Research Article

Jesús Robledano Arillo*

Quality control of digital representations of manuscript texts: proposal of a standards-based method

<https://doi.org/10.1515/opis-2021-0003>

received November 18, 2019; accepted October 5, 2020.

Abstract: This study aims to propose a quality control method for digitized versions of manuscript documents that will be relevant for paleographical and codicological analysis. The methodology applied consisted of a systematic review of papers related to automated analysis of the physical characteristics of handwritings and document supports in the field of digital paleography, as well as of the numerous standards that have been emerging in the field of image engineering for quality assessment in digital image recordings. We also worked with a sample of 275 digital representations of pages or double pages of manuscript documentation dating to between the 12th and 17th centuries. As a result of this study, we propose a taxonomy of physical attributes of the handwritings and of their documentary supports that must be represented in the digital image with a high level of fidelity and without any distortions that could lead scholars to erroneous interpretations of the physical and formal characteristics of the original documents. On the basis of this taxonomy, we identified a set of typical distortions caused by digitization processes that can affect the recording quality of the physical attributes previously proposed, as well as a set of parameters and metrics for measuring quality that can be used to create a sufficiently exhaustive quality model. We also detected a series of limitations which, if not properly managed, can compromise the effectiveness of these types of controls.

Keywords: Image quality assessment; Digitization; Digital image; Digital Paleography; Codicology; Manuscripts; Medieval codices.

1 Introduction

The impact and presence of the digital humanities are experiencing significant growth thanks to the ever-increasing availability of resources in digital format on the Internet and social media. Within this unstoppable trend, the publication of historical sources in digital format has now become a vital element in various fields of scientific research and in teaching of the humanities, especially in the areas of historiographical sciences and techniques. Authors like Sinn and Soares (2014) have highlighted how the collections of digital archives are important sources for historical studies, demonstrating that, while many historians prefer the documents in their original form, their preferences for the sources in digital format are changing as they increase their consultation of digitized archives online. The availability of documentary sources in digital versions online is obviously affecting, and will continue to affect, the humanities in various ways (Brügger, 2016). One of them is the possibility of undertaking meticulous codicological or paleographical studies and using the digital substitutes for the original documents as a source of analysis without the need for the researcher to travel.

In the field of inquiry of the disciplines we just mentioned, the technical quality of the digital images that represent the original documents is essential, as it will guarantee access to an information source that is free of any distortions that might lead the scholar to interpret them erroneously. An accurate image, free of distortion, can be used with

*Corresponding author: Jesús Robledano Arillo, Department of Library and Information Sciences. Universidad Carlos III de Madrid, Getafe, Madrid, Spain, E-mail: jroble@bib.uc3m.es

confidence to obtain precise physical data about the original documents, such as the exact measurements of the text characters and spacing, or of any other element of the handwriting or physical support; changes in the density of the strokes that make up the text; the exact color of the inks; the changes of material from one page to another of a codex; the types of deterioration or manipulation the document has undergone; or the use of inks with a different chemical composition. But digitization of historic heritage documents is not always understood as a systematic representation of the physical and formal characteristics of the original documents; it is more common to see in many projects a biased interest towards representation of the intellectual content, with less attention paid to precise recording of the material characteristics of the document. This criterion can impair the accuracy and fidelity of the capture of many formal and physical aspects that make it possible to reliably undertake paleographical and codicological studies.

One of the scientific fields most impacted by the preoccupation with the quality of the digital surrogates is digital paleography. It has been pointed out the wide range of scientific disciplines, techniques and knowledge that converge in it, beyond the mere pairing of computer science and paleography (Hassner, Sablatnig, Stutzmann, & Tarte, 2012), such as engineering and software development, linguistics, paleography, history, or the disciplines related to the management, conservation and dissemination of documentary heritage with historical value (archival science, library science and museology). In this paper, moreover, we assert the need for image quality control techniques from the field of imaging science to also form part of this array of disciplines. The reason is that accuracy in the capture of all formal and material elements of the documents is essential for application of the automated analysis techniques of digital paleography, such as the identification of scribes, recognition of the manuscript text and its characterization using quantitative measures. Along these lines, Arabadjis, Faigenbaum, Sablatnig, & Stinson (2012) assert that image acquisition should follow a rigorous protocol that includes practices such as the use of color and grayscale charts, documentation of the lighting system used, references to the size of the original objects, information about the image capture equipment, metadata and file-naming conventions.

Our study attempts to define a quality assurance and post-capture quality assessment model for raster digital images of historical written sources that will be used as work objects in paleographical or codicological analysis procedures. This model will be especially useful for application of the automated analysis techniques of digital paleography itself. We first created the proposal of a taxonomy of the formal and physical attributes of the manuscript text and its physical support whose accurate recording must be guaranteed in the digital image of the document. We then attempted to identify a set of distortions of the digital signal that are frequent in capture procedures and which can affect the quality of the digital recording of the previously identified attributes. Lastly, we propose a set of quality control parameters and metrics to help detect the entire range of distortions, such that application of a quality control method that includes them can guarantee that the digital images of the documents are free of them or that their degree of affectation is acceptable for paleographical and codicological studies or for application of the automated procedures of digital paleography. We end with a critical analysis of the proposed model and a series of conclusions that provide a synthesis of our contributions in summary form.

2 Literature review and hypothesis

Understanding the digital image as a substitute for the physical document implies the application of a capture standard that requires that a high level of fidelity in the digital signal be obtained (Robledano Arillo, 2017). This standard makes it possible for the scholar to determine, directly from the digital image, physical attributes relative to its spatial details, geometry, densitometry, colorimetry and, in cases where multispectral capture is used, spectrometry. The pursuit of maximum accuracy in the digital representation of the physical attributes of documents is not new; it has been advocated over the past two decades in the community of documentary heritage experts. In the widely disseminated and influential papers by Kenney (2000) and Rieger (2000), we already see a perfectly defined method for deriving technical capture variables based on physical measurements taken of the document itself to guarantee the fidelity of its digital recording, and the proposal of objective metrics applicable to the quality control procedures for capture devices and digital images.

The objectification of quality assessment of the digital recording has been related in imaging science to the concept of the physical image parameter (Engeldrum, 2004), understood as a function that enables measurement of the fidelity of the digital recording of a determined characteristic of the original physical image, such as its range of optical

densities, its color, its contrast or its spatial detail. Theoretically, good performance of the image capture system in these types of parameters guarantees fidelity in the digital image with respect to the digitized physical object. The degree of fidelity achieved can be quantified objectively, as these parameters represent measurable physical characteristics. In the field of digitization of documentary heritage, interest in the application of physical image parameters in the quality assessment process of the resulting images has been evident since the mid-1990s in the work of Reilly and Frey (1996), Frey and Reilly (1999, 2006) and Williams (2002, 2003, 2010). These studies have tried to build on the standardization work of the ISO/TC42 committee in the field of standards for the evaluation of performance in digital image capture by proposing standardized metrics for quality assessment in digitization process for documents with heritage value. Since the beginning of the current decade, we have been seeing attempts to develop exhaustive and systematic quality control models for digitized versions of heritage documents based on the idea of applying a predetermined set of physical parameters sustained by standardized or commonly used metrics in the context of image engineering (FADGI 2010, 2016; Dormolen, 2012; Nationaal Archief, 2010). The process of standardization in the field of documentary heritage culminated with the publication of the ISO heritage standards (International Organization for Standardization, 2017a, 2017b, 2015; Wueller & Keijser, 2016). This is a set of three standards aimed at unifying metrics, vocabulary, methodologies and tools for specifying and measuring the quality of a digital capture system for raster images applied to cultural heritage documents.

The methods established by the aforementioned specifications and standards are of great utility in the processes of quality assurance and post-capture quality control. The former is used to evaluate the devices and capture procedures and techniques for processing the captured digital signal. The latter is used to determine the quality achieved in the digitized images. The creation of a global framework for quality assessment capable of effectively interrelating the different quality measurement parameters in order to calculate a global quality value of the image has been called a quality model. Already in the late 1980s, we had systematic contributions in the definition of the requirements and constituent elements of these types of models, such as the *Image Quality Circle* (Engeldrum, 1995, 1999, 2004). This proposal is exhaustive enough to lay the groundwork for a global model. These types of models adopt a mathematical formulation in those of their elements that utilize quantitative parameters. Two of the most common mathematical approximations for this type of model are the *Generalized Weighted Mean* and the Minkowski metrics (Engeldrum, 1995). Quality in this type of model can be approximated as a function that calculates the Euclidean distance of the degraded images with respect to an ideal image in an n -dimensional space, the dimensions being the parameters to be measured that were included in the experiments (Robledano Arillo, Moreno Pelayo, & Pereira Uzal, 2016). According to Robledano Arillo, Moreno Pelayo, & Pereira Uzal, 2016 (2016), an approximation of them is shown in equation 1.

$$C(x, y) = \sqrt{\sum_{i=1}^n ((x_i - y_i) \cdot p_i)^2}$$

Equation 1. Quality (C) understood as the Euclidean distance between a digital image (x) and its ideal reference image (y), based on its attributes (i), weighted using their weighting coefficients (p).

These are the models currently available to us in the heritage field and to which we just referred, given the ease of computing the attributes and physical distortions by means of standardized metrics. These models fit into the category of *stimulus models*, as they use physical image parameters instead of perceptual attributes to predict quality. In addition to measurement parameters and their metrics, stimulus models incorporate a set of ranges of permissible values for each of the quality measurement metrics. The quality compliance level is obtained by comparing the results of applying the metrics to their corresponding ranges of permissible values. An important research line in the development of stimulus models that work well at the perceptive level has focused on how to connect the physical performance levels with the perceptual performance levels such that the global quality of an image can be estimated automatically at the perceptive level by means of physical measurements. The term visual algorithm is commonly used to refer to these types of mathematical models. The difficulty of developing visual algorithms is derived from the fact that performance in a set of physical parameters does not necessarily have a proportional correspondence to the subjective perception

of quality due to the multiplicity of elements and complex interrelationships that underly the phenomenon of human visual perception (Engeldrum, 2004; Wang, Bovik, & Lu, 2002).

Despite the limitations of the stimulus models, our hypothesis is that they are very useful for measuring the degree of fidelity with which the physical attributes of the original documents (that are essential for visual or automated analysis in the field of paleography and codicology) have been digitally represented in the image. In these fields access to objective data on the physical characteristics take priority over a global visual perception of similitude between the image of the document and its corresponding view of the original guided by aesthetic criteria.

3 Methodology

In a first work phase, we conducted a literature review to identify the most useful scientific contributions and standards for helping to make our quality model as technologically up to date as possible. This review enabled us to develop a first approximation of the taxonomy of representative attributes for the paleographical analyses we present later in the results section (Table I). We subsequently identified, based on the knowledge obtained in this way and from working with a documentary sample of digitized versions of historic manuscripts, a classification and collection of samples of the typical distortions in digitization processes that can affect them. In this study, we focused only on attributes of the handwriting and of the support, and therefore we did not analyze other physical characteristics that can be specific to old manuscripts, such as illustrations, three-dimensional decorative elements superimposed on the pages or bindings, hanging seals or bindings. Nor did we cover specific characteristics of the multispectral images or specific distortions of this capture technology or its measurement metrics.

The sample was the set of documents digitized in the *CRELOC*¹ and *Digitalización de documentos procedentes de San Pedro de Arlanza conservados en el Archivo de los Duques de Alba*² projects, which make up a 275-page volume of images of pages or double pages of documentation dating to between the 7th and 12th centuries. These documents are characterized by wide paleographic and calligraphic diversity. The sample was captured digitally with a high-performance planetary scanner (Digibook® 10000 RGB made by the company i2S) and a high-performance medium format camera system and digital back (Hasselblad® H1D SLR-D with an HV90x viewfinder and HC 4/120mm-II macro lens, equipped with a high-resolution Phase One® P45+ digital back).

A proposal of objective quality control parameters and quality control metrics was later developed. We applied quality measures to the documentary sample to test the effectiveness of the use of the proposed physical parameters using Imatest Master® software³ and a set of control charts for control of tonal, chromatic, resolute and geometric aspects and uniformity. This procedure helped us to detect a set of problems and inconsistencies in the use of these types of techniques that we have reflected in the results and conclusions.

¹ The reader can get further information about these documents at *CRELOC (Clientela y Redes Locales en la Castilla medieval). Estudio histórico y Tecnologías Documentales. Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica: MEC (Ref. BHA2003-2006) and MICINN (Ref.: HUM206-04554)*. <http://www.creloc.net/>. The documents in particular are: España. Ministerio de Cultura. Archivo Histórico Nacional. Diversos y Colecciones, Códice 91, Cartulario medieval del monasterio de Santa María de Rioseco; España. Ministerio de Cultura. Archivo Histórico Nacional. Diversos y Colecciones, Códice 998, Cartulario medieval del monasterio de San Miguel de Villamayor de Treviño; España. Ministerio de Cultura. Archivo Histórico Nacional. Clero secular-regular. Libro 1375, Libro registro de San Miguel de Villamayor de Treviño.

² The documents included in the study are: ADA (Archivo de los Duques de Alba) C.21-3, dated in 1164; ADA (vitrina 13), dated in 1172; ADA C.145-1, dated in 1177; ADA C.21-4, dated in 1178; ADA C. 145-2, dated in 1214; ADA (vitrina 13) , dated in 1220; ADA Bandejero nº 1, dated in 1224; ADA C.21-9, dated in 1233; ADA C.145-3, dated in 1235; ADA C.145-4, dated in 1235; ADA C.145-5, dated in 1239; ADA C.145-6, dated in 1251; ADA C.145-7, dated in 1256; ADA C. 21-10, dated in 1323.

³ Images and transcripts of the sample or quality test results can be requested at jroble@bib.uc3m.es

4 Findings

4.1 Taxonomy of digital representation attributes in the manuscript text

Characterization of the physical attributes of the handwriting in a manuscript has broad utility in paleographical analysis. Already in the middle 1950s, we had works that promoted the use of handwriting analysis methods based on objective and quantifiable measurements, such as the angle of inclination of the letters with respect to their baseline, the dimensions of the letters and the variations in stroke density. Mallon (1952), for example, proposes the use of these types of measurements for identifying different scribal hands. Numerous subsequent works explore in depth the types of measurement parameters that can be used to identify hands and falsifications (Morris, 2000; Rajan, 2017). In the field of automated image analysis, there is a long history, going back to the 1980s, of proposals of handwriting analysis processes based on the identification and automatic extraction of measurable and quantifiable formal characteristics that are then analyzed computationally in an interrelated manner (Sirat, 1981; Plamondon & Lorette, 1994; Bulacu & Schomaker, 2007; Brink, Smit, Bulacu, & Schomaker, 2012; Rajan, 2017; Dahllof, 2014). Many of these techniques were originally developed initially in the field of expert handwriting analysis, but they have played an important role in the emergence of digital paleography (De Stefano, Maniaci, Fontanella, & Scotto di Freca, 2018a).

These and other studies in the same line of research demonstrate more or less directly the utility of a faithful digital representation of the formal characteristics of the strokes that make up the written characters or graphic representations in the documents, and even of the details of the surface and treatments of the writing supports, in both automated analysis and visual analysis. Some of the metrics or analytical procedures, automated or visual, rely on the identification of subtle differences in the density, shape and width of the stroke even in ranges that are barely perceptible to the naked eye. Therefore, only distortion-free digital representations of the documents, down to a very minute level of detail, can serve as a reliable and robust resource.

The aforementioned automatic analysis methods rely on the concept of metrics. In the context of paleographical studies, this is a measurement that attempts to quantify a particular property of a written character (Rajan, 2017). This author has identified a set of quantitative metrics more directly related to the properties of written characters according to the static shape of the characters (Rajan, 2017). Length, Divergence, Size, Length-breadth Index, Average Curvature, Compactness, Openness, Distinctivity, Ascendancy and Descendancy, and Circularity and Rectangularity. If we break these types of metrics down into the basic elements on which they are based, we can identify a series of underlying basic graphic characteristics of representation of the characters. We have attempted to systematize and present them in Table I. Those same characteristics are also the basis for visual analysis; for example, they are essential for determining the ductus of the handwriting of a text. By applying objective quality measurement techniques to the digital recording, we can even determine with some approximation the degree of distortion we might find in the digital representation of these types of characteristics following the process of digital capture of the document.

Building on the study by Rajan (2017) and other papers in the same line of research (Schomaker, 2008; Brink, 2011; Stokes, 2007, 2009; De Stefano, Maniaci, Fontanella, & Scotto di Freca, 2018a; De Stefano, Maniaci, Fontanella, & Scotto di Freca, 2018b; Papaodysseus, Rousopoulos, Giannopoulos, Zannos, Arabadjis, Panagopoulos, Kalfa, Blackwell, & Tracy, 2014; Liang, Fairhurst, Guest, & Erbilek, 2016; Bulacu & Schomaker, 2003; He & Schomaker, 2015), we attempted to identify the essential graphic elements that can be used as a graphic information source for the application of automated metrics or visual analyses. We subsequently analyzed the most common distortions that can affect them in digital capture procedures and developed a proposal of quality control methods for digital recordings designed to help detect them. The analysis of the documentary sample described earlier in this document helped us to complete this set and develop a taxonomy of formal and physical characteristics of the handwritings and of their documentary supports that must be recorded in the digital images of the documents with the least possible distortion, which we show in Tables I and II. Distortions are identified by a code. The distortion code set appears in the following section.

Table 1: Attributes of the handwriting.

Category	Attributes	Distortions
Spatial	Width of the individual strokes that form the characters or connect them (t-crosses, i-dots, accent marks, curved or straight lines, etc.)	D2, D3, D5, D7, D8, D9, D10, D11, D14, D16, D17, D18
	Length of the individual strokes that form the characters	D2, D3, D5, D7, D8, D9, D10, D11, D14, D16, D17, D18
	Shapes of the strokes that form the characters	D2, D3, D5, D7, D8, D14, D15, D16, D17, D18
	Location of the strokes that form the characters	D7, D16, D17, D18
	Number of strokes that form a character (evidence of the number of hand movements involved in writing a character)	D2, D3, D5, D8, D9, D10, D11, D16, D17, D18
	Angle of inclination of the individual strokes that form the characters	D7, D8, D16, D18
	Height of the characters	D7, D18
	Width of the characters	D7, D18
	Exact shape of the characters	D7, D8, D9, D10, D11, D13, D14, D15, D16, D17, D18
	Angle of inclination of the characters with respect to their baseline	D7, D18
	Spacing between characters	D7, D8, D9, D10, D11, D14, D15, D16, D17, D18
	Proportions of the characters (width/height ratio)	D7, D8, D18
	Spacing between words	D7
	Length of the words	D7
	Proportions of the word (width/height ratio)	D7
	Compactness of the words (ratio between the number of letters and word length)	D7
	Surface area occupied by margins and the space between columns of text	D7
	Distance between text columns	D7
	Line spacing	D7, D18
Tonal	Fluctuations in stroke density	D2, D3, D4, D5, D6, D12, D13, D16, D17, D18
	Density range of the strokes (greater and lesser density)	D2, D5, D6, D12, D18
Chromatics	Color of inks and their variations in any of the degrees of density of the strokes (colorimetric data according to a CIE standard, such as CIELAB or CIE XYZ, or to an unambiguous color space such as ROMM RGB or Adobe RGB (1998))	D1, D2, D3, D4, D5, D6, D12, D13, D14, D15, D17
	Spectral data of the inks and their variations (this can help to distinguish their nature and chemical composition)	D1, D2, D3, D4, D5, D6, D12, D13, D14, D15, D16, D17

4.2 Taxonomy of typical digitization distortions that can affect representation of the identified physical attributes

We understand distortion to mean an alteration of the fidelity ratio, understood as a correspondence at the physical level which must exist between the original element and its digital representation. Various approximations have emerged that have attempted to create a complete and systematic taxonomy of distortions that affect the quality of the digital signal and which can appear in document digitization processes, such as the *Taxonomy of Digital Imaging Performance* (FADGI, 2010, p. 7) or those proposed by Burns (2015) and Frey and Reilly (1999, 2006). From these types

Table 2: Attributes of the support.

Category	Attributes	Distortions
Spatial	Texture details of the support (these can be used to identify the hair and flesh sides of a parchment support, for example, or to distinguish materials and treatments applied in the fabrication of the support)	D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D12, D13, D14, D15, D16, D17, D18, D19
	Ruling of the support (lined)	D2, D3, D5, D7, D8, D9, D10, D11, D12, D13, D14, D15, D16, D17, D18
	Pricking of the support (perforation or puncturing)	D2, D3, D5, D7, D8, D9, D10, D11, D12, D13, D14, D15, D16, D17, D18
	Details of the manufacture of the codex or bound document based on elements visible on the surface of the document when opened to double page (observable based on the representation of the binding, the space between one page and the next, the order of the pages or the texture of the surface of each page, and which are helpful for determining aspects such as how many times the initial sheet was folded, identification of joined quires or the how the quires are joined)	D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D13, D14, D15, D16, D17, D18
	Alterations caused by the loss or addition of sheets (presence of stubs from cut out sheets, presence of stubs or guards for the addition of sheets, presence of binding reinforcements, details of the boards, covers, etc.)	D7, D8, D9, D10, D11, D13, D14, D15, D16, D17, D18
	Three-dimensional details of the surface of the support (dry seals in relief or other thin elements superimposed or marked in relief)	D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D13, D14, D15, D16, D17, D18
	Watermarks on the paper (requires digital recording using translucent light)	D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D13, D14, D15, D16, D17, D18
	Paper molding details (surface texture of the paper support, they require translucent light for best digital recording, although they can also be captured with a certain level of detail using reflected light)	D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D13, D14, D15, D16, D17, D18
	Structure of the paper pulp used in fabricating the support (requires translucent light for digital recording)	D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D13, D14, D15, D16, D17, D18
	Shape of the pages	D7
	Absolute dimensions of the pages	D7
	Absolute dimensions of the surface occupied by the handwriting and by the graphics present on the page	D7
	Detail of fingerprints present on the surface of the support (due to ink stains from pressure of fingertips stained with wet ink)	D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D13, D14, D15, D16, D17, D18
	Detail of other types of stains and deterioration present on the surface of the support	D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D13, D14, D15, D16, D17, D18, D19
	Alterations of the support involving mechanical actions that leave a mark (tears, gouges, scrapes, cuts, holes, etc.)	D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D13, D14, D15, D16, D17, D18, D19
Tonal and chromatic	Range of densities of the surface of the support considering any of its elements (handwriting, graphics, stains and other forms of deterioration)	D2, D5, D6
	Optical density variations on the surface of the support	D2, D3, D4, D5, D6, D18
	Signs of natural or intentional degradation that involve a change in color or optical density on the surface of the support (foxing, ink corrosion, mold, insect excrement, acidification or oxidation of the paper, ink fading, etc. They can be useful for dating or identifying the type of writing support or detecting changes in the support or falsifications)	D1, D2, D3, D4, D5, D6, D8, D9, D10, D11, D12, D13, D14, D15, D16, D17, D18
	Color of the support and its variations on the surface (colorimetric or multispectral data)	D1, D6, D15, D17, D18

of systematizations, we can abstract six major categories of digital recording problems according to the main graphic information type affected: chromatic, tonal, geometric, spatial detail, noise and artifacts. The oversimplification of this classification makes it practical for presenting digital recording quality problems succinctly, but it entails the limitation that, in many of the effects discussed later on in this document, various types of information are involved which could be classified in several of these categories at once. To overcome this drawback, we have based the classification of the distortions on the information type most severely affected. Likewise, we have simplified the number of distortions so that we can objectively analyze any defect for which we can supply objective quality measurement metrics.

The information about the physical attributes of the original document that are affected by each distortion are reflected in Tables I and II, where the distortions are identified by their corresponding code according to the list shown below.

- a) Color distortions. These are problems that affect accurate recording of the color of the physical elements captured in the digital image.
 - D1. Lack of accuracy at the colorimetric or spectral level in the digital representation of the color. This is manifested in a discrepancy between the colorimetry or spectrometry of the surface of the original and that of its digital recording which precludes correct visualization or reproduction of the color or accurate analysis of the color variations that the inks or surface of the support may display. Color is a plastic attribute that is important not only for the most precise understanding of the physical characteristics of the original document but also for interpretation of the plastic or symbolic message conveyed through the overall aesthetic of the page or of any illustrations or decorative motifs it may present.
- b) Tonal distortions. These distortions affect accurate recording of the optical density changes presented by the surface of the document in any of its elements (handwriting, illustrations, support, etc.). When the digital representation of the changes in stroke density does not correspond to the original text, or when it has not been possible to record densities, it is difficult to obtain accurate information about the writing process, such as the pressure and movement of the hand, which can compromise the accuracy of the paleographical analysis.
 - D2. Failure to reproduce the range of optical densities in the document. The range of densities refers to the highest and lowest densities that the surface of the document presents in any of its elements. If this is not reproduced in the digital image, there will be tonal information that is not captured on one or both ends of the scale of densities the document presents (in the darkest and lightest parts).
 - D3. Poor tonal sensitivity of the capture device. Tonal sensitivity refers to the capacity of the digital capture system to detect and record subtle changes in light intensity. Poor tonal sensitivity can result in failure to record changes in density that are small but of great importance for paleographical analysis. Figure 1 illustrates this problem. The top image shows a capture with good tonal sensitivity, which makes it possible to appreciate the difference in density of the outline with respect to the interior of the stroke, something that is impossible to appreciate in the image on the bottom, which has less tonal sensitivity.
 - D4. Absence of linearity. As a result of this distortion, the contrast of the surface of the original document does not correspond to that of the digital image, such that slight density changes of the surface of the original may be represented as sharp differences or vice versa.
 - D5. Clipping of lights and darks due to incorrect post-processing of the capture. This tends to be caused by overly aggressive contrast adjustments.
 - D6. Lack of uniformity in the intensity of the capture lighting. The intensity of the illuminant is not uniform across the entire capture area. This problem can be caused by a variety of factors: poor quality in the lenses of the optical system of the capture device (vignetting), use of a non-uniform or defectively configured lighting system, or by the irregular curvature of the page at the moment of capture. This problem affects recording of the color and densities of the surface of the document. Given that, in general, when we use a digital image of a document, it is because we do not have the original in front of us, we have to assume that the digital image was captured with perfectly uniform light intensity and that this intensity is not distorted by defects in the optics used in the capture. Otherwise, it will be impossible for us to accurately interpret the color or tonal characteristics of the entire surface of the document.

In Figures 1 and 2, we see examples of typical information losses due to tonal problems in the digital capture and processing of images of documents. In the incorrect recordings, the growth in the density value of the densest zones has distorted the characteristic density pattern of the scribal hand.

Figure 1: Top, accurate recording of the density levels of the original writing. Bottom, inaccurate recording of stroke densities due to excessive contrast adjustment.

Figure 2: Two examples of inaccurate recording of the stroke densities. This defect was caused by the low dynamic range and poor tonal sensitivity of the capture system. The problem was exacerbated by an excess of lateral chromatic aberrations.

Figure 3: Example of inaccurate recording of small spatial details and density variations of the stroke of the writing and of the support due to focusing problems accentuated by incorrect exposure of the original during the digital photographic scan.

- c) Geometric optical distortions. These are due to factors like the lack of correction of the lenses of the optical devices, warping of the pages at the moment of capture, or lack of parallelism between the surface of the page and the focal plane of the capture device.
- D7. Geometric and dimensional deformations. These are reflected as alterations in the shape and proportions of the pages and their content. When they are caused by correction defects of the lenses, they appear as barrel or pincushion type deformations.
- d) Distortions in the capture of spatial detail. This problem affects recording of the detail present in the original, such as the strokes of the writing or the texture pattern of the surface of the support.
- D8. Lack of resolution capacity. This is produced when the digital capture has not recorded the smallest capturable detail with the spatial sampling value applied. It can be due to various factors: poor quality of the optics attached to the capture devices, inadequate focus, insufficient depth of field or movement of the capture device or document during the digitization process.
- D9. Problems or irregular focus across the surface of the page due to its excessive curvature.
- D10. Lack of focus due to defective focus at the moment of capture which affects the entire surface of the page.
- D11. Geometric optical aberrations (spherical aberration, astigmatism, coma, field curvature). These appear as a lack of sharpness that can unevenly affect the digitally captured surface. They are caused by the optics and mirrors attached to the digital cameras and scanners due to insufficient correction during their fabrication process.
- e) Noise. Random variation in the color or brightness of the image pixels, which do not correspond to original information.
- D12. Color or luminance noise. This appears as a distracting random dispersion pattern of colored or bright points that do not correspond to the color of the surrounding pixels.
- f) Artifacts. These are graphic patterns that appear in the digital images but which do not correspond to the original information or which, in doing so, show obvious deformation. They affect the tonal and chromatic and information and spatial detail. There is a wide range of artifact types, and so we have only included the most frequent ones we have identified in the documentary sample used in the empirical work.
- D13. Presence of dirt on the sensor of the capture device or on the document. This appears as more or less unfocused dots or lines dispersed on the surface of the image.
- D14. Chromatic optical aberrations. These are produced when the optics of the capture device are incapable of locating on the same point the different wavelengths of the light that the image carries; this creates an outline of red, green, blue, cyan, magenta or yellow around the edge of the objects present in the image.
- D15. Misalignment of color channels or spectral bands (color misregistration). This produces a misalignment of the channels or bands of the image, generating a distortion in color and shape that is very obvious on the edges of the elements represented.

- D16. Aliasing. This is produced in the digital recording of small repetitive graphic patterns when the sampling resolution is insufficient. These patterns disappear in the digital image, replaced by a moiré pattern or similar.
- D17. Compression artifacts. These are produced by excessive compression of the image using compression algorithms with loss. They tend to appear as distortions on the edges of the shapes and lines, loss of small spatial detail and even as alteration of the color in the zones most affected.
- D18. Banding. Appearance in the digital image of horizontal or vertical lines that present a recording that does not correspond to the original information.
- D19. Defective pixels. Due to defects in the fabrication of the sensor, some pixels do not record the signal properly, with their values being completely unexpected.

In addition to the problems reflected in the previous list, it is common to find the defect of an excessive degree of inclination of the document at the time of its digital capture. We have not included it in the taxonomy of distortions as it is considered a problem that is not related to the performance of the device or digital processing but to a poor positioning of the document. We also consider that it not alters any formal or physical attribute of the document. This problem requires an additional post-capture digital processing for its elimination that can lead to changes in some of the physical characteristics of the document, although usually not very relevant. As Brink, Niels, Batenburg, Heuvel and Schomaker (2011) have shown, the angle of inclination as an isolated factor is not essential for good performance analysis of writer verification.

4.3 Taxonomy of physical parameters and their capture metrics. Proposal of pre- and post-capture quality control methods and metrics

In Tables III to V, we present a structured list of physical image parameters and their metrics. We have included the ones we considered most useful for creation of a quality control model suitable for the quality assurances processes for document images according to the exacting standards required for paleographical and codicological analysis. In presenting them, we identify the following data: parameter type, parameter, distortion(s) related to them, standards or metric specifications considered suitable for measurement of the parameter, procedure or metric recommended for their quality assessment and control charts we can use to apply the metric in an automated or manual manner.

We have sought to ensure that all these parameters can be measured with the commercial or free applications most commonly used in the cultural heritage field: FADGI OpenDICE and AutoSFR⁴, Golden Thread⁵, Imatest Master⁶, and IQ Analyzer⁷, although other applications can also be used. When a metric does not appear in any of the five products just listed, we have indicated it in the Parameter column.

Tables III, IV and V show that we currently have a wide variety of standards, measurement metrics, commercial and free software applications, and control charts to enable us to develop systematic controls of quality that can help control maximum fidelity in the digital representation of physical characteristics of documents that are relevant for paleographical and codicological studies. Nonetheless, application of all the parameters and metrics proposed has led us to identify certain limitations.

⁴ OpenDICE is a tool for analysis and measurement of conformity according to the FADGI (FADGI, 2010, 2016) quality specifications, and it is distributed free of charge as open-source code by the Federal Agencies Digitization Guidelines Initiative. It can be downloaded from <http://www.digitizationguidelines.gov/guidelines/digitize-OpenDice.html>. AutoSFR is a product of the same initiative and can be downloaded from the same site.

⁵ Golden Thread[®] is a commercial software application made by the company Image Science Associates and is specifically designed to support quality assessment processes of capture devices and digital images in heritage documentation projects. Further information about this product can be obtained at http://www.imagescienceassociates.com/mm5/merchant.mvc?Screen=CTGY&Store_Code=ISA001&Category_Code=GT.

⁶ Imatest Master[®] is a commercial software application from Imatest for quality analysis of digital capture devices intended for the image engineering field, although fully applicable in the area of digitization of heritage documents. Extensive information about this product is available at <http://www.imatest.com/>.

⁷ IQ Analyzer[®] is a commercial software application from Image Engineering focused on quality assessment of image capture devices, but specially adapted for the quality specifications of the ISO 19263 and 19264 standards. Further information can be obtained at <https://www.image-engineering.de/products/software/376-iq-analyzer>.

Table 3: Chromatic and tonal physical image parameters, their distortions and quality control procedures.

Parameter type	Parameter	Distortions	Sources of measurement procedures and metrics	Recommended procedures and metrics	Control charts
Chromatic	Accuracy in color reproduction or registration	D1	ISO 19264, ISO/CIE11664-6, ISO/CIE 11664-4, ISO 17321-1, ISO 17321-2, ISO 17321-3, ISO 13655	ISO 19264-Color Reproduction (CIEDE 2000 Delta E), CIE 1976 Delta-C (a^*b^*), CIE 1976 Delta-E (a^*b^*), CIEDE2000 Delta C, CIEDE2000 Delta E	UTT*, ColorCheckr Classic®**
	White balance	D1	ISO 19264, ISO 14524	ISO 19264-White Balance	UTT
Tonal	OECF (Opto Electronic Conversion Function)	D4, D2, D5	ISO 14524, ISO 21550, ISO 19264	ISO 19264-Tone reproduction, Imatest Density response,	UTT, Q13***
	Gain modulation	D4	ISO 19264	ISO 19264- Gain modulation, Imatest® Density Response	UTT, ColorCheckr Classic®
	Dynamic range	D2, D5	ISO 19264, ISO 21550, ISO 15739	ISO 19264-Dynamic range	UTT, Q13
	Local Contrast	D3	ISO 14524, ISO 21550	Imatest® Local Contrast (Density response slope)	UTT, Q13
	Uniformity	D6	ISO 19264, ISO 17957	ISO 19264-Illumination non-uniformity, ISO 19264-Chrominance non-uniformity	UTT, an evenly illuminated uniform object

* UTT (Universal Test Target). Its specifications are accessible at <http://www.universaltesttarget.com/>

** ColorCheckr Classic is distributed by X-Rite. Its specifications are accessible at <https://xritephoto.com/colorchecker-classic>

*** Q-13 Chart. Tiffen® (Kodak®) Color Separation Guide with Grey Scale.

In the first place, we must stress that control methods based on measurement of parameters against control charts placed on the margins of the documents, or which are captured previously encompassing the entire surface where the original documents to be digitized will be placed, do not allow total control of quality assurance. This is because there are determining quality factors that are activated during the capture process and which, moreover, vary from one set of captures to another. These factors are very difficult to control, as they depend on the action of the capture operator or on other unpredictable circumstances, such as involuntary movements or vibrations of the capture platform, of the document itself or of the device; degree of inclination or warping of the page; deficient focus; diachronic variations in the intensity of the capture lighting; changes in the pattern of noise production due to diachronic differences of the temperature of the sensor; noise caused by electromagnetic contamination; accumulation of dirt on the documents, lenses or sensor that is variable according to the capture time of a batch; etc. Nonetheless, post-capture control, image by image, does make it possible to discard images where factors of this type have resulted in quality problems.

Another problem that can skew the pre- or post-capture quality assurance controls is that the surface of the documentary supports or the original writing inks do not, in general, have the same physical properties as the surfaces and pigments or inks used to produce the quality control charts used in the tests. This can lead to skewed results in chromatic or tonal quality measurements, as these are more related to the capture quality of the control charts than to that of the original document itself. For example, the color profiles of the capture devices are generally created using color patterns from standardized color charts, the spectral response of which may not coincide with that of the document, thereby compromising accuracy in color and tone in its printed reproduction or visualization.

We have also considered that the curvature that the surface of the document page may present at the moment it is digitized will affect the surface of the document differently in terms of focusing quality, the presence in the digital image of diverse types of geometric distortions, or problems of uniformity of lighting. Therefore, post-capture control by means of charts placed next to the margins of the documents may not be sufficiently representative of the quality of

Table 4: Physical image parameters related to optical distortions and spatial detail and its quality control procedures.

Parameter type	Parameter	Distortions	Sources of measurement procedures and metrics	Recommended procedures and metrics	Control charts
Optical distortions	Geometric distortion	D7	ISO 19264, SMIA TV Distortion, ISO 17850, ISO 9039	ISO 19264-Distortion, Imatest® SMIA TV Distortion	UTT, Squares grid, Distortion grid
	Curvature degree of the document (It measures the relationship between the actual size of the horizontal dimension of the page to be captured and the size with which it is registered in the digital image). It is not usual to find this metric in image quality control applications.	D7		Curvature degree = $I / R * 100$ Where I = size of the horizontal dimension of the page in the digital image; and R = real size of the horizontal dimension of the page.	
Spatial detail	Sampling rate	D8	ISO 19264	ISO 19264–Sampling rate	UTT
	Limiting resolution	D8, D9, D10, D11, D17	ISO 19264, ISO 12233, ISO 16067-1	ISO 19264–Limiting resolution	UTT
	Sharpening (software focus).	D16, D4, D3, D5	ISO 19264, ISO 12233, ISO 16067-1	ISO 19264–Sharpening	UTT
	MTF50% (optical resolution value at MTF 50%).	D8, D9, D10, D11, D17	ISO 12233, ISO 16067-1	ISO 19264–MTF 50	UTT
	Noise frequency spectrum (the noise versus the spatial frequency value. It illustrates the possible application of low-pass filters (antialiasing and noise reduction) that manufacturers of digital cameras and scanners, or RAW processing software, apply to images. These filters can cause the loss of fine spatial detail in uniform areas of the image or in low contrast transitions)	D8, D17	Imatest®	Imatest®-Noise frequency spectrum	UTT

Table 5: Physical image parameters related to noise and artifacts and their quality control procedures.

Parameter type	Parameter	Distortions	Sources of measurement procedures and metrics	Recommended procedures and metrics	Control charts
Noise	Noise	D12	ISO 19264, ISO 15739, ISO 21550	ISO 19264-Noise	UTT, Q13
Artifacts	Banding	D18	ISO 19264	ISO 19264-Banding	UTT
	Defect pixels	D19, D13	ISO 19264, ISO 9241-302	ISO 19264-Defect pixels	UTT
	Color mis-registration	D15	ISO 19084, ISO 19264	ISO 19264-Color mis-registration	UTT
	Chromatic aberration	D14		Imatest®-SFR area of crossing chromatic aberration, IQ Analyzer®-Distortion	UTT

the recording in other zones of the captured surface and, as a result, may not provide representative data on the overall quality of their capture. Another problem that can skew the quality assurance control is the existence of temporal variations in some quality factors, such as the uniformity of lighting (Robledano Arillo & Navarro Bonilla, 2017), which may be due to elements that are difficult or impossible to control.

Research on digital image processing of documents has generated a comprehensive set of algorithms for detecting and correcting quality problems that are easily detectable in an automated way, such as problems of uniformity of lighting (Zeng, 2017; Robledano Arillo & Navarro Bonilla, 2017; Yu, 2004; Arjona, 2004; Schoonees & Palmer, 2009; Shi & Lu, 2009), or geometric distortions caused by warping of the pages of the documents at the moment of capture (Zhang, Zhang, & Tan, 2008; Brown & Seales, 2004, 2001; Cao, Ding, & Liu, 2003; Meng, Pan, Xiang, & Duan, 2012; Meng, Xiang, Pan, & Zheng, 2016). But in the context of paleographical or codicological studies, it is necessary to guarantee, prior to their application, their innocuousness with respect to alteration of other physical characteristics of the originals.

We must also stress limitations that are inherent to the principles of the stimulus models we have discussed in this paper, whose image quality control is based on physical image parameters measured on test targets. Quality controls by themselves do not measure the registration quality needs of the original documents, but rather the capacity of the scanner or digital camera to meet the quality aims of the digitizer. Therefore, we have a bias involved in this process, the objectives of the digitizer. That is, what those responsible for the digitization project consider relevant to be registered or feasible with the project budget.

Other limitation is the determination of ranges of permissible values for the metrics that make it possible to obtain the quality value in the different physical parameters. Some widely used specifications and standards —like ISO 19263 and ISO 19264, FADGI (2010) or *Metamorfoze* (Dormolen, 2012)— provide tables showing these values. The ISO 19263 heritage standard itself describes how the ranges of values for each physical parameter to be measured were established: “These aims and tolerances have been derived via extensive testing and feedback from cultural heritage imaging users and program managers” (International Organization for Standardization, 2017a, p. 7). But this conception is overly simplistic. In reality, we do not have systematic information about these studies or, as a result, about the correspondence of the values of the ranges used to establish the quality ratings with the levels of perceptual affectation of the problem, or about the importance that the loss of accuracy indicated by these ranges has for systematic studies on the digital images of the documents. We need to understand that the ISO heritage standards have inherited those values from the aforementioned FADGI and *Metamorfoze* specifications, which did not specify how they were obtained either. A possible solution in this regard is to carry out a rigorous analysis procedure on the original documents themselves based on a representative sample, contrasting physical information with values of physical image parameters or capture variables that allow it to be reproduced correctly. Through this procedure it will be possible to determine with greater precision the values to be achieved in the quality control. Even so, we will always have a subjective factor involved or even limitations caused by the feasibility of certain information records.

5 Conclusions

The paleographical studies based on direct visual or manual analysis of the raster digital images of the original documents and automated analysis by applying procedures developed in the field of digital paleography have been gaining broad relevance in recent years. Both types of procedures require the application of digital capture and processing techniques that produce digital images that are highly representative of their originals at the densitometric, colorimetric and spatial level. The rigor and effectiveness of these techniques lies in the application of quality control procedures to the capture process based on physical parameters on the basis of sufficiently reliable metrics. This is the only way that analysis of the fidelity of the digital recording can be made objective with respect to its original and to have assurances that it is possible to use the digital substitutes with full confidence in the analyses carried out on the documents and their results.

In this paper, we have demonstrated that we already have enough quality control tools and procedures that can be applied with low resource consumption to be able to systematize the captures and guarantee their level of representativeness in a high percentage of them. Nonetheless, some contingencies that are very complicated to control

during capture can ultimately cause the performance in the images of the documents to not be exactly equivalent to that obtained in the quality control processes due to the different nature of the control charts used for the quality assessments and the impossibility of achieving complete control of certain mechanical aspects that are fundamental for correct performance of the capture procedures. Likewise, we must stress limitations in the principles and foundations of the type of quality model on which these control systems are based, the commonly called stimulus models. One of the most controversial aspects of these models is the difficulty or lack of a systematic approach in the development of the ranges of values for the different performance quality ratings in the different control physical parameters.

To improve quality control procedures, further research is needed in areas such as the fabrication of test targets adapted to the spectral characteristics of the surfaces of the supports and inks of the original physical documents, or systems to detect or correct instances of irregular and local effects of problems related to focusing, geometric distortion or uniformity in lighting. Also, it is necessary to advance in the development of tests that allow determining with precision and objectivity the ranges of values in capture variables and physical image parameters that are suitable to achieve the complete and correct registration of the relevant physical attributes of the documents. These tests should also help to know what parameters or procedures can avoid or limit distortions for different capture conditions.

Funding Information: Publication was made possible by the funding of the R&D+i project within the *Programa Estatal de Investigación, Desarrollo e Innovación* oriented towards the *Retos de la Sociedad, Vocabularios para una Red de Archivos y Colecciones de Media Art y sus efectos: Metaliteracy y Turismo del Conocimiento*. Reference: HAR2016-75949-C21-R. Funding Body Ministerio de Economía y Competitividad.

Conflict of Interest: Author states no conflict of interest.

List of referred standards

- ISO 9039:2008. Optics and photonics — Quality evaluation of optical systems — Determination of distortion.
- ISO 9241-302:2008. Ergonomics of human-system interaction — Part 302: Terminology for electronic visual displays.
- ISO/CIE 11664-4:2019 [CIE LEAD]. Colorimetry — Part 4: CIE 1976 L*a*b* colour space.
- ISO 12233:2017. Photography — Electronic still picture imaging — Resolution and spatial frequency responses.
- ISO 13655:2017. Graphic technology — Spectral measurement and colorimetric computation for graphic arts images.
- ISO 14524:2009. Photography — Electronic still-picture cameras — Methods for measuring opto-electronic conversion functions (OECFs)
- ISO 15739:2017. Photography — Electronic still-picture imaging — Noise measurements.
- ISO 15739:2017. Photography — Electronic still-picture imaging — Noise measurements.
- ISO 16067-1:2003. Photography — Spatial resolution measurements of electronic scanners for photographic images — Part 1: Scanners for reflective media.
- ISO 17321-1:2012. Graphic technology and photography — Colour characterisation of digital still cameras (DSCs) — Part 1: Stimuli, metrology and test procedures.
- ISO/TR 17321-2:2012. Graphic technology and photography — Colour characterization of digital still cameras (DSCs) — Part 2: Considerations for determining scene analysis transforms.
- ISO/TR 17321-3:2017. Graphic technology and photography — Colour characterization of digital still cameras (DSCs) — Part 3: User controls and readouts for scene-referred imaging applications
- ISO 17850:2015. Photography — Digital cameras — Geometric distortion (GD) measurements.
- ISO 17957:2015. Photography — Digital cameras — Shading measurements.
- ISO 19084:2015. Photography — Digital cameras — Chromatic displacement measurements.
- ISO/TR 19263-1:2017. Photography — Archiving systems — Part 1: Best practices for digital image capture of cultural heritage material.
- ISO/TS 19264-1:2017. Photography — Archiving systems — Image quality analysis — Part 1: Reflective originals.
- ISO 21550:2004. Photography — Electronic scanners for photographic images — Dynamic range measurements.
- SMIA (Standard for Mobile Imaging Architecture) 1.0 Part 5: Camera Characterisation Specification. 2004. Nokia & ST Microelectronics, p. 61.

References

- Arabadjis, D., Faigenbaum, S., Sablatnig, R., & Stinson, T. (2012). Working Groups. 4.1 Acquisition of Images. Seminar 18–21. September, 2012. *Computation and Palaeography: Potentials and Limits. Dagstuhl Reports*, 2(9), 184–199. Retrieved from https://www.academia.edu/2575333/Computation_and_Palaeography_Potentials_and_Limits?auto=download
- Arjona, M., Novella, M.L. Lasarte, M., Pujol, J., & Vilaseca, M. (2004, May 19–21). Caracterización espacial de dispositivos de adquisición de imágenes para su utilización en la medida de color. Paper presented at VII Congreso Nacional de Color, Pamplona: Universidad de Navarra, 173–174.
- Brink, A. A. (2011). Robust and applicable handwriting biometrics (PhD thesis). Groningen: University of Groningen.
- Brink, A. A., Niels, R. M. J., van Batenburg, R. A., van den Heuvel, C. E., & Schomaker, L. R. B. (2011). Towards robust writer verification by correcting unnatural slant. *Pattern Recognition Letters*, 32(3), 449–457. <https://doi.org/10.1016/j.patrec.2010.10.010>
- Brink, A. A., Smit, J., Bulacu, M. L., & Schomaker, L. R. B. (2012). Writer identification using directional ink-trace width measurements. *Pattern Recognition*, 45(1), 162–171. <https://doi.org/10.1016/j.patcog.2011.07.005>
- Brown, M. S., & Seales, W. B. (2001). Document restoration using 3D shape: a general deskewing algorithm for arbitrarily warped documents, In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV '01)*, 2. Paper presented at International Conference on Computer Vision (ICCV '01), 367–374. <https://doi.org/10.1109/ICCV.2001.937649>
- Brown, M. S., & Seales, W. B. (2004). Image restoration of arbitrarily warped documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10), 1295–1306. <https://doi.org/10.1109/TPAMI.2004.87>
- Brügger, N. (2016). Digital Humanities in the 21st Century: Digital Material as a Driving Force. *Digital Humanities Quarterly*, 10(3), 39–53.
- Bulacu, M., & Schomaker, L. (2003). Writer style from oriented fragments, In: N. Petkov, & M. Westenberg (Ed.). *Proceedings of the Tenth International Conference on Computer Analysis of Images and Patterns (Groningen - The Netherlands, August)*. Paper presented at Tenth International Conference on Computer Analysis of Images and Patterns, Groningen: Springer, 460–469. https://doi.org/10.1007/978-3-540-45179-2_57
- Bulacu, M., Schomaker, L., & Vuurpijl, L. (2003). Writer-identification using edge-based directional features, In: *Proceedings of the Seventh International Conference on Document Analysis and Recognition*. Paper presented at Seventh International Conference on Document Analysis and Recognition, Edinburgh: IEEE, 937–941. <https://doi.org/10.1109/ICDAR.2003.1227797>
- Bulacu, M., & Schomaker, L. (2007). Text-independent writer identification and verification using textural and allographic features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4), 701–717. <https://doi.org/10.1109/TPAMI.2007.1009>
- Burns, P. D. (2015). Image Quality Concepts. In M. Kriss (Ed.), *Handbook of Digital Imaging* (pp. 325–372). Hoboken, NJ: John Wiley & Sons, Ltd.
- Cao, H., Ding, X., & Liu, C. (2003). A cylindrical surface model to rectify the bound document image, In: *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV '03)*. Paper presented at 9th IEEE International Conference on Computer Vision (ICCV '03), Nice (France): IEEE, 228–233.
- Dahllof, M. (2014). Scribe attribution for early medieval handwriting by means of letter extraction and classification and a voting procedure for larger pieces. In: *Proceedings of the 22nd International Conference on Pattern Recognition*. Paper presented at 22nd International Conference on Pattern Recognition, Stockholm: IEEE Computer Society, 1910–1915. <https://doi.org/10.1109/ICPR.2014.334>
- De Stefano, C., Maniaci, M., Fontanella, F., & Scotto di Freca, A. (2018a). Reliable writer identification in medieval manuscripts through page layout features: The “Ávila” Bible case. *Engineering Applications of Artificial Intelligence*, 72, 99–110. <https://doi.org/10.1016/j.engappai.2018.03.023>
- De Stefano, C., Maniaci, M., Fontanella, F., & Scotto di Freca, A. (2018b). Layout measures for writer identification in mediaeval documents. *Measurement*, 127, 443–452. <https://doi.org/10.1016/j.measurement.2018.06.009>
- Dormolen, H. (2012). *Metamorfoze Preservation Imaging Guidelines. Image Quality. Version 1.0, January 2012*. The Hague, NL: National Library of the Netherlands; https://www.metamorfoze.nl/sites/metamorfoze.nl/files/publicatie_documenten/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf
- Engelrum, P. G. (1995). A framework for image quality models. *The Journal of Imaging Science and Technology*, 39(4), 312–318.
- Engelrum, P. G. (1999). Image Quality Modeling: Where Are We? In: *IS&T's 1999 PICS Conference*. Paper presented at IS&T's 1999 PICS Conference, Savannah, Georgia: The Society for Imaging Science and Technology, 251–255.
- Engelrum, P. G. (2004). A Theory of Image Quality: The Image Quality Circle. *The Journal of Imaging Science and Technology*, 48(5), 446–456.
- FADGI [Federal Agencies Digitization Initiative] - Still Image Working Group (2010). *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files for the Following Originals - Manuscripts, Books, Graphic Illustrations, Artwork, Maps, Plans, Photographs, Aerial Photographs, and Objects and Artifacts*. http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf
- FADGI [Federal Agencies Digitization Initiative]- Still Image Working Group (2016). *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files*. http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image_Tech_Guidelines_2016.pdf
- Frey, F., & Reilly, J. (1999). *Digital Imaging for Photographic Collections: Foundations for Technical Standards*. Rochester, NY: Image Permanence Institute.

- Frey, F., & Reilly, J. (2006). *Digital Imaging for photographic collections: foundations for technical standards* (2nd ed.). Rochester, NY: Image Permanence Institute.
- Hassner, T., Sablatnig, R., Stutzmann, D., & Tarte, S. (eds.) (2012). *Report from Dagstuhl Seminar 14302 Digital Palaeography: New Machines and Old Texts*. http://drops.dagstuhl.de/opus/volltexte/2014/4793/pdf/dagrep_v004_i007_p112_s14302.pdf
- He, S., & Schomaker, L. (2015). A Polar Stroke Descriptor for classification of historical documents. In: *13th International Conference on Document Analysis and Recognition, ICDAR 2015*, 23-26 Aug. 2015. Paper presented at 13th International Conference on Document Analysis and Recognition, ICDAR 2015, Nancy (France): IEEE Computer Society, 6-10. <https://doi.org/10.1109/ICDAR.2015.7333715>
- International Organization for Standardization. (2017a). *Photography — Archiving systems — Part 1: Best practices for digital image capture of cultural heritage material* (19263-1:2017).
- International Organization for Standardization. (2017b). *Photography — Archiving systems — Image quality analysis — Part 1: Reflective originals* (19264-1:2017).
- International Organization for Standardization. (2015). *Photography — Archiving Systems – Vocabulary* (19262:2015).
- Kenney, A. R. (2000). Digital Benchmarking for Conversion and Access. In *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (pp. 24–60). Mountain View, CA: Research Libraries Group.
- Liang, Y., Fairhurst, M. C., Guest, R. M., & Erbilek, M. (2016). Automatic Handwriting Feature Extraction, Analysis and Visualization in the Context of Digital Palaeography. *International Journal of Pattern Recognition and Artificial Intelligence*, 30 (4), 1653001-1- 1653001-26.
- Mallon, J. (1952). *Paléographie Romaine. (Scripturae. Monumenta et studia, III)*. Madrid: Consejo Superior de Investigaciones Científicas, Instituto Antonio de Nebrija de Filología.
- Meng, G., Pan, C., Xiang, S., Duan, J., & Zheng, N. (2012). Metric rectification of curved document images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4), 707–722. <https://doi.org/10.1109/TPAMI.2011.151>
- Meng, G., Xiang, S., Pan, C., & Zheng, N. (2016). Active Rectification of Curved Document Images Using Structured Beams. *International Journal of Computer Vision*, 122(1), 34–60. <https://doi.org/10.1007/s11263-016-0952-z>
- Morris, R. (2000). *Forensic handwriting identification: fundamental concepts and principles*. San Diego: Academic press.
- Nationaal Archief (2010). *Digitisation of photographic materials*. Guidelines. http://www.nationaalarchief.nl/sites/default/files/docs/guidelines_digitisation_photographic_materials.pdf
- Papaodysseus, C., Rousopoulos, P., Giannopoulos, F., Zannos, S., Arabadjis, D., Panagopoulos, M., . . . Tracy, S. (2014). Identifying the writer of ancient inscriptions and Byzantine codices. A novel approach. *Computer Vision and Image Understanding*, 121, 57–73. <https://doi.org/10.1016/j.cviu.2014.01.003>
- Plamondon, R., & Lorette, G. (1994). Automatic signature verification and writer identification: The state of the art - 1989-1993. *Pattern Recognition*, 22(2), 107–131. [https://doi.org/10.1016/0031-3203\(89\)90059-9](https://doi.org/10.1016/0031-3203(89)90059-9)
- Rajan, V. (2017). Quantifying scripts: Defining metrics of characters for quantitative and descriptive analysis. *Digital Scholarship in the Humanities*, 32(3), 602–631.
- Reilly, J., & Frey, F. (1996). *Recommendations for the Evaluation of Digital Images Produced from Photographic, Microphotographic, and Various Paper Formats*. Report to the Library of Congress National Digital Library Project. Contract # 96CLCSP7582. <http://lcweb2.loc.gov/ammem/lpireprt.pdf>
- Rieger, O. Y. (2000). Establishing a Quality Control Program. In *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (pp. 61–83). Mountain View, CA: Research Libraries Group.
- Robledano Arillo, J., Moreno Pelayo, V., & Pereira Uzal, J. M. (2016). Aproximación experimental al uso de métricas objetivas para la estimación de calidad cromática en la digitalización de patrimonio documental gráfico. *Revista Española de Documentación Científica*, 39(2), 1–17. <https://doi.org/10.3989/redc.2016.2.1249>
- Robledano Arillo, J. (2017). *La captura digital de patrimonio cultural. Experiencia de trabajo con el fondo del archivo Espacio P*. Granada: Ars Activus Ediciones.
- Robledano Arillo, J., & Navarro Bonilla, D. (2017). Aproximación sistemática a la creación de versiones digitales de negativos fotográficos históricos. *El Profesional de la Información*, 26(6), 1172–1183. <https://doi.org/10.3145/epi.2017.nov.16>
- Schomaker, L. (2008). Writer Identification and Verification. In N. K. Ratha & V. Govindaraju (Eds.), *Advances in Biometrics* (pp. 247–264). London: Springer. https://doi.org/10.1007/978-1-84628-921-7_13
- Schoonees, J.A., & Palmer, G.T. (2009). Camera shading calibration using a spatially modulated field, In: Bailey, D. (Ed.). 2009 24th International Conference Image and Vision Computing New Zealand. Paper presented at 24th International Conference Image and Vision Computing New Zealand (IVCNZ 2009), Wellington, New Zealand: IEEE, 191-196. <https://doi.org/10.1109/IVCNZ.2009.5378412>
- Shi, X., Diao, C., & Lu, D. (2009). Photo Vignetting and Camera Orientation Correction for High Precision Acquisition, In: *10th Pacific Rim Conference on Multimedia, Bangkok, Thailand*, December 15-18, 2009 Proceedings. Paper presented at Pacific-Rim Conference on Multimedia PCM 2009: Advances in Multimedia Information Processing - PCM 2009, Springer, 155-166. https://doi.org/10.1007/978-3-642-10467-1_13
- Sinn, D., & Soares, N. (2014). Historians' use of digital archival collections: The web, historical scholarship, and archival research. *Journal of the Association for Information Science and Technology*, 65(9), 1794–1809. <https://doi.org/10.1002/asi.23091>
- Sirat, C. (1981). *L'examen des_écritures: l'oeil et la machine. Essai de methodologie* (PhD. Thesis), Paris: Editions du Centre National de la Recherche Scientifique.
- Stokes, P. A. (2007). Palaeography and image-processing: Some solutions and problems. *Digital Medievalist*. <https://journal.digitalmedievalist.org/article/10.16995/dm.15/>

- Stokes, P. A. (2009). Computer-Aided Palaeography, Present and Future. In *Kodikologie und Palaeographie im digitalen Zeitalter-Codicology and Palaeography in the Digital Age* (309–338). Norderstedt: Herstellung und Verlag: Books on Demand GmbH.
- Wang, Z., Bovik, A.C., & Lu, L. (2002). Why is image quality assessment so difficult? In: *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Volume 4). Paper presented at Acoustics, Speech, and Signal Processing (ICASSP), Montreal: IEEE, IV-3313-IV-3316. <https://doi.org/10.1109/ICASSP.2002.5745362>
- Williams, D. (2002). Image quality metrics. *RLG Diginews*, 4 (4). <http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file1806.html>
- Williams, D. (2003). Debunking of Specsmanship: Progress on ISO/TC42 Standards for Digital Capture Imaging Performance, In: *PICS 2003: Image: Processing, Quality, Capture, Systems Conferece*. Paper presented at PICS 2003: Image: Processing, Quality, Capture, Systems Conferece, Rochester, New York: Society for Imaging Sciences and Technology, 77-81.
- Williams, D. (2010). *Imaging Science for Archivists*. <http://www.presentica.com/ppt-presentation/imaging-science-for-archivists-101-don-williams-image-science-associates-d-williamsimagescienceassociates-terms>
- Wueller, D, Bøgvad Kejser, U. (2016). Standardization of Image Quality Analysis – ISO 19264, In: *Archiving Conference, Archiving 2016 Final Program and Proceedings*. Paper presented at Archiving Conference 2016, Rochester, New York: Society for Imaging Science and Technology, 111-116. <https://doi.org/10.2352/issn.2168-3204.2016.1.0.111>
- Yu, W. (2004). Practical anti-vignetting methods for digital cameras. *IEEE Transactions on Consumer Electronics*, 50(4), 975–983. <https://doi.org/10.1109/TCE.2004.1362487>
- Zeng, F., Liu, S., & Xiao, K. (2017). Adaptive uneven illumination correction method of document images. *Journal of Computational Methods in Sciences and Engineering*, 17(3), 533–544. <https://doi.org/10.3233/JCM-170740>
- Zhang, L., Zhang, Y., & Tan, C. (2008). An improved physically-based method for geometric restoration of distorted document images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4), 728–734. <https://doi.org/10.1109/TPAMI.2007.70831>