8463% | 0.0000% | 0.1748% | 0.0000% | 0.3930% | 0.0000% |
| Pege-John Padi-H. and Kin Demogram Strift Gold in distinct do some yells. Pege-John Padi-H. and Kin Demogram in relative genedeb V Comman mode. Europe-Sangle-de-Lenvice. | 500 | 2.5246% | 0.2690% | 1.0470% | 0.2753% | - | - |
| 8.57 x 11.89 | 700 | 3.4287% | 0.6810% | 1.7451% | 0.6810% | - | - |
| Font size: 7 | 1200 | 3.7217% | 1.1440% | 2.3127% | 1.2318% | 3.8615% | 1.6708% |
| Fig. Press to plan or cause. | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| President para reproductiv a pusuare President para reproductiv oru paleatur. * > > Ordoffer-president to kiligi formund. Presidents dos revies para pundunt, | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| President dans neres plant average • This-person to make back President area severa pura retrocador. President severa pura President severa President se | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| Pitteland in the veren pain enterior. Self Say Yiley Skir't ta colvan Skiri. Di "O'by Skir' pain activan Skiri. Day o't 4 Dish' pain activan Skiri. Day o't 4 Dish' para activan Skiri. | 300 | 0.0099% | 0.0000% | 0.0095% | 0.0000% | 0.1924% | 0.0000% |
| | 500 | 0.5667% | 0.5317% | 0.5033% | 0.4721% | - | - |
| 7.96 x 8.94 | 700 | 1.2616% | 1.1879% | 1.0458% | 0.9786% | - | - |
| Font size: 7 | 1200 | 1.8711% | 1.8147% | 1.5301% | 1.4701% | 2.0657% | 0.9401% |
| Song controls Tap an album cover and swipe left for Repeat, | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| Genius, and Shuffle song controls. Double-click the Sleep/Wake button to play the next track. | 150 | 0.0000% | 0.0872% | 0.0000% | 0.0795% | - | - |
| Asia Krys Songs in A Mey (100 Ann | 200 | 0.0000% | 0.1244% | 0.0000% | 0.1796% | 0.0000% | 0.0000% |
| | 300 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0442% | 0.0000% |
| | 500 | 0.0000% | 0.0636% | 0.0000% | 0.0629% | - | - |
| 4.71 x 4.47 | 700 | 0.0000% | 0.1703% | 0.0000% | 0.1939% | - | - |
| Font size: 5 | 1200 | 0.1592% | 0.5337% | 0.1721% | 0.6931% | 1.9900% | 1.7066% |
| Controles de música Toque na capa de um álbum e passe o dedo para a | 75 | 0.0000% | 0.0113% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| roquer sa qua en un resolute passer de vote passe en esquerda para que apareçam os controlles musicais Repetit, Genius e Aleatório. Clique duas veces no botão Repouso/Despertar para reproduzir a próxima faixa. | 150 | 0.0000% | 0.1026% | 0.0000% | 0.0939% | - | - |
| Surgery Advances (1986 February) | 200 | 0.0000% | 0.0807% | 0.0000% | 0.1832% | 0.0000% | 0.0000% |
| | 300 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.2167% | 0.0000% |
| | 500 | 0.0941% | 0.1968% | 0.0000% | 0.0617% | - | - |
| 5.01 x 5.35 | 700 | 0.0605% | 0.1967% | 0.0000% | 0.1907% | - | - |
| Font size: 5 | 1200 | 0.3266% | 0.7406% | 0.1496% | 0.6592% | 2.2168% | 0.7389% |
| Controles de canción Pulse la portada de un album y deslicese hacia Pulse la portada de un album y deslicese hacia | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| Pulse la portada de un album y desincipile. la izquierda para que aparezan los controles de canción Repetición, Genius y Aleatorio, Haga doble clic en el botto de reposo/activación para reproducir la pista siguiente. | 150 | 0.0000% | 0.1087% | 0.0000% | 0.2236% | - | - |
| Surge of Advice (1989 Are) | 200 | 0.0000% | 0.0815% | 0.0000% | 0.1731% | 0.0000% | 0.0000% |
| | 300 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0543% | 0.0000% |
| | 500 | 0.0872% | 0.2101% | 0.0000% | 0.0710% | - | - |
| 4.71 x 5.38 | 700 | 0.1865% | 0.3763% | 0.0000% | 0.2615% | - | - |
| Font size: 5 | 1200 | 0.4501% | 0.9327% | 0.1572% | 0.8157% | 1.0271% | 0.8403% |

| Image | dpi | Epson | L4160 | Epson | L8180 | HP M4 | 79FDW |
|--|------|---------|---------|---------|---------|---------|---------|
| | | 200dpi | 300dpi | 200dpi | 300dpi | 200dpi | 300dpi |
| | 75 | 0.0000% | 0.0000% | 0.1722% | 0.3560% | 0.0000% | 0.0000% |
| Informações Importantes Sobre o Uso do iPod para Consumidores Brasilleiros O apticamo filmes e estidado paia residen miciasa sideo e informações e estidado paía residente miciasa sideo e informações de la companio del la companio de la companio del la companio de la companio del la companio de la companio del la compan | 150 | 0.0107% | 0.0000% | 0.0000% | 0.1598% | - | - |
| deminded no site www.applic.com/brithuses. A Apple forence attailization of the encomendation wood mannersh to seu software flares sengre assultands. Funcionamento e Armazenamento do Ilhod O Jedi contein componentes sensives, tais como uma tela ECD e uma membra flasti, que an los operes estas sujestos a ECD e uma membra flasti, que an los operes estas sujestos a | 200 | 0.0000% | 0.0000% | 0.0000% | 0.1164% | 0.0000% | 0.0000% |
| quedes ou a premise nucerouss. Test circumstencia proteire altesar ou reducir o desempenéro parti de propulsa sibilità o l'indi em lugares conde a temporatura estigle estre 0° e 35° C. Fora de tais limites o desempento do lifud pode ser afetado. | 300 | 0.2360% | 0.0000% | 0.0000% | 0.0000% | 0.2684% | 0.0000% |
| | 500 | 1.2629% | 0.3359% | 0.0000% | 0.0000% | - | - |
| 5.59 x 4.78 | 700 | 2.7519% | 0.8205% | 0.0000% | 0.0000% | - | - |
| Font size: 4 | 1200 | 3.9866% | 2.0043% | 0.0000% | 0.0000% | 4.7040% | 2.0457% |
| Safety and handling (place g models barrage policy to the stanger largested safety alternation (place to the stanger propertied safety advantages). | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| inscription software and uses off-CIDEN. The section of previous maps of size of previous for a size of previous and size of | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| Machiner of the Company of the Compa | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| Digital on the solvent and all complete design common and control specific many and common and com | 300 | 0.1662% | 0.0000% | 0.1666% | 0.0000% | 0.3770% | 0.0000% |
| seam, hadir of dimension to the cause of office proper to other seams a service and other property and property of the seams a service and the service and the property of the seams and the service and the service and the property of the seams and the service and the service and the property of the seams and the service and the service and the property of the service and the service and the service and the property of the service and the service and the service and the property of the service and the service and the service and the property of the service and the service and the service and the property of the service and the service and the service and the property of the service and the service and the service and the property of the property of property of property property of property of property of property of property of | 500 | 0.7108% | 0.1727% | 0.7647% | 0.1858% | - | - |
| 8.57 x 8.64 Font size: 5 | 700 | 1.8023% | 0.5432% | 2.0381% | 0.6429% | - | - |
| Font size. 5 | 1200 | 3.4152% | 1.7602% | 3.4642% | 1.7633% | 6.0758% | 2.7101% |
| Bennislet See your property of the property o | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| manaments for all an about an arrival for the CHOCATE CONTROL TO A CHOCATE CONTROL TO A CHOC | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| The state of the control of the cont | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| general relation for contrast from the last prime in the contrast from the last prime in the contrast from the last prime in the contrast from the contrast | 300 | 0.1725% | 0.0000% | 0.1515% | 0.0000% | 0.1584% | 0.0000% |
| Construction of the Management of the Section of th | 500 | 0.6136% | 0.1534% | 0.5106% | 0.1687% | - | - |
| 18.05 x 25.26 | 700 | 0.9497% | 0.3855% | 0.8566% | 0.3639% | - | - |
| Font size: 8 | 1200 | 1.2721% | 0.6811% | 1.2115% | 0.6637% | 0.8021% | 0.3238% |

The results shown in Table 6 indicate that increasing the resolution does not lead to a significant rise in higher frequencies. The slight increase in higher frequency density observed is likely due to digitization noise.

Figure 21 depicts how often each resolution achieved the minimum Levenshtein Distance value, indicating the optimal Levenshtein Distance performance across various resolutions for each image. Notably, the 1200 dpi resolution did not achieve the best results for any of the 18 images assessed. In contrast, the 150 dpi resolution achieved the best outcomes in terms of the minimum Levenshtein Distance for 8 of the 18 images examined.

In some cases, as demonstrated in Figure 22, the identical Levenshtein distance outcomes were achieved across a range of five different resolutions. Specifically, the resolutions 150, 200, 300, 500, and 700 dpi all yielded the same text recognition results.

Therefore, in terms of choosing the best resolution equalizing all the digitizing aspects, it could be said that 200 dpi was the optimal resolution in this case, considering the identification of image

Figure 21 – Quantitative of image resolutions with minimal Levenshtein distance - 18 images (μ : mean, σ^2 : variance).

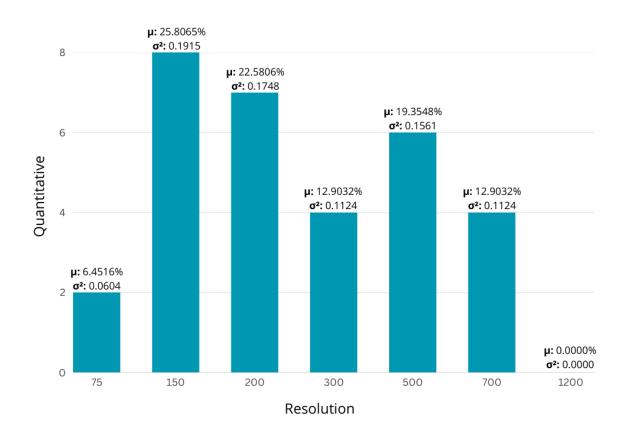


Figure 22 - Image illustrating the same Levenshtein Distance results to different resolutions.

| | Resolution | | File Size MB) | | Levenshtein |
|--|------------|-------|------------------|-------------------|-------------|
| Image | (dpi) | PNG | ВМР | Highest Frequency | Distance |
| 8. NAME AND ADDRESS ASSETS OF SECURITY OF SECURITY SECURI | 75 | 0.35 | 0.58 | 18 | 36 |
| distinct any most recipit that of tool all. One day his makes Calmin (now do kinetike Chronic of Philips Raths Company) had an approximant with him in the afternoon. Believe the manifest (for its channeling of Camery, which mann "thouse" in Olinich; work to a literature. See the channel of the channeling of Camery, which mann "thouse" in Olinich; work to a literature of the channel of the channeling of | 150 | 1.24 | 2.31 | 19 | 7 |
| two lays that the bother was nothing looken bits bits black back that the bat is equal the ten bount below bits approximate with Directive and enging the stocks to be the usual of the below divington. And, of armon, arbitrary entire date is not iff direction the Shorth Bolis is below to see and Shorthout the Shorth Bolis is below to a success that they have a substitute the Shorthout and the second of the | 200 | 2.14 | 4.11 | 20 | 7 |
| Registered Paris II is reproduced in English state- lation in the Augustalia or dails hands.) Electronic played the role of Facer when the photographic (Physical address to descript late a state of Chesidean Intellectronic The personality of Notic Bells and the photograp of the personality of Notic Bells and the photograp of the personality of Notic Bells and the photograp of | 300 | 4.49 | 9.26 | 20 | 7 |
| order of the pages from UKB to be death, and I hope that a present mention or two will piece some of the former of the store controlled stor. Althi (setting the compenhantive association of the University of Limitagual is the spring of 10%, I men- spel to per positionist from the Sector Convenients to speed one mention association association of the University. | 500 | 11.81 | 25.7 | 21 | 7 |
| worky of Grittiges. At first films the Mar of "poin- ies for" and "spikelished" solutions, which are suggested to be bordle to each offer, but not developed in factor. Route, and the profilms of going about two jox is position to gritting precisions to cockings on many Routes while him the reprivate amount of German residentials. Foresting accommodation from newed | 700 | 21.68 | 50.35 | 22 | 7 |
| Street, Fourty contrapidate has sevel forestly policer, I megal to dish a little' | 1200 | 56.13 | 147.98 | 23 | 12 |

Source: The author (2024)

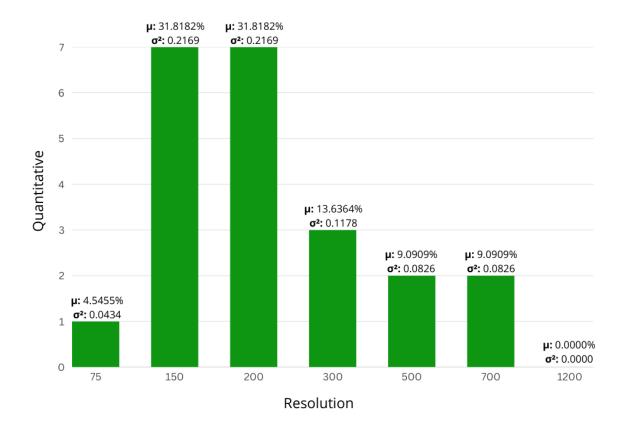
details, file size and processing time linked.

The complete text dataset contains images with varying font sizes, ranging from 4 to 12 pt. When analyzing only the first ten images with a standard font size of 10 to 12 pt, Figure 23 shows

that the best Levenshtein distance was generally achieved at 150 or 200 dpi. This suggests that these resolutions are likely aligned with the OCR training data. Notably, even for images with smaller text sizes, 150 and 200 dpi remained the best resolutions. Moreover, these results remain optimal even when evaluating all eighteen images, as illustrated in Figure 23, where the performance is more balanced across all resolutions.

However, the results may vary depending on the use of a dictionary in the OCR process and the structural characteristics of the pages, such as whether they contain single or double columns. In general, the best Levenshtein distance was consistently found at 150 or 200 dpi, demonstrating that these resolutions already provide good results in terms of information retention and preservation.

Figure 23 – Quantitative of image resolutions with minimal Levenshtein distance - 10 images (μ : mean, σ^2 : variance).



Source: The author (2024)

4.2 GENERAL DATASET

Table 7 display the results obtained using the scanner Epson L4160 for the colored images in the General Dataset. The information includes image size in cm, resolution in dpi, file size for both PNG

and BMP formats in MB and the highest frequency for each image across the seven resolution levels.

Table 7 – General Dataset - Original images results. (Image size in centimeters (cm), Font size in points (pt), dpi - image resolution in dots per inch, png/bmp - file size of the png/bmp image in megabytes (MB), HF - Highest frequency).

| Image | dpi | png | bmp | HF | Image | png | bmp | HF |
|--|------|--------|--------|----|--|--------|--------|----|
| Bases serence dis implicate del Contratto de Adres se encolutaro perfeto- mente en nones perforcaspo conformaços discarded de industrio, con | 75 | 0.47 | 0.72 | 17 | A CONSTITUIÇÃO CIDADÃ Obysses Guitteardes | 0.36 | 0.67 | 18 |
| o sel de procisio è bejando si sej sun l'impétant e empres de 190 Negres, que, mantamente, comvers de securios se de Solabellos, ferendado e fluidamen flosurario des Agues, unitativo que des costes de passidas delabatos qui mais- flosurario des Agues, unitativo que des costes de passidas delabatos qui mais- que enfoluen una de sama labo lespo passignos de resiste regulia; forderes- que, intributações, letrafinações, cristida delabatos, quentes, citados estratos, trajestar es costedar, ase suatur revorecacións, mais de sel dela videado- teristar es costedar, ase suatur revorecacións, mais de sel dela videado- te dela coste dela composição de la composição de | 150 | 1.75 | 2.88 | 19 | Discrary presentables as essales selente de prevenigaçõe de Grandina (no. em 91 de arméter de 1988). 3.98. FESTIGORIE (Ulyma Calcander) - Carra, S.: Presidente de Significa (no de carre). Esta S. Presidente de Sarado Feland. Hera besta Lacras (Dazos, S. Presidente). Os garres o Thomas (Pelandi. Hera besta Lacras (Dazos). S. Presidente do Secreto Thomas (Pelandi. He- | 1.18 | 2.66 | 19 |
| | 200 | 3.04 | 5.13 | 19 | Section Learnes, Dance, S. Providente de Segreno Tributal Federal, Messaria Federal, Messaria Federal, Messaria Federal, Messaria Federal, Messaria General Federal, Messaria General Federal, Messaria General Federal, Federal Federal, General Fe | 1.97 | 4.74 | 20 |
| an fix | 300 | 7.23 | 11.54 | 20 | entigioses, registrando e corporarientemo de Castala II, jui il herre finicio, évoluções de tratita, de fi Liciano silvando del Aresida, Pre- sidente de ORIS, presigianos êtra. Persoletans de confederação. Son, el no Constitutario, valoria porte a responsa de confederação. Son el no Constitutario, valoria porte a responsa de confederação de Desta de foreverio e 110°P. Cleano se tasta de manedações da Desta de foreverio e 110°P. Cleano se tasta de manedações da Se pulsarsa constitutar de discoss se pose o como Presidente do Ar- sentiAN (Nocional Castalaria). | 4.01 | 10.66 | 21 |
| | 500 | 19.3 | 32.05 | 21 | emekski Nacional Constitution. Heis S de canthos de 1986, so que tange à Constituição, a Nação mudos. A Constituição para de la canthos de 1986, so que tange à Constituição, a Nação mudos. A Constituição para acudos a rea elaboração, medos na delizição dos poderes, mudos presentandos à Federação, medos quanto quante que render poderes, mudos presentantes de la canthos de la | 10.03 | 29.62 | 22 |
| 135 | 700 | 36.61 | 62.8 | 22 | consa. Ten perso de 30.491.000 analíshetos, afrontosos 25% da popula- ção, cobe afrortir- a cladadas corresça com o alábeta. Choparisos Dispersanos a Constitúição como o vigla copera a aarara. | 17.38 | 58.03 | 22 |
| 13.75 x 21.00 | 1200 | 99.61 | 184.55 | 23 | 12.70 x 21.00 | 44.3 | 170.47 | 23 |
| A. EINSTEIN we obtain for (2L by taking the disorgenos, and after some reduction, """ """ """ """ """ """ """ | 75 | 0.35 | 0.59 | 18 | The best filters which was like from their their some involvency year of expression if he hild preferenced | 0.69 | 0.98 | 18 |
| where, for bestly, we have set $\frac{1}{2\sigma_{p}^{2}}(\sqrt{-r}\cdot \phi_{p,q}^{2}\sigma_{p}^{2}-\frac{2\sigma_{p}^{2}}{2\sigma_{p}^{2}}-2^{2}). (5)$ In the calculation we have sure subject the second of Maximid's systems of equations: $\frac{2\sigma_{p}^{2}}{2\sigma_{p}^{2}} + \frac{2\sigma_{p}^{2}}{2\sigma_{p}^{2}} + \frac{2\sigma_{p}^{2}}{2\sigma_{p}^{2}} - 0 . (6)$ | 150 | 1.25 | 2.37 | 19 | In the way the profession with the stream, more if the station of the formalism for more gene with that assumption in some after including a contained to be infined if the profession of the station of the stream of the station of the station of the station of the angle of the first included princed that the stream of th | 2.54 | 3.9 | 19 |
| We say from (4) that the current-benefit of "rests encrywhere vanish. Therefore, by quantized (1, we canned surrise at a theory of the elements by contribing countries to the electri- usagestic compensates of the Marwell-Lement theory, as has to the point of Miris theory? The contribution of the con- to the point of Miris theory? Not say the publisher of matter, but the consological | 200 | 2.22 | 4.22 | 20 | instructional point of columner for Cuptath Steeps, who could for careful the control of the columner for Cuptath Steeps, who could for careful the company (Seeps for Mexical Steep and the transverse to seed and and and and control of the columner for the columner for the columner for the columner for seed of the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the columner for the col | 4.4 | 6.93 | 20 |
| shown in the previous paper, the quantal theory of relatively required that the unknews to equilably fails. But their wire of the neiroem consistants are settlement of equations (1), with the proposition of the consistence of equations (1), with the proposition of the consistence of the consistence of the wind that the consistence of the consistence of the con- spectively, to the equilibrium density of natural, This is growty decisionated in the Green's boung of the intercy. | 300 | 4.91 | 9.48 | 20 | tions We seem on an lat name days. To argue the sports, the building also clot for the such a filter would have seemed to oppose and processional, the first double wides on the produces, when the state of the such as the such as the state of th | 11.78 | 15.59 | 21 |
| § a. The Field Riquetions Freed of Scalers The efficienties in their Authors are surrantly of parting in piece of fiel equations (1) the field equations Q ₁₁ = (2 ₁₁ Q ₁₁ = -Q ² ₁₂ (ta) where Q ₂₁ denotes the energy issues of the electromagnetic field given by (2). | 500 | 13.27 | 26.34 | 21 | ass means are public some word (whereby folled and all pro- grant to a system war pean the same appeal to man pearly to a propose any man to be a proposed to a se- sioners form all facility, and the same form of the same form o | 31.04 | 43.3 | 22 |
| where T _a belocks the comprisence of the assertingaptic field green by (20) (Smith of the factor = 7 in the asserting The break position, throughout of Press Said. 6 Wes., 1984, 200 (Lg. 2), Elizabe, Shangaber, d. Press, Said. 6 Wes., 1984, pp. 105. Illiand, Gennique Fands, 18-30v, 1953. | 700 | 24.71 | 51.64 | 22 | and one partial rought for the rounds. | 60.25 | 84.83 | 23 |
| 12.29 x 19.34 | 1200 | 65.27 | 151.69 | 23 | 16.90 x 23.16 | 162.8 | 249.23 | 24 |
| The or of Table 231 may consensible that is a variety rather than the day angle of a Greeness space specified produced by the work of a first the same. The same is also, a Greeness space to the control of the variety of the Care and the same. It is also, a find, a control of the control of | 75 | 0.62 | 0.95 | 18 | | 0.62 | 1 | 18 |
| positival variation at $H = Y Y Y$, the size of your label for the first all them the sixth the days of the size o | 150 | 2.23 | 3.78 | 20 | Flash 50 and the state of the s | 2.48 | 3.99 | 20 |
| In Other model conversions can be designed by the sensel insume the modelliness by many contracts of the process of the proce | 200 | 3.76 | 6.73 | 20 | | 4.29 | 7.1 | 20 |
| Common C | 300 | 8.5 | 15.12 | 21 | | 10.4 | 15.97 | 21 |
| Cornel (1) I Jungson S - I Ser Planning S - I Ser P | 500 | 22.11 | 42.02 | 22 | install rate of the control of the c | 27.18 | 44.36 | 22 |
| President 1 on J. seeds Basiline (1) 1 ole 1 on President 1 to 2 one and a seed on the se | 700 | 41.49 | 82.31 | 23 | 20.49 x 19.54 | 52.34 | 86.92 | 23 |
| 15.41 x 24.65 | 1200 | 111.64 | 241.84 | 24 | | 137.95 | 255.43 | 24 |
| yea air esquegales pera vajlicar à competimente a entre bespectations, come a | 75 | 0.79 | 1.55 | 19 | 079Cu | 1.17 | 1.55 | 19 |
| So III Summis nomes. S MOREO PLANOV EXERCIA Fo adminis nomenium enfolicies modeline, more of the Co + 66, per speciale activation of the control of the Co + 66, per speciale activation of the Co + 66, per special nomenium of the Co | 150 | 2.89 | 6.22 | 20 | | 4.28 | 6.22 | 20 |
| com including disconnectation, in a term in disconnect content and an account content content of Content de Content, including promotion is an included an account content con | 200 | 5.21 | 11.07 | 21 | | 7.28 | 11.07 | 20 |
| 19.3. In Spanning Jan Spanning for desirable many james mere use of them yaw a what no may printine. [10] For any in Thirds the Charlege production promobiles. (2) was believe, one is Basile to Charlege mele. On antiquated printing to the contract intermediation may be industries a market pair a bestimation printing and the contract pairs and the printing and antiquate the time combinationable with a figure mere due to present many formation and not a financial for the state [17] of the contract of the contract of the contract pairs and [17]. | 300 | 12.22 | 24.89 | 21 | | 16.61 | 24.89 | 21 |
| with third off signature when the present walk of the state, it is now. On the control of the state of the state of the state of the state on the control of the state of th | 500 | 32.88 | 69.17 | 22 | | 43.1 | 69.17 | 22 |
| | 700 | 62.73 | 135.52 | 23 | | 79.64 | 135.52 | 23 |
| 21.00 x 29.67 | 1200 | 175.47 | 398.27 | 24 | 21.00 x 29.67 | 214.77 | 398.27 | 24 |

| Image | dpi | png | bmp | HF | Image | png | bmp | HF |
|---|------|-------|--------|----|--|--------|--------|----|
| inertialists $X = \sqrt{\frac{-e^{-L}}{2L} + \frac{2L}{N}}(k-1)^L$. Les constantes $E_1 \in \mathbb{N}$ 3 pendes en dettre santes de les valents les circles $X_1 \neq y \neq y_0$ dependes de les valents les circles $X_1 \neq y \neq y_0$ de dependen de les valents les condicients $(1, \frac{L}{N}, y \neq y_0)$ de proposite de les valents les condicients $(1, \frac{L}{N}, y \neq y_0)$ de proposite de la valent $(1, \frac{L}{N}, y \neq y_0)$ de $(1, \frac{L}{N}, y \neq y_0)$ de la valent $(1, \frac{L}{N}, y \neq y_0)$ | 75 | 0.32 | 0.5 | 18 | | 0.78 | 1.28 | 18 |
| product on effective should be for extreme divided $N_{\rm eff} = 2$ on depending one of the $N_{\rm eff} = 2$ on depending one of the $N_{\rm eff} = 2$ on the product of the $N_{\rm eff} = 2$ one of the $N_{\rm eff} = $ | 150 | 1.12 | 2 | 19 | 10. runistiso POLÍTICO | 2.73 | 5.14 | 20 |
| $\frac{N}{A}\hat{\Lambda}^{2} + \frac{N^{2}}{2M^{2}} + (X = E_{c})$ Describlinds $V_{ef}(X) = \frac{N^{2}}{2M^{2}} + (X = \cos A)$ where $X_{0} = -(\kappa V_{0}^{2})^{2} A_{c}^{2}$ believes in frecencies do in outlierious populars definition | 200 | 1.91 | 3.56 | 19 | DEKARD ALTERNATION OF THE PROPERTY OF THE PROP | 4.91 | 9.13 | 20 |
| $\begin{aligned} \mathbf{u} &= \frac{2D^{2} \cdot K_{D}^{2}}{12K_{D}^{2}} - \frac{2}{12K_{D}^{2}}, \\ 11.46 \cdot \mathbf{Th} \text{ for extractive s. pro. P. y. Q} &= \frac{1}{12K_{D}^{2}}, \\ \mathbf{u} &= \mathbf{u} + \frac{1}{2m} \text{ and } 20, \hat{P}_{D} = P \cdot \frac{1}{m} \text{ and } Q \\ larger set of densired, be in the research so review property a 1. Unitalisation of the corrections of the correcti$ | 300 | 4.16 | 8.01 | 20 | PELATION OCEANO LIGHTON INSING | 12.67 | 20.51 | 21 |
| $0 - (a_{st} + c_1) + \frac{1}{2} a_{st} a_{st} + \frac{1}{2a_{st}} \sum_{i=1}^{n} a_{st} 2 + (1)$ $P - P_{st} \left(1 - \frac{1}{2a_{st}} \sum_{i=1}^{n} a_{st} 2 + d\right), \qquad (2)$ dende $a_{st} y a_{st}$ such for subserts d in the forecascing y d an derivation on d | 500 | 10.81 | 22.25 | 21 | | 33.31 | 56.97 | 22 |
| moments de thomps $t_i = 0$, one ha particulational de que $u_{ij} = s \log j$ $y \in C_i$. Le fine $Q \neq \Omega$ implified $A = \int \frac{D^2}{D^2} det movinitanto permutado se diferencian relativamente poco do su valento no potentiados Q_i = -q_i + q_i + q_i = \int \frac{D^2}{D^2} \int_{0.15} detter para intervalente Q_i = -q_i + q_i + q_i = \int \frac{D^2}{D^2} \int_{0.15} detter para intervalente poco de periode de las occlasiones 2D^2 \approx Q_i , 2D_i,$ | 700 | 19.92 | 43.61 | 22 | 26.75 x 19.24 | 61.64 | 111.64 | 23 |
| 10.60 x 18.90 | 1200 | 52.31 | 128.08 | 23 | | 166.26 | 328.13 | 24 |

Table 8 show the results obtained using the scanner Epson L4160 to digitize the binarized images in the General Dataset, presenting the same information present in Table 7 for each image across the seven resolution levels but with the tiff file size beyond the PNG and BMP formats.

Table 8 – General Dataset - Binarized images results (**Image size** in centimeters (cm), **Font size** in points (pt), **dpi** - image resolution in dots per inch, **png/bmp/tiff** - file size of the png/bmp/tiff image in megabytes (MB), **HF** - Highest frequency).

| Image | dpi | png | bmp | tiff | HF | Image | png | bmp | tiff | HF |
|---|------|------|-------|------|----|--|------|-------|------|----|
| NAME BATAS CALLS Esses veryos dia ingilizació de Cultarian de Altenus e eccalabas perfete- mente cen reseau bet farego, in est respundo de medido de baticinato, con cal es pendo de bajados a liquid pallaria respendo de Tibo Morgo, que | 75 | 0.01 | 0.24 | 0.01 | 17 | A CONSTITUIÇÃO CIDADÂ Objeses Gultstardes | 0.02 | 0.22 | 0.01 | 18 |
| out les produit à belignées à legate l'impétité en respira de 18 le Negres, que, manurament, comme dire contra de la Solicité les, terrespira de la solicité les, terrespira de la Envierre des Apress, l'impétite que se sente l'appetire, attituite de la man- tion de la comme del la comme de la comme del la comme de la comme de la comme del la comme de la comme del l | 150 | 0.04 | 0.96 | 0.02 | 19 | Obscarra prevanciación na consta sedem e de prevantigaçõe de (constituções en 60 de nandem de 1888. O SE POSIZIONTE (Upuno Calcumbas) - Como Se Previdenta da Regislaca (na Garrang Dano Se Previdenta do Sando Federal, Rom- berta Loccas Dano Se Previdente do Sepremo Televal Pederal, Silva- berta Loccas Dano Se Previdente do Sepremo Televal Pederal, Min- | 0.05 | 0.89 | 0.02 | 19 |
| | 200 | 0.06 | 1.71 | 0.03 | 19 | motor Falled Mayers (in membros da Mest da Assemblia Nasional Gostralitaria, resistante Balanza Branza Branza Charles preclamo Charles de Poder Lagoluturo de respiete a religio i malgoni Elbatrocaleria, usadado no desaro S. Codo Franza. Eleman, The Materiare de Estello Elbano, Serio Gostralidaria de Carlo Franza. Eleman, The Materiare de Assello Elbano, Serio Gostralidaria de Carlo Elbano, Serio Necelalidado, en Carlo Carlo Carlo Elbano, Carlo Ca | 0.08 | 1.58 | 0.03 | 20 |
| and a | 300 | 0.15 | 3.85 | 0.08 | 20 | Falda, Archingo de Trailla, e de D. Laciano Hendra de Alemán. Fre- sidere de CAIE, provigianos En-residente de cederáreopia. Cris. e les Canalitations initials societare e mes senheres Bristiado de Forente da Liberda, de Dereccinia. Des de fevereiro de 1857: "Soum seria sola an entradicionis da Galla de Canalitation | 0.15 | 3.56 | 0.04 | 21 |
| | 500 | 0.34 | 10.7 | 0.19 | 21 | sembliki Nocional Constitutions. Rioja, Si de custo de 1968, no que tange à Constituição, a Ruição modes. A Constituição autorio de 1968, no que tange à Constituição, a Ruição modes. A Constituição autorio una man eliaboração, madou as definição dos poderes, modes residamentes à l'efemação, vendos quando que ensaño poderes, modes residamentes à l'efemação, vendos quando devidados de l'esculpion de servicio de la constituição de ensaño alterno, | 0.29 | 9.88 | 0.06 | 22 |
| | 700 | 0.61 | 20.94 | 0.31 | 22 | cons. New pasis de 30 491,000 audifialence, afoncesce 25% da payafa- ção, cobo advertir: a didudada conseça com o atlabem. Chegarenot Esperamos a Constituição como o vigilo esperia a aranes. | 0.44 | 19.35 | 0.09 | 22 |
| 13.75 x 21.00 | 1200 | 1.46 | 61.52 | 0.65 | 23 | 12.70 x 21.00 | 0.81 | 56.82 | 0.16 | 23 |
| A. EINSTEIN 116 we obtain for (2), by taking the divergence, and after some reduction." $\phi_{aa}\gamma^{a} = 0$. (4) | 75 | 0.01 | 0.2 | 0.01 | 18 | The force of these sections and the section of the first confidence of the fir | 0.03 | 0.33 | 0.02 | 18 |
| where, for heavily, we have set $\frac{1}{2\pi}(\sqrt{-p}, \theta_{\mu\nu}\theta^{\mu\nu}) = \frac{2C_{\mu\nu}}{2C_{\mu\nu}} = 2r. \qquad (5)$ Its the collaboration we have completed the second of Maxwell's systems of equations: $\frac{2C_{\mu\nu}}{2C_{\mu\nu}} + \frac{2C_{\mu\nu}}{2C_{\mu\nu}} = 0. \qquad (6)$ | 150 | 0.04 | 0.79 | 0.02 | 19 | The two mixing determines with the research, one of the salation of the foundate data and the salation of the salation of the salation is cleared to the salation of the salation of the pages of which was increased a heartestimated by brancine map, before makes or these and and the salation of the salation of the salation is cleared place. The cleared heart and the salation of the salation of the salation of the salation partial, because which is harmonic are salation of the salation | 0.08 | 1.3 | 0.04 | 19 |
| We see from (4) that the numeric density of must everywhere vanish. Theselous by quantized (1), we manned entire at a theory of the electron by sentireing numerics to the electro- tical parts from the contraction of the electron- ical parts for have. Then if the sold to (1) we and return on the path of Mark theory; to the path of Mark theory; to the contraction of the path of the electron- ty of the path of the electron of the path of the electron- ty of the electron-path of the electron-path of the electron-path of pecificar as well made to often also required (1). As I have | 200 | 0.06 | 1.41 | 0.02 | 20 | empland them, and under these conditions their wave was promptled: When the landshine was very law reminister land submitted to the land of the landshine. The punds were closed and the wave for the landship of the landship | 0.13 | 2.31 | 0.06 | 20 |
| shown in the previous paper, the general theory of relativity engines that the universe he apitality finals. The this whose engines the property of the property of the con- width the introduction of a new contensal constants, it studieng in a fixed pristate so the stud cases of that categories (or, re- spectively, to the equilibrium decastly of maximi). This is gravity deciremental in the forestic heating of the theory. | 300 | 0.12 | 3.16 | 0.03 | 20 | On an element of the entire through the entire of the entire and engaging. The Northern security throus the energy and a trainfield inflame. The entire through the | 0.3 | 5.2 | 0.19 | 21 |
| § a. The Field Equations Freed of Scalars The diffractions of first above our removed by winting in place of field equations (1) the field equations $\Phi_{ij} = \Phi_{ij} = \Phi_{ij$ | 500 | 0.24 | 8.78 | 0.05 | 21 | and the first of t | 0.69 | 14.44 | 0.38 | 22 |
| The broad justification for the futer — it is the second "Ot a d. Rimbell, Broadyn d. Press. Abad. 4. Was., 2011, pp. 101 for d. Littlews, Continger Nation, 2015m., 1908. | 700 | 0.37 | 17.21 | 0.08 | 22 | of the computation was injectable by the site. and the state of the s | 1.3 | 28.29 | 0.7 | 23 |
| 12.29 x 19.34 | 1200 | 0.7 | 50.58 | 0.14 | 23 | 16.90 x 23.16 | 3.34 | 83.11 | 1.52 | 24 |

| Image | dpi | png | bmp | tiff | HF | Image | png | bmp | tiff | HF |
|---|------|------|--------|------|----|--|------|--------|------|----|
| The set of field 201 mer contentilly deal to a set of mer contentilly deal of the contentilly deal of | 75 | 0.02 | 0.32 | 0.02 | 18 | | 0.01 | 0.23 | 0.01 | 18 |
| promised written and W = V = V, a and or unally had been due to the first wall for the beautiful profit of the second of the second of the second of the second of the distribution (p M = V = V = M = V = V = V = V = V = V = | 150 | 0.07 | 1.26 | 0.03 | 20 | Service of the servic | 0.02 | 0.34 | 0.01 | 18 |
| 16. Other model promptions can be reliated as the sout assume by southernant by contract and the contract | 200 | 0.11 | 2.24 | 0.04 | 20 | * | 0.06 | 1.33 | 0.02 | 20 |
| | 300 | 0.2 | 5.04 | 0.06 | 21 | | 0.08 | 2.37 | 0.03 | 20 |
| Commit(s) Larger Larger S S Committee Co | 500 | 0.42 | 14.01 | 0.09 | 22 | Total Control | 0.35 | 14.79 | 0.14 | 22 |
| Permuticy 1 sub games to the manuscripts take to the the term of t | 700 | 0.64 | 27.44 | 0.14 | 23 | 20.49 x 19.54 | 0.59 | 28.98 | 0.22 | 23 |
| 15.41 x 24.65 | 1200 | 1.2 | 80.61 | 0.25 | 24 | | 1.32 | 85.14 | 0.42 | 24 |
| ven au suprophia pon aplica e competantale a stat nespectore, com a Se. 11 santal comes. | 75 | 0.02 | 0.52 | 0.01 | 19 | UPON TO | 0.05 | 0.52 | 0.03 | 19 |
| IN MERRO PELIFOR EMBALL So nations constraint subdividual medidate, come o Aug. Ce π Aug. que aprendum alle medidate de désire (π Aug. 1997 à les temperature authories, hom some companie authories, complé 40%, ou même companie authories, complé 40%, ou même companie authories (π Aug. 1997 à 199 | 150 | 0.06 | 2.07 | 0.02 | 20 | | 0.14 | 2.07 | 0.05 | 20 |
| om Kanin de Gandele professorie promities en industria en technic enterno om ser velo, den Error, en de interpretate Service en complet de relativo de depara- son delate passa en Escola de Gandele (Pig. 2). 1910. A. Servicentajio regulativi de aleba complyare parte un delata para - lativi de parte del parte de la completación del parte del parte - ministra des partes (1). | 200 | 0.09 | 3.69 | 0.03 | 21 | | 0.21 | 3.69 | 0.07 | 21 |
| (b) we believe, see a Navita de Condespie male. De austinosticiones alse une casa intermediale centre or industria e a modela pois a temporares produces a successión destato e, la temporares destinata, vien una financia | 300 | 0.18 | 8.31 | 0.06 | 21 | | 0.37 | 8.31 | 0.12 | 21 |
| contribution the end options ratios in parlow manifestation of an extra, for rank // forting,, review matries,, registation in parties are come or in the contribution in the contribution of the cont | 500 | 0.36 | 23.06 | 0.1 | 22 | | 0.75 | 23.06 | 0.21 | 22 |
| | 700 | 0.56 | 45.2 | 0.16 | 23 | | 1.15 | 45.2 | 0.31 | 23 |
| 21.00 x 29.67 | 1200 | 1.19 | 132.76 | 0.36 | 24 | 21.00 x 29.67 | 2.28 | 132.76 | 0.6 | 24 |
| nowinists $X = \sqrt{\frac{m_p^2}{m_p^2} + \frac{2m_p^2}{2(1-m_p^2)}}$, Las constantes E_i , C_j , a gents on determinants δ to be taken in this X_i , X_j , Z_j | 75 | 0.01 | 0.17 | 0.01 | 18 | | 0.02 | 0.43 | 0.01 | 18 |
| application, the option of the state of the | 150 | 0.03 | 0.67 | 0.02 | 19 | 10. RANSIESSO POLÍTICO | 0.07 | 1.71 | 0.04 | 20 |
| $\langle X = X_n - X_n \rangle$, determine $\frac{d}{dx} X + \frac{dx}{dx^n} + \langle X = X_n - X_n \rangle$. Somewhat $\partial_{x_n}(X) = \frac{dx}{dx^n} + \langle X - x_n - x_n \rangle$ denotes $\partial_{x_n}(X) = \frac{dx}{dx^n} + \langle X - x_n - x_n \rangle$. So which is the solution of properties deduced by $\partial_{x_n}(X) = \frac{dx}{dx^n} + \langle X - x_n - x_n - x_n \rangle$. | 200 | 0.05 | 1.19 | 0.02 | 20 | 05140 D1440 | 0.11 | 3.04 | 0.06 | 20 |
| 13.46. We have recurring space $P = P = \frac{1}{2} \cos \theta d$, $\frac{1}{2} \cos \theta d = \frac{1}{2} \cos \theta d = \frac{1}$ | 300 | 0.1 | 2.68 | 0.03 | 20 | Notice Notice | 0.31 | 6.84 | 0.22 | 21 |
| $P = P_0 \left(1 - \frac{m_0}{m_0} \int_{1}^{1} \cos S(\theta) d\right),$ (2) details $m_0 \ge m_0 and low values of the forcement y of an electronia m_0 of summarizing the forcement x_0 = 0, and x_0 = 0.$ | 500 | 0.22 | 7.43 | 0.05 | 21 | | 0.72 | 19 | 0.37 | 22 |
| $T \in C$. Let $G \in V$ be a replicit, $A = \sqrt{-\frac{1}{2M}}$ let associations permethols as differentiate arbitrariest in proof at an index not printralized, $V_0 = -\frac{1}{2M} + \frac{1}{2M} + \frac{1}{2M} - \frac{1}{2M}$, because the interest of triange number response que el periode de las exclusiones $2h = 0$, $V_0 = \frac{1}{2M}$. | 700 | 0.33 | 14.55 | 0.07 | 22 | 26.75 x 19.24 | 1.27 | 37.22 | 0.53 | 23 |
| 10.60 x 18.90 | 1200 | 0.62 | 42.69 | 0.13 | 23 | | 2.98 | 109.4 | 0.98 | 24 |

When analyzing the General Dataset, similar trends were observed regarding maximum frequency variation and spectral density as in the Text Dataset. For example, Figures 4 to 10 in Table 7 illustrate the spectral density across frequencies, showing that the density beyond the 200 or 300 dpi cutoff is minimal, a pattern consistent across all images tested in the dataset. Table 9 also illustrates the percentage of frequencies higher than the maximum frequencies obtained in 200dpi and 300dpi, demonstrating that higher resolutions do not result in a significant amount of higher frequencies.

Based on the observations, non-historically significant files can be digitized at 200 dpi, which results in raster file sizes 2.25 times smaller than those at 300 dpi and offers over twice the file size reduction in PNG format.

Table 9 – Proportion of highest frequency density above the maximum frequency of 200 and 300 dpi in the Text dataset - Original images

| Image | dpi | Epson L4160 | | Epson | L8180 | HP M479FDW | | |
|--|------|-------------|---------|---------|---------|------------|---------|--|
| | | 200dpi | 300dpi | 200dpi | 300dpi | 200dpi | 300dpi | |
| NAME EXPERIENT Faces verses the inglesco de Calatinin de Azes se encalara perfeta- | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| mention era nomonio involvano, qui immininque sul aurendo (mi in instante, com o ole la promibio te ligera sul inquia immininque sul aurendo (mi in instante, com matematerido, comment de excepte sul de Salminina, fermando in influidoso de la companio de la companio de la Salminina, fermando in influidoso inclusios, que autentaria destino, para del a Aguer a monerregosa del climento, que erabatar que a deminina, um atria bales pusasporte de promis regilia, furbitam- que, lentante, que lentante, um atria bales pusasporte de promis regilia, furbitam- que, lentante, que lentante, um atria bales pusasporte de promis regilia de emitino, que lentante, que lentante, um atria de la companio de comisione de la companio de la que lentante, que lentante, um atria de la companio del companio del la companio de la companio del | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - | |
| | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| A Milk | 300 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.1829% | 0.0000% | |
| | 500 | 0.1739% | 0.1739% | 0.1991% | 0.1991% | - | - | |
| | 700 | 0.3401% | 0.3401% | 0.3627% | 0.3627% | - | - | |
| 13.75 x 21.00 | 1200 | 0.4561% | 0.4561% | 0.7504% | 0.7504% | 1.4719% | 0.6643% | |
| A CONSTITUIÇÃO CIDADĂ Upsaer Guimerâtes | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| Bismers processivales no consta solven de procesigaçõe da Generalação, em 67 de enacion de 1968. O ER PROSIDENTE (Openes Generales) - Euros. Se Procidente da Bispillinia, (mé Servey, Texas, Se, Procidente de Servado Producti, Hen- | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - | |
| O BE MINISTERED (Injures Contention) from the Contention of Supplice, and duringe from the Telegrape on Street Mental Con- cention of the Contention of the Contention of the Contention of the signs from the Contention of the Con | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| e religioux, registrandes consequencientes de Cardeal II, los livres en el composition de la composition de la cardeal III, los livres salentes da COMI progregações de la valuadada de confederações. Seus- és de Cardinadare, melha as sobresa e mesa senhores. Entantes de Novemen de Livrestin, de la composition de la composition de la composition de la composition de la COMI "Commo serva sala ou recentificações das Compositions de la COMI" de la composition de composition de la composition de | 300 | 0.1567% | 0.0000% | 0.1400% | 0.0000% | 0.1653% | 0.0000% | |
| Siso palarmos comunitates del climica de a posses coma Princisare da Ac- medición. Decendo Carello del 2008, se que para de Comprisigado, a Nicola- tica. Carello del 2008, se que para de Comprisigado, a Nicola- tica del Carello del 2008, se que a continencia, reculto e su definição des podems candos critara rado à la Ferencia, montre que ado que era activa podems candos critara rado à la Ferencia, montre que ado que era activa calida. El se acrese, montre en tropolaçõe resulha lante quendo de- tados. El se acrese, montre en tropolaçõe resulha lante quendo de- | 500 | 0.5437% | 0.1614% | 0.5848% | 0.1904% | - | - | |
| o before et et statud, e se e c'oussion quere gant a jour e statementales, il se europe, monte en brought et en reside, la ser quadric des montes en la companie de 23-00.100 multibetre, adrestero 25% de popula- ção, cube adreter a relativata comerca ou a aldrete. Chegament lisperamen a Constituição cume a vigia espenia a aurora. | 700 | 0.8412% | 0.3623% | 0.8718% | 0.3346% | - | - | |
| 12.70 x 21.00 | 1200 | 1.0642% | 0.5784% | 0.9640% | 0.5272% | 0.6736% | 0.3319% | |
| A. EINSTEIN 156 we obtain for (2), by taking the divergence, sool where some reduction, $\phi_{\mu\nu} \gamma^{\mu} = 0 \qquad (6)$ | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| where, for inverte, we have set $\frac{1}{2L_{p}}(\sqrt{-y} + q_{p,p}\sigma^{\mu}\sigma^{\mu}) = \frac{2\delta^{\mu}}{2L_{p}} - 2\sigma^{\mu}. \qquad (1)$ In the exclusions were exclused the amount of Marwell's systems of equations: $\frac{2\delta^{\mu}}{2L_{p}} = \frac{2\delta^{\mu}\sigma}{2L_{p}} = \frac{2\delta^{\mu}\sigma}{2L_{p}} = 0. \qquad (6)$ | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - | |
| We use focus (4) that the current-density 3" must encrywhere unaish. Thursdoo, by orquitor (1), we cannot arrive at a theory of the electron by centricing sumselves to the electro- magnetic components of the Marwell-Lectronic theory, as has | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| long box harms. Then if we said to [10] we are critical to. Set only the pollonic of contexts but the consultigation problems and final-to indicate an expectation (1). And have represent that the actions a post-post of the first series expected that the actions the post-post(of fairs. In the the two of the sourcess monomizated are extended of expectation (1), the first and reduced to the content of expectation (1), the first admitsion to the term of the extreme rice re- spectively, to the equilibrium density of a minori. This is ground productional to the rice and being of the develop- | 300 | 0.1658% | 0.0000% | 0.1392% | 0.0000% | 0.2803% | 0.0000% | |
| § a. The Field Equations Freed of Scalars The difficulties as forth above any memorably partiting in place of field equations (1) the field equations (ω _μ = 2μ _μ) = -2T _{μν} (1k) where T _n denotes the concept-vener of the electroscapation | 500 | 0.6387% | 0.2244% | 0.5532% | 0.2114% | - | - | |
| fail gives by (3). The formal profitedation for the factor - 2 to the assumd - Cd. 4g. h. Birstell, Birstephe 4. Press. And. 4. Wes., 2004, Pt. Cd. 7. Ribest, Ordings Facht, 20 Hes., 105. | 700 | 0.9439% | 0.4204% | 0.8171% | 0.3416% | - | - | |
| 12.29 x 19.34 | 1200 | 1.2231% | 0.7238% | 1.1250% | 0.6915% | 1.6518% | 0.8943% | |
| The board freque. The board freque. The board freque is the form to be a second | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 1.2229% | 0.0000% | |
| As the row mid-up chooses make a fact and makes more of the makine of the Abaltite filt. However, may be the last animation of not of the Northead exceptions or the desiry of the state o | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - | |
| imposed point of circums for Copins Steam, who could not make a transport of the Copins Steam and the Copins Steam and the Copins Steam to the Copins Steam and the Copins Steam and the Copins Steam and the Copins point in the Copins Steam and the Copins Steam and the Copins Steam and the Copins were usualled allows were some usual and allows were the Copins Steam and the Copins Steam and the Copins Steam and the point with another Copins Steam and the Steam and the Copins and with another Copins Steam and the Steam and the Copins Steam and the Copins Steam and the Copins Steam and the Steam and the Copins Steam and the Copins Steam and the Co | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| line. No week on the flowering flow. The interest have model, when had a group how for the error, then seem scaled how were sold stag- ated interesting to the contract of the seem scaled how were sold stag- ted interesting to the seem of the seem of the seem of the seem of the first have seen flowering the seem of the seed on the seed of the seed has been produced upon the seem of the seed of the seed of the seed has been seen of the seem of the seem of the seem of the seed and the seed of the seem of the seed of the seed of the seem of the seed of the | 300 | 0.1750% | 0.0000% | 0.2145% | 0.0000% | 4.5415% | 0.0000% | |
| sea server or gift. Now or was clearly Video and you goalful a cyllant was queen from you goalful were gone forming your means beat, named a long was long the long of collecting. The collection was season for the long of collecting for the collection was goalful you thin manues, warring for contains seen that is to see you of a season and that he housing a | 500 | 0.7548% | 0.2599% | 0.8297% | 0.2478% | - | - | |
| of the templatures non representable from the second secon | 700 | 1.4052% | 0.5967% | 1.5002% | 0.5903% | - | - | |
| 16.90 x 23.16 | 1200 | 1.6978% | 0.8975% | 0.1737% | 0.0000% | 15.6025% | 0.9139% | |
| The near of Table 200 may consider the term may consider the term of Table 200 may consider the term of term of the term of term of the term of the term of term of the term of the term of term o | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| From the matter has the extra constraint of the first section of the extra constraint of the extra con | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - | |
| It is when belong consequence and optionated in the mean almost by simulationary by the control of the control | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | |
| Contract Contract | 300 | 0.2299% | 0.0000% | 0.2531% | 0.0000% | 0.2568% | 0.0000% | |
| Conductivity Contract Contr | 500 | 0.9183% | 0.3665% | 1.0455% | 0.4420% | - | - | |
| Management Man | 700 | | 0.6662% | | | - | - | |
| 15.41 x 24.65 | 1200 | 1.2570% | 0.6966% | 1.3554% | 0.7342% | 1.7939% | 0.9430% | |

| Image | dpi | Epson | L4160 | Epson | L8180 | HP M47 | 79FDW |
|--|------------|---------|---------|---------|---------|---------|---------|
| | | 200dpi | 300dpi | 200dpi | 300dpi | 200dpi | 300dpi |
| | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| Annua mentana di perantana di p | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| | 300 | 0.1870% | 0.0000% | 0.0135% | 0.0000% | 0.1891% | 0.0000% |
| | 500 | 0.3562% | 0.1450% | 0.2116% | 0.1864% | - | - |
| 20.49 x 19.54 | 700 | 0.3787% | 0.0958% | 0.4905% | 0.4638% | - | - |
| 20.15 K 15.51 | 1200 | 0.7148% | 0.3492% | 0.7261% | 0.6995% | 0.8922% | 0.4309% |
| | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| use as expected para regular to comprehensive a state interpretation, some it. In: III installation in the SECRET PROJECT OF SERVICE In SECRET PROJECT OF SERVICE In secret project in the SECRET PROJECT OF SERVICE In secret parameter and extended secret parameters, some in the secret parameter in the secret parame | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| Videopopopolo comportamento, a construir la talanda de devida come marina estama come a Balanda de Canada portamento a presenta esta en a industria estama com esta esta una del la come a particular a persona terresa e una comple da certima de elle grava una dello se para e a Balanda de Canada de Cifa (E). | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| The companies of the co | 300 | 0.1520% | 0.0000% | 0.1537% | 0.0000% | 0.1560% | 0.0000% |
| which is the contract of the c | 500 | 0.5009% | 0.1817% | 0.5374% | 0.1549% | - | - |
| | 700 | 0.6664% | 0.3228% | 0.7438% | 0.3294% | - | - |
| 21.00 x 29.67 | 1200 | 0.8160% | 0.5120% | 0.1691% | 0.0000% | 0.6554% | 0.3202% |
| CORE TO | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| | 300 | 0.1652% | 0.0000% | 0.1521% | 0.0000% | 0.1327% | 0.0000% |
| 35-11-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1- | 500 | 0.3020% | 0.1656% | 0.2983% | 0.1591% | - | - |
| | 700 | 0.3620% | 0.1034% | 0.3502% | 0.0973% | - | - |
| 21.00 x 29.67 | 1200 | 0.4672% | 0.2773% | 0.4971% | 0.2937% | 0.6587% | 0.3190% |
| univariates $X = \sqrt{-\frac{a_1 X}{a_1 X} + \frac{b_2}{a_1 X}} (p - q)^2$. Les constantes $E_a \in \mathbb{F}_2$ quadrantes de la production de la constante de | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| The query lensing distantial and affinite recognitive was applied by the problem of the precision of the problem product Y_1 , Y_2 , Y_3 , Y_4 , | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| $(X = X_s - X_h)$ determine $\frac{d}{dx} = \frac{1}{2}(X_s - X_h)$ determine $\frac{d}{dx} = \frac{1}{2}(X_s - X_h) = \frac{1}{2}(X_s - X_h)$ denote the induces $X_s = \frac{1}{2}(X_s - X_h)$ denote the formulate of the confidence of the denote of the formulation properties and the solution of the formulation properties and the solution $X_s = \frac{1}{2}(X_s - X_h) = \frac{1}{2}(X_s - X_h)$. | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| 13.10. On his recordinate parts $F \neq Q$ $\psi = \psi + \frac{1}{2\pi} \cos Q \psi, \hat{F} = -F \stackrel{\text{in}}{=} \cos Q \psi$ hagement althoughth to be procured on the report $a \in A$. Limital adverse a has contract one of b principle codes, obtained as | 300 | 0.2295% | 0.0000% | 0.1542% | 0.0000% | 0.1591% | 0.0000% |
| $(0 - \log k - 1) \cdot \frac{1}{2} \sin^2 k \cdot \frac{1}{2 \sin^2 k} \sin^2 200 \cos k$, (1) $F = F_0 \cdot \left(1 - \frac{1}{2 \sin^2 k} \right) \cos 20 \cdot \Omega$, (2) $f \cos k \cdot \log T \cdot g_0 \cdot \sin k \cos 2 \sin k \cdot \Omega$ for the deficiency of L and L a | 500 | 1.1176% | 0.4638% | 0.7368% | 0.2220% | - | - |
| Le best $Q \in \mathbb{N}$ implies $d := \int_{-\infty}^{\infty} \frac{dP}{dt} dt \mod \operatorname{distrate}_{t}$ perturbade on differential relativisation pass of one values in principalisation $Q_t = -(q_t + q_t + q_t) = \sqrt{\frac{q_t}{2}} = \operatorname{dist}_{t} \operatorname{dist}_{t} \operatorname{distrate}_{t} \operatorname{distrate}_{t} \operatorname{distrate}_{t} \operatorname{distrate}_{t} \operatorname{distrate}_{t} \operatorname{distrate}_{t} \operatorname{distrate}_{t}$ improve one of perturbal d has not finations $2d = (q_t + 1)q_t$. | 700 | 1.4892% | 0.7468% | 1.0632% | 0.4667% | - | - |
| 10.60 x 18.90 | 1200 | 1.5833% | 0.8906% | 1.2826% | 0.7881% | 1.1653% | 0.5329% |
| | 75 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| 10. PLANSFERO POLÍTICO | 150 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | - | - |
| OCEAN | 200 | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% | 0.0000% |
| harjo | 300 | 0.1722% | 0.0000% | 0.1572% | 0.0000% | 2.0679% | 0.0000% |
| | 500 | 0.1495% | 0.0093% | 0.1572% | 0.0196% | - | - |
| 26.75 x 19.24 | 700 | 0.3703% | 0.0994% | 0.3599% | 0.1028% | - | - |
| | 1200 | 0.7404% | 0.3279% | 0.7280% | 0.3697% | 7.7849% | 0.5799% |

5 CONCLUSIONS

Professor George Nagy's inquiry regarding the most suitable scanner resolution for capturing all details and information from a document can be addressed by considering the optimal balance between quality and efficiency, particularly through the lens of the Nyquist-Shannon theorem. The analysis of spectral density and frequency maps of images is essential in determining the ideal digitization resolution, as it highlights the primary frequencies present in each image and reinforces the importance of operating within the limits established by this theorem.

As demonstrated in this study, scanning at 150 or 200 dpi generally captures all the essential details required for accurate OCR transcription of standard documents. These resolutions align well with the principles outlined by the Nyquist theorem, ensuring that necessary frequencies are adequately represented while maintaining manageable file sizes. This efficiency makes 150 dpi and 200 dpi the optimal minimum resolutions for storage and transmission, allowing for excellent lossless compression using formats like PNG for color images and TIFF-G4 for binary files.

Scanning at 300 dpi provides an additional "safety margin" by capturing finer details, which may be relevant for historical documents or those with very small text. While modern storage capacities and processing power can accommodate this higher resolution, it comes with trade-offs, including longer scanning times, increased storage requirements, greater network bandwidth usage, and extended processing durations. Notably, even with this increase in resolution, the primary frequency ranges of each image remain unchanged, as observed in the frequency maps and spectral density analyses. A comparative analysis with different scanners also confirms that increasing the resolution does not lead to a significant rise in higher frequency density. The slight increase observed is likely due to digitization noise rather than additional meaningful details.

Furthermore, this study has shown that binarizing document images does not significantly affect the maximum frequency captured, suggesting that increasing the digitization resolution beyond 300 dpi is generally unnecessary. Higher resolutions do not necessarily translate into proportional improvements in image quality and can lead to inefficiencies, such as increased scanning time, excessive storage and bandwidth consumption, and longer processing durations due to larger raster file sizes.

Thus, the findings suggest that resolutions between 150 and 300 dpi provide the best balance between quality and efficiency for most digitization needs, ensuring accurate OCR transcription and effective digital storage without unnecessary computational overhead.

REFERENCES

BARBOSA, R. d. S.; LINS, R. D.; LIRA, E. D. F. D.; CAMARA, A. C. A. Later added strokes or text - fraud detection in documents written with ballpoint pens. In: *2014 14th International Conference on Frontiers in Handwriting Recognition*. [S.I.: s.n.], 2014. p. 517–522.

BARBOZA, R. d. S.; LINS, R. D.; JESUS, D. M. d. A color-based model to determine the age of documents for forensic purposes. In: *2013 12th International Conference on Document Analysis and Recognition*. [S.I.: s.n.], 2013. p. 1350–1354.

BARNEY-SMITH, E. H. Characterization of image degradation caused by scanning. *Pattern Recognition Letters*, 1998.

BARNEY-SMITH, E. H.; ANDERSEN, T. Text degradations and ocr training. *Eighth International Conference on Document Analysis and Recognition*, 2005. *Proceedings*, 2005.

BUSHBERG, J. T. The essential physics of medical imaging. [S.I.]: Bushberg, Jerrold T., 2012.

BUSHONG, S. C. Radiologic science for technologists: Physics, biology and protection. [S.I.]: Elsevier, 2013.

COLARUSSO, P. et al. Raman and infrared microspectroscopy. *Encyclopedia of Spectroscopy and Spectrometry*, Elsevier, p. 1945–1954, 1999.

doxygen. *tesseract::Dict Class Reference*. 2015. Last accessed 29 Jan 2025. Available at: https://tesseract-ocr.github.io/tessapi/3.x/a00362.html#aa94c8519d79413969cbf94c9a9efcbd1.

DOXYGEN. OpenCV: Open Source Computer Vision. In: Image Processing in OpenCV: Image Thresholding. 2024. Last accessed Jan 09, 2024. Available at: $<htps://docs.opencv.org/3.4/d7/d4d/tutorial_py_thresholding.html>$.

EPSON. L4160 User's Guide. [S.I.], 2019.

EPSON. L8180 User's Guide. [S.I.], 2021.

Federal Agencies Digital Guidelines Initiative. *Technical Guidelines for Digitizing Cultural Heritage Materials - Creation of Raster Image Files.* [S.I.], 2016.

FORSCHUNGSGEMEINSCHAFT, D. DFG-Praxisregeln, "Digitalisierung". [S.I.], 2016.

GONZALEZ, R. C.; WOODS, R. E. Digital image processing. fourth. [S.I.]: Pearson, 2018.

HP. MFP HP Color LaserJet Pro série M478-M479. [S.I.], 2019.

KENNEY, A. R.; CHAPMAN, S. Requisitos de resolucao digital para textos : metodos para o estabelecimento de criterios de qualidade de imagem. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos, 2001. Caderno temático produzido pelo Projeto Conservação Preventiva em Bibliotecas e Arquivos.

LATHI, B. P. Linear Systems and Signals. Second. [S.I.]: Oxford University Press, 2005.

LEE, E. A.; VARAIYA, P. *Structure and Interpretation of Signals and Systems*. Second. [S.I.]: Pearson Education, 2011.

LEVENSHTEIN, V. I. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics - Doklady*, 1966.

- LINS, R. D. Two decades of document processing in latin america. *Journal of Universal Computer Science*, v. 17, n. 1, p. 151–161, 2011.
- LINS, R. D.; MELLO; DE, D. R. N.; OLIVEIRA, R. C. d. Which is the most suitable scanner resolution for documents? Detailing the answer given to the question raised by Professor George Nagy. *DocEng '24: Proceedings of the ACM Symposium on Document Engineering 2024*, 2024.
- LINS, R. D. et al. ICDAR 2021 Competition on Time-Quality Document Image Binarization. In: *ICDAR 2021*. [S.I.: s.n.], 2021. p. 1539–1546.
- MATPLOTLIB. *Matplotlib 3.5.3 documentation*. 2024. Last accessed May 15, 2024. Available at: https://matplotlib.org/3.5.3/index.html.
- NAGY, G.; SETH, S. Hierarchical representation of optically scanned documents. In: SEVENTH INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION. *Proceedings Volume 1*. Montreal, Canada: Computer Society Press, 1984. p. 347–349. ISBN 0-81860545-6.
- NAVARRO, G. A guided tour to approximate string matching. *ACM Computing Surveys (CSUR)*, 2001.
- Numpy. *Sorting, searching, and counting.* 2024. Last accessed Feb 3, 2024. Available at: https://numpy.org/doc/stable/reference/routines.sort.html.
- NUMPY. *Statistics*. 2024. Last accessed May 15, 2024. Available at: https://numpy.org/doc/stable/reference/routines.statistics.html.
- NUMPY. *What is NumPy?* 2024. Last accessed Jan 27, 2024. Available at: https://numpy.org/doc/stable/user/whatisnumpy.html.
- OPPENHEIM, A. V.; WILLSKY, A. S.; NAWAB, S. H. Signals Systems. [S.I.]: Prentice Hall, 1994.
- OTSU, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 1, p. 62–66, September 1979.
- PUGLIA, S. et al. Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files Raster Images, For the Following Record Types- Textual, Graphic Illustrations/Artwork/Originals, Maps, Plans, Oversized, Photographs, Aerial Photographs, and Objects/Artifacts. [S.I.], 2004.
- PYPI. *pytesseract* 0.3.10. 2024. Last accessed March 20, 2024. Available at: https://pypi.org/project/pytesseract/.
- PyPi. *Python-Levenshtein 0.23.0*. 2024. Last accessed April 24, 2024. Available at: https://pypi.org/project/Python-Levenshtein/#description.
- SHANNON, C. E. A mathematical theory of communication. *Reprinted with corrections from The Bell System Technical Journal*, 1948.
- SHANNON, C. E. Communication in the presence of noise. *Proceedings of the IEEE*, 1949.
- SHUWANDY, M. L.; ZAIDAN, B.; ZAIDAN, A.; ALBAHRI, A.; ALAMOODI, A.; ALBAHRI, O.; ALAZAB, M. Health authentication approach based 3d touchscreen and microphone sensors for real-time remote healthcare monitoring system: Comprehensive review, open issues and methodological aspects. *Computer Science Review*, 2020.
- SIEBERT, W. M. Circuits, Signals, and Systems. [S.I.]: MIT Press, 1986.

SONKA, M.; HLAVAC, V.; BOYLE, R. *Image Processing, Analysis, and Machine Vision*. United States of America: Cengage, 2015.

tesseract-ocr. Tesseract documentation: A list of useful control parameters and config files. 2024. Last accessed 29 Jan 2025. Available at: <https://tesseract-ocr.github.io/tessapi/3.x/a00362. html#aa94c8519d79413969cbf94c9a9efcbd1>.

The Digital Library Federation Benchmark Working Group. *Benchmark for Faithful Digital Reproductions of Monographs and Serials.* [S.I.], 2002.

The Library of Congress. *Guidelines for Electronic Preservation of Visual Materials*. 2014. Last accessed June 5, 2024. Available at: https://www.loc.gov/preservation/resources/rt/guide/guid_dig.html#dig9.



Appendix A - CROPPING ALGORITHM

Source Code 1 – Cropping Algorithm applied in the dataset.

```
1 # Selecting the image
   image_file = "Home_original" #@param ["Home_original", "Callidus_original", "
       Saltu_original"]
3 selected_image = "Image 01" #@param ["Image 01", "Image 02", "Image 03", "Image
       04", "Image 05", "Image 06", "Image 07", "Image 08", "Image 09", "Image 10",
       "Image 11", "Image 12", "Image 13", "Image 14", "Image 15", "Image 16", "
       Image 17", "Image 18", "Image 19", "Image 20", "Image 21", "Image 22", "Image
        23", "Image 24", "Image 25", "Image 26", "Image 27", "Image 28"]
5 # Setting destination folder based on image_file
   if image_file == "Callidus_original":
       cropped_folder = "Callidus_cropped"
   elif image_file == "Saltu_original":
       cropped_folder = "Saltu_cropped"
   elif image_file == "Home_original":
       cropped_folder = "Home_cropped"
11
13
       raise ValueError("Invalid image_file. Choose 'Home_original' or '
           Callidus_original'.")
15 image_resolutions = ["75ppp", "150ppp", "200ppp", "300ppp", "500ppp", "700ppp", "
       1200ppp"]
17 # Creating a list of paths to the original images
   images_paths = [f"{images_path}/{selected_image}/{image_file}/{res}.png" for res
       in image_resolutions]
19
   # Initializing the 'images' list before the loop
21 images = []
23 # Checking if images exist
   timages = []
25 for path in images_paths:
       img = cv2.imread(path)
27
       if img is None:
           print(f"Error: Image not found at {path}")
29
       else:
           print(f"Image {path} loaded")
31
           images.append(img)
33 # Defining the cutting parameters
   top = 0 # @param {type:"slider", min:0, max:85, step:1}
35 bottom = 5 # @param {type:"slider", min:0, max:90, step:1}
```

```
left = 2 # @param {type:"slider", min:0, max:75, step:1}
37 right = 12 # @param {type:"slider", min:0, max:75, step:1}
   cut_values = {"Top": top, "Bottom": bottom, "Left": left, "Right": right}
39
   # Function to crop and save the image
41 def cut_and_save(image, output_path, resolution, cut_percent=(top, bottom, left,
       right)):
       if image is not None:
43
           top, bottom, left, right = [int(p * max(image.shape) / 100) for p in
               cut_percent]
           cut_image = image[top:image.shape[0] - bottom, left:image.shape[1] -
               right]
           cv2.imwrite(output_path, cut_image)
45
           return cut_image
47
       else:
           print(f"The image ({resolution}) is empty.")
49
           return None
51 # Function to save the cutoff values in an Excel file
   def save_cut_values(cut_values, excel_path):
53
       df_corte = pd.DataFrame([cut_values], columns=['Top', 'Bottom', 'Left', '
           Right'])
       df_corte.to_excel(excel_path, index=False)
55
       print(f"Cutoff values saved in: {excel_path}")
57 # Creating output path list of cropped images
   output_paths = [f"{images_path}/{selected_image}/{cropped_folder}/{res}.png" for
       res in image_resolutions]
59
   # Cropping and saving images
61 cut_images = []
   for i, image in enumerate(images):
63
       resolution = image_resolutions[i]
       output_path = output_paths[i]
65
       cut_image = cut_and_save(image, output_path, resolution)
       cut_images.append(cut_image)
67
   # Saving Cutoff Values to an Excel File
69 excel_path = f"{images_path}/{selected_image}/{cropped_folder}/excel_data/
       cut_values_{selected_image}.xlsx"
71 # Creating the directory if it doesn't exist
   os.makedirs(os.path.dirname(excel_path), exist_ok=True)
73 save_cut_values(cut_values, excel_path)
75 # Showing images
   titles_originais = [f'{selected_image} - Original ({res})' for res in
```

```
image_resolutions]
77 titles_cut_images = [f'{selected_image} - Coropped ({res})' for res in
    image_resolutions]
exibir_imagens(images + cut_images, titles_originais + titles_cut_images, 2, 7)
```

Appendix B - OTSU BINARIZATION ALGORITHM

Source Code 2 – Otsu binarization Algorithm applied in the dataset.

```
# Selecting the image
2 image_file = "Home_cropped" #@param ["Home_cropped", "Callidus_cropped", "
       Saltu_cropped"]
   selected_image = "Image 01" #@param ["Image 01", "Image 02", "Image 03", "Image
       04", "Image 05", "Image 06", "Image 07", "Image 08", "Image 09", "Image 10",
       "Image 11", "Image 12", "Image 13", "Image 14", "Image 15", "Image 16", "
      Image 17", "Image 18", "Image 19", "Image 20", "Image 21", "Image 22", "Image
        23", "Image 24", "Image 25", "Image 26", "Image 27", "Image 28"]
   image_resolutions = ["75ppp", "150ppp", "200ppp", "300ppp", "500ppp", "700ppp", "
       1200ppp"]
6
   # Creating a list of paths to the original images
8 images_paths = [f"{images_path}/{selected_image}/{image_file}/{res}.png" for res
       in image_resolutions]
10 # Initializing the 'images' list before the loop
   images = []
12
   # Checking if images exist
14 timages = []
   for path in images_paths:
16
       img = cv2.imread(path)
       if img is None:
18
           print(f"Error: Image not found at {path}")
           print(f"Image {path} loaded")
20
           images.append(img)
   # Function to binarize and save the image
24 def binarize_and_save(image, output_path, resolution):
       if image is not None:
           # Converting the image to grayscale
26
           gray_image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
28
           # Applying Otsu thresholding to grayscale image
           _, binarized_image = cv2.threshold(gray_image, 0, 255, cv2.THRESH_BINARY
30
               + cv2.THRESH_OTSU)
           cv2.imwrite(output_path, binarized_image)
32
           return binarized_image
34
       else:
           print(f"The image ({resolution}) is empty.")
```

```
36
           return None
38 # Binarizing the images and saving them to the binarized_images list
   binarized_images = []
40 for i, image in enumerate(images):
       resolution = images_paths[i].split("/")[-1].split(".")[0]
       output_path = f"{images_path}/{selected_image}/{image_file}/binarized_{
42
           resolution \}. png"
       binarized_image = binarize_and_save(image, output_path, resolution)
44
46
       binarized_images.append(binarized_image)
48 # Showing the original and binarized images
   titles_originais = [f'{selected_image} - Original ({images_paths[i].split("/")
       [-1].split(".")[0]})' for i in range(len(images))]
50 titles_binarized = [f'{selected_image} - Binarized ({images_paths[i].split("/")
       [-1].split(".")[0]})' for i in range(len(binarized_images))]
   exibir_imagens(images + binarized_images, titles_originais + titles_binarized, 2,
       7)
```

Appendix C - SPECTRAL ANALYSIS ALGORITHMS

Source Code 3 – Fourier Transform Algorithm applied in the dataset.

```
1 # Selecting parameters
   selected_image = "Image 01" #@param ["Image 01", "Image 02", "Image 03", "Image
       04", "Image 05", "Image 06", "Image 07", "Image 08", "Image 09", "Image 10",
       "Image 11", "Image 12", "Image 13", "Image 14", "Image 15", "Image 16", "
       Image 17", "Image 18", "Image 19", "Image 20", "Image 21", "Image 22", "Image
        23", "Image 24", "Image 25", "Image 26", "Image 27", "Image 28"]
3 image_file = "Home_cropped" #@param ["Home_cropped", "Callidus_cropped", "
       Saltu_cropped"]
   color_image = "Color Image" #@param ["Color Image", "Binarized Image"]
   # Image resolutions for analysis
7 image_resolutions = ["75ppp", "150ppp", "200ppp", "300ppp", "500ppp", "700ppp", "
       1200ppp"]
9 # Image paths
   images_path = "/content/drive/MyDrive/ScanImages"
11 if color_image == "Color Image":
       images_paths = [f"{images_path}/{selected_image}/{image_file}/{res}.png" for
           res in image_resolutions]
13 else:
       images_paths = [f"{images_path}/{selected_image}/{image_file}/binarized_{res
           }.png" for res in image_resolutions]
   # Loading images
17 images = []
   for path in images_paths:
19
       img = cv2.imread(path, cv2.IMREAD_GRAYSCALE)
       if img is not None:
           images.append(img)
21
       else:
           print(f"Error: Image not found at {path}")
23
25 def plot_images(images, titles, rows, cols, variables, frequencies):
       plt.figure(figsize=(18, 6 * rows))
       for i in range(len(images)):
27
           plt.subplot(rows, cols, i + 1)
           if i % cols == 1: # Frequency Map
29
               im = plt.imshow(images[i], cmap='jet')
               plt.colorbar(im, ax=plt.gca(), fraction=0.05, pad=0.04)
31
           else: # Original and Reconstructed Image
               plt.imshow(images[i], cmap='gray')
33
           plt.title(titles[i])
           plt.axis('off')
35
```

```
if variables and frequencies and i % cols == 0 and i // cols < len(
               variables) and i // cols < len(frequencies):</pre>
               plt.text(0.5, 1.1, f"Variable: {variables[i // cols]}\nMax Frequency:
37
                    {int(frequencies[i // cols])}",
                         horizontalalignment='center', verticalalignment='center',
                            transform=plt.gca().transAxes)
       plt.tight_layout()
39
       plt.show()
41
43 def save_to_excel(variables, frequencies, excel_path):
       os.makedirs(os.path.dirname(excel_path), exist_ok=True) # Ensure directory
           exists
45
       data = {'Variable': variables, 'Max Frequency': frequencies}
       df = pd.DataFrame(data)
       df.to_excel(excel_path, index=False)
47
       print(f'Data saved to {excel_path}')
49
   def save_frequency_maps(frequency_maps, frequency_map_dir, image_variables):
51
       os.makedirs(frequency_map_dir, exist_ok=True)
       for i, freq_map in enumerate(frequency_maps):
53
           freq_map_path = os.path.join(frequency_map_dir, f"{image_variables[i]}
               _frequency_map.png")
           plt.imsave(freq_map_path, freq_map, cmap='jet')
       print(f'Frequency maps saved to {frequency_map_dir}')
55
57 # Fourier Transform and Frequency Analysis
   def calculate_max_frequency(magnitude_spectrum):
59
       row_center, col_center = np.unravel_index(np.argmax(magnitude_spectrum),
           magnitude_spectrum.shape)
       return magnitude_spectrum[row_center, col_center]
61
   # Configuring layout for data visualization
63 rows = len(images)
   cols = 3  # Original Image, Frequency Map, Reconstructed Image
65 all_images = []
   image_variables = []
67 frequencies = []
   frequency_maps = []
69
   for image_path in images_paths:
71
       image_variables.append(image_path.split("/")[-1].split(".")[0])
73 # Processing each resolution
   for image in images:
       transform_image = np.fft.fft2(image)
75
       center_transform_image = np.fft.fftshift(transform_image)
```

```
77
       magnitude_spectrum = np.log(np.abs(center_transform_image) + 1)
       # Image Reconstruction and Frequency Map Saving
79
       reconstructed_image = np.abs(np.fft.ifft2(np.fft.ifftshift(
           center_transform_image))).astype(np.uint8)
81
       max_frequency = calculate_max_frequency(magnitude_spectrum)
83
       frequencies.append(max_frequency)
       frequency_maps.append(magnitude_spectrum)
85
       all_images.extend([image, magnitude_spectrum, reconstructed_image])
87
   titles = ["Original Image", "Frequency Map", "Reconstructed Image"] * len(images)
89
   # Display images and max frequency
91 plot_images(all_images, titles, rows, cols, image_variables, frequencies)
93 # Save results to Excel
   output_excel_path = os.path.join(images_path, selected_image, image_file, "
       excel_data", f"{selected_image}_frequencies.xlsx" if color_image == "Color
       Image" else f"binarized_{selected_image}_frequencies.xlsx")
95 save_to_excel(image_variables, frequencies, output_excel_path)
97 # Save frequency maps
   frequency_map_dir = os.path.join(images_path, selected_image, image_file, "
       frequency_maps")
99 save_frequency_maps(frequency_maps, frequency_map_dir, image_variables)
```

Source Code 4 – Spectral Density Plot Algorithm applied in the dataset.

```
1 # Selecting parameters
   selected_image = "Image 01" #@param ["Image 01", "Image 02", "Image 03", "Image
       04", "Image 05", "Image 06", "Image 07", "Image 08", "Image 09", "Image 10",
       "Image 11", "Image 12", "Image 13", "Image 14", "Image 15", "Image 16",
       Image 17", "Image 18", "Image 19", "Image 20", "Image 21", "Image 22", "Image
        23", "Image 24", "Image 25", "Image 26", "Image 27", "Image 28"]
3 image_file = "Home_cropped" #@param ["Home_cropped", "Callidus_cropped", "
       Saltu_cropped"]
   color_image = "Color Image" #@param ["Color Image", "Binarized Image"]
5
   # Image resolutions for analysis
7 image_resolutions = ["75ppp", "150ppp", "200ppp", "300ppp", "500ppp", "700ppp", "
   resolutions_dpi = ["75dpi", "150dpi", "200dpi", "300dpi", "500dpi", "700dpi", "
       1200dpi"]
   # Create image paths
11 if color_image == "Color Image":
       images_paths = [f"{images_path}/{selected_image}/{image_size}/{res}.png" for
           res in resolutions]
13 else:
       images_paths = [f"{images_path}/{selected_image}/{image_size}/binarized_{res
           }.png" for res in resolutions]
15
   # Load images, ensuring all are correctly loaded
17 images = [cv2.imread(path, cv2.IMREAD_GRAYSCALE) for path in images_paths]
19 # Define the default output format for images
   pio.kaleido.scope.default_format = "png"
21
   # Function to calculate spectral energy
23 def calculate_spectral_energy(magnitude_spectrum):
       return np.sum(magnitude_spectrum**2)
25
   # Function to compute the maximum frequency
27 def calculate_max_frequency(magnitude_spectrum):
       row_center, col_center = np.unravel_index(np.argmax(magnitude_spectrum),
           magnitude_spectrum.shape)
       return magnitude_spectrum[row_center, col_center]
29
31 # Function to calculate the frequency distribution
   def calculate_frequency_distribution(magnitude_spectrum):
33
       hist, bins = np.histogram(magnitude_spectrum.flatten(), bins=50, range=(0,
           magnitude_spectrum.max()))
       hist = hist / hist.sum()
       predominant_frequencies = bins[:-1][hist > 0.1]
```

```
return hist, bins, predominant_frequencies
37
   # Function to display and save frequency recurrence using Plotly
39 def plot_frequency_recurrence(hist, bins, title, xlabel, ylabel, save_path):
       fig = go.Figure()
41
       bin_width = bins[1] - bins[0]
43
       fig.add_trace(go.Bar(
           x=bins[:-1],
           y=hist,
45
           marker=dict(color='#008000', line=dict(width=0)),
47
           width=bin_width,
           opacity=1.0
49
       ))
       fig.update_layout(
51
           title=title,
           xaxis_title=xlabel,
53
           yaxis_title=ylabel,
55
           plot_bgcolor="#E5E5E5",
           xaxis=dict(showgrid=False),
57
           yaxis=dict(showgrid=True, gridcolor="white"),
           bargap=0,
59
           bargroupgap=0,
       )
61
       fig.write_image(save_path, engine="kaleido")
63
       fig.show()
65 # Directory to save the graphs
   save_dir = os.path.join(images_path, "espectral_density", selected_image,
       color_image, image_size, "original")
67 os.makedirs(save_dir, exist_ok=True)
69 # Apply processing steps for each resolution
   frequencies = []
71 energies = []
   titles = [f"{selected_image} - {res}" for res in resolutions_dpi]
73
   for i, (image, res, res_dpi) in enumerate(zip(images, resolutions,
       resolutions_dpi)):
75
       if image is None:
           print(f"Image {res} was not loaded correctly, skipping...")
           continue
77
79
       transform_image = np.fft.fft2(image)
       center_transform_image = np.fft.fftshift(transform_image)
```

```
81
        magnitude_spectrum = np.log(np.abs(center_transform_image) + 1)
        max_frequency = calculate_max_frequency(magnitude_spectrum)
83
        frequencies.append(max_frequency)
85
        energy = calculate_spectral_energy(magnitude_spectrum)
        energies.append(energy)
87
89
        hist, bins, _ = calculate_frequency_distribution(magnitude_spectrum)
        save_path = os.path.join(save_dir, f"spectral_density_{res_dpi}.png")
91
        print(f"Saving graph at: {save_path}")
93
        plot_frequency_recurrence(hist, bins, f"{image_size} | Spectral Density - {
            titles[i]}: {color_image}", "Frequency", "Recurrence", save_path)
95 # Create DataFrame with results
    data = {
97
       'Resolution': resolutions_dpi,
        'Max Frequency': frequencies,
        'Energy': energies
99
101 df = pd.DataFrame(data)
103 # Path to save the Excel file
    output_excel_path = os.path.join(images_path, selected_image, image_size, "
       excel_data", f"{selected_image}_spectral_analysis.xlsx")
105 os.makedirs(os.path.dirname(output_excel_path), exist_ok=True)
    df.to_excel(output_excel_path, index=False)
```

Appendix D - OCR AND LEVENSHTEIN DISTANCE ALGORITHM

Source Code 5 - OCR and the Levenshtein Distance Algorithm applied in the dataset.

```
# Selecting parameters
2 selected_image = "Image 18" #@param ["Image 01", "Image 02", "Image 03", "Image
       04", "Image 05", "Image 06", "Image 07", "Image 08", "Image 09", "Image 10",
       "Image 11", "Image 12", "Image 13", "Image 14", "Image 15", "Image 16", "
       Image 17", "Image 18", "Image 19", "Image 20", "Image 21", "Image 22", "Image
        23", "Image 24", "Image 25", "Image 26", "Image 27", "Image 28"]
   image_file = "Home_cropped" #@param ["Home_cropped", "Callidus_cropped", "
       Saltu_cropped"]
4 color_image = "Color Image" #@param ["Color Image", "Binarized Image"]
6 # Image resolutions for OCR
   image_resolutions = ["75ppp", "150ppp", "200ppp", "300ppp", "500ppp", "700ppp", "
       1200ppp"]
   # Image paths
10 images_path = "/content/drive/MyDrive/ScanImages"
   if color_image == "Color Image":
       images_paths = [f"{images_path}/{selected_image}/{image_file}/{res}.png" for
12
           res in image_resolutions]
   else:
14
       images_paths = [f"{images_path}/{selected_image}/{image_file}/binarized_{res
           }.png" for res in image_resolutions]
16 # Loading images
   images = []
18 for path in images_paths:
       img = cv2.imread(path, cv2.IMREAD_GRAYSCALE)
20
       if img is not None:
           images.append(img)
22
       else:
           print(f"Error: Image not found at {path}")
   # Directory to save OCR results
26 output_dir = f"/content/drive/MyDrive/ScanImages/OCR_Results/{selected_image}/{
       image_file } / { color_image } "
   os.makedirs(output_dir, exist_ok=True)
28
   # Tesseract configuration to disable dictionary
30 tesseract_config = "--psm 6 load_system_dawg=0 load_freq_dawg=0"
   ocr_texts = []
   # Performing OCR and saving results
34 for i, image in enumerate(images):
```

```
if image is None:
36
           print(f"Error processing image at {images_paths[i]}")
38
       # Performing OCR
       text = pytesseract.image_to_string(image, config=tesseract_config)
40
42
       # Saving OCR result
       output_file = os.path.join(output_dir, f"{image_resolutions[i]}_nd_ocr_result
           .txt")
44
       with open(output_file, 'w') as file:
           file.write(text)
       ocr_texts.append(text)
46
       print(f"Extracted text for resolution {image_resolutions[i]} saved to {
           output_file \}")
48
   # Reading the reference text
50 def read_file(file_path):
       with open(file_path, 'r', encoding='utf-8') as file:
52
           return file.read()
54 reference_path = f"/content/drive/MyDrive/ScanImages/References/{selected_image}
       _reference.txt"
   reference_text = read_file(reference_path)
56
   # Computing Levenshtein distances
58 distances = []
   for i, ocr_text in enumerate(ocr_texts):
60
       distance = levenshtein(reference_text, ocr_text)
       distances.append({
           'Resolution': image_resolutions[i],
62
           'Levenshtein Distance': distance
64
       })
66 # Saving results to Excel
   output_lev_dir = f"/content/drive/MyDrive/ScanImages/Levenshtein_Dist/{
       selected_image}/{image_file}/{color_image}"
68 os.makedirs(output_lev_dir, exist_ok=True)
   output_file_path = os.path.join(output_lev_dir, f"{selected_image}_nd_levenshtein
       .xlsx" if color_image == "Color Image" else f"binarized_{selected_image}
       _levenshtein.xlsx")
70 df = pd.DataFrame(distances)
   df.to_excel(output_file_path, index=False)
72
   print(f"Levenshtein distances saved to: {output_file_path}")
```