CONTENTS

Page

INTRODUCTION 1
0.1 Context of the thesis 1
0.2 Problem statement 3
0.3 Objectives of the thesis 9
0.4 Outline of the thesis 12

CHAPTER 1 LITERATURE REVIEW 13
1.1 Intensity based document image enhancement 13
1.1.1 One-sided document image enhancement methods 13
1.1.2 Double-sided document image enhancement methods 14
1.1.3 Criticism 14
1.2 Multispectral Imaging based historical document image restoration 15
1.2.1 Electromagnetic radiation and optical proprieties of objects 15
1.2.2 Multispectral imaging 16
1.2.3 MS Images 17
1.2.4 Historical document image analysis 18

CHAPTER 2 METHODOLOGY AND CONTRIBUTIONS 23
2.1 Intensity-based binarization of historical document images 23
2.2 Multispectral restoration of historical document images 25
2.3 Reference data estimation for historical document image binarization 26

CHAPTER 3 ARTICLE I: A SPATIALLY ADAPTIVE STATISTICAL METHOD FOR HISTORICAL DOCUMENT IMAGE BINARIZATION 29
3.1 Introduction 29
3.2 Related work 31
3.3 Problem statement 34
3.4 Formulation 35
3.5 Methodology 36
3.5.1 Sauvola binarization algorithm 37
3.5.2 Spatially adaptive model 38
3.5.3 Computing the fields of μ_t, μ_b, and σ_b 38
3.5.4 Estimation of the σ_t field 42
3.5.4.1 Estimation of the global σ_t: S_t 42
3.5.4.2 Spatial adaptation of σ_t 43
3.5.5 Estimation of u_{BW} 44
3.6 Experimental results and discussion 45
3.6.1 Subjective evaluation 46
3.6.2 Objective evaluation against DIBCO'09 (Gatos *et al.*, 2009a) 50
3.6.2.1 Evaluation setup 51

3.6.2.2 Performance measures 52
3.6.2.3 Comparison with the state of the art 54
3.6.3 Computational cost and complexity of the method 55
3.7 Conclusions and future prospects 55

CHAPTER 4 ARTICLE II: DOCUMENT IMAGE RESTORATION USING MULTI-SPECTRAL IMAGING SYSTEM 59
4.1 Introduction 60
4.1.1 Difficulty in analyzing degraded document images 60
4.1.2 Objective of the paper 61
4.2 Related work 62
4.2.1 Hyperspectral remote sensing image enhancement 62
4.2.2 Multispectral imaging in the area of document analysis 63
4.3 Multispectral Image Acquisition 64
4.3.1 Characteristics of the MS degraded document image 66
4.4 Proposed restoration model 66
4.5 Parameter estimation and model optimization 68
4.5.1 Unsupervised IR band selection 68
4.5.2 Semi-local correction of slight degradations 71
4.5.3 Correction of strong degradations 72
4.5.3.1 Estimation of the binary mask 73
4.5.3.2 TV denoising and inpainting problem 75
4.6 Experimental result 78
4.6.1 Parameters setup 80
4.6.2 Subjective and objective evaluation 80
4.7 Conclusion 84

CHAPTER 5 ARTICLE III: REFERENCE DATA ESTIMATION 91
5.1 Introduction 91
5.2 Reference estimation methodology and its evaluation 94
5.2.1 General framework 94
5.2.2 Evaluation 97
5.3 Application: historical document image analysis 98
5.4 Conclusion 107

CHAPTER 6 GENERAL DISCUSSIONS 109
6.1 Adaptive soft thresholding for intensity-based HDI binarization 109
6.2 Variational method of multispectral HDI restoration 111
6.3 Reference data estimation in a multispectral representation space 112

GENERAL CONCLUSION 113

ANNEX I MS IMAGING SYSTEM, SET-UP AND ACQUISITION 119

ANNEX II AUTOMATIC FINDING OF THE THRESHOLD τ. 129

ANNEX III EXPERIMENTAL SET-UP FOR **IRR**, **UVR** AND **UVF** IMAGING TECHNIQUES..... 131

BIBLIOGRAPHY 132

LIST OF TABLES

Page

Table 3.1 Performance in terms of precision, recall, and *F-measure* of the proposed algorithm 52

Table 3.2 Comparison of the performance of the proposed algorithm and the others against DIBCO'09. For Sauvola's method, the performances reported in (Farrahi Moghaddam and Cheriet, 2010b) have been used 55

Table 3.3 Performance in terms of precision, recall, and *F-measure* of the proposed algorithm. For the Otsu (Otsu, 1979) and Sauvola (Sauvola and Pietikainen, 2000) methods, the performances reported in (Farrahi Moghaddam and Cheriet, 2010b) have been used 57

Table 4.1 F-measure and the gain ($\triangle_{Ai,i=1..9}$) values of the algorithms: A1 (Lu *et al.*, 2010), A2 (Farrahi Moghaddam and Cheriet, 2010b), A3 (Gatos *et al.*, 2004), A4 (Wolf *et al.*, 2002), A5 (Farrahi Moghaddam and Cheriet, 2012), A6 (Rivest-Hénault *et al.*, 2011), A7 (Toennies, 2005, P.202), A8 (Abutaleb, 1989b) and A9 (O'Gorman, 1994). The means (**Mean**) and standard deviations (**STD**) are calculated for more insight on the performance of each algorithm 82

Table 5.1 Divergence based comparison between different ground-truths 100

Table 5.2 F-measure based comparison between the reference contour and those of different ground-truths 102

Table 5.3 NRM based comparison between the reference contour and those of different ground-truths 103

Table 5.4 PSNR based comparison between the reference contour and those of different ground-truths 103

Table 5.5 Classification error rate (%) 107

LIST OF FIGURES

Page

Figure 0.1 Examples of degradation types 7

Figure 0.2 Summary of our research problem 9

Figure 0.3 The two first specific objectives of our research problem 10

Figure 0.4 Objective 3 12

Figure 1.1 Electromagnetic spectrum ranges(http://www.robertdalsanto.com/cosmology.php) 16

Figure 1.2 Spectral bands at different wavelengths from VIS to IR. Image from (Klein *et al.*, 2008) 21

Figure 1.3 The use of UV fluorescence to reveal the faded ink and highlight integrated strips 21

Figure 3.1 A sample set of some historical document images. (a) the original image (jum), which suffers from low-contrast degradation, (b) the original image (Gatos *et al.*, 2009a), which suffers from a degraded background, (c) the original image (Drira, 2007), which suffers from bleed-through degradation, (d) the original image (Gatos *et al.*, 2009a), which suffers from bleed-through and show-through degradation 31

Figure 3.2 An example of rough initialization. (a) original image, (b) initialization map (Sauvola binarization), the arrows showing the lost parts of the text, (c) ground truth map 32

Figure 3.3 (a) An example of degraded text. This is a part of one image of the data set used in this work. It is zoomed to display the local variation in intensity where some text pixels are brighter than the degraded background. (b) An example of text containing low-intensity pixels. For better visualization only a part of the input image (Figure 3.14) is shown 37

Figure 3.4 An example of how we produce the masked data (text and background). (a) original image, (b) initialization map (Sauvola binarization), (c) masked text data, the white area containing NaN values, this image is used to compute μ_t and σ_t. (d) masked background data, the black areas containing NaN values: this image is used to compute μ_b and σ_b 39

Figure 3.5 An example showing how to compute the local parameters. (a) computing $\mu_b(x_j)$ and $\sigma_b(x_j)$ for the background data. In practice, the black area contains NaN values which are not considered for the calculation of the parameters. The parameter of the pixel x_j in patch J is computed by using only the gray text pixels (not the black ones) belonging to the corresponding patch. (b) the same applies to estimating the text parameters: for example, to estimate the $\mu_t(x_j)$ in patch J, we take only the gray pixels in that patch and ignore the white pixels, which contain NaN values (painted in white to distinguish them from the gray text) 40

Figure 3.6 An example of how to inpaint the data. (a) μ_t only for masked text data, the white area containing NaN values, (b) inpainted mean text data values μ_t 41

Figure 3.7 Estimation of the global S_t. **(a)** histogram of the pixels of Fig. 3.8(c), (**b**) histogram of the pixels of Fig. 3.8(d). We can clearly see that $S_t^2 > \sigma_t^2$, and this is because of the low intensity text pixels 43

Figure 3.8 An example of how low intensity text pixels are recovered. (a) part of the original image shown in Figure 3.14, (b) the output of Sauvola's method (initialization map), (c) text stroke masked by (b), (d) text pixels masked by the output dilation of (b), (e) the result of the binarization method. The output of the whole image is presented in Figure 3.14 44

Figure 3.9 Example of how the grid works. Left, a sample image, the black squares representing the nodes of the grid, and the distance between two successive square equal to d_c. Right, the values in the black squares represent the calculated values of the grid. Each of them is computed by the function performed on all values belonging to the windows shown on the left. In this example, the input function is the mean function 46

Figure 3.10 An example of degraded shapes with variable intensities. (a) part of a degraded image (selected for better visualization), (b) the output of Otsu's method, (c) the output of Sauvola's method, (d) the output of the proposed method, which is continuous and smooth 47

Figure 3.11 A second example of degraded shapes with variable intensities. (a) the original image, (b) the output of Otsu's method, (c) the output of Sauvola's method, (d) the output of the proposed method which is continuous and smooth 47

Figure 3.12 A third example of degraded shapes with a strong interference of bleed-through. (a) the original image, (b) the output of Otsu's method, (c) the output of Sauvola's method, (d) the output of Drira's method (Drira, 2007), (e) the output of the proposed method 48

Figure 3.13 Binarization results of the images in Figure 3.1 49

Figure 3.14 Subjective comparison with the state of the art, (a) original image, (b) Sauvola thresholding result, (c) Su and Tan (Su *et al.*, 2010) result, (d) the proposed method result 50

Figure 3.15 Printed and handwritten ink degradation, **(a)** handwritten document, **(b)** printed document. The arrows indicate the missing ink or areas with little ink 54

Figure 3.16 Visual comparison of the binarization algorithms on the image shown in 3.2(a). **(a)** Sauvola's binarization, **(b)** Lu and Tan's binarization, **(c)** Rivest-hénault, Farrahi and Cheriet's binarization, **(d)** the proposed binarization 56

Figure 3.17 Time cost of the proposed algorithm. The continuous line indicates the processing time per pixel (the left axes), while the dashed line shows the image processing time 58

Figure 4.1 Some samples of degraded document images collected from the BAnQ 61

Figure 4.2 Document image enhancement using multispectral imaging system. (a) Color band. Due to the degradation of the ink, the details of the objects become invisible to the naked eye in the 400nm and 600nm band (b,c); but they can be distinguished in the infrared band at 1000nm (d). The images (a-d) are from (Klein *et al.*, 2008). The UV fluorescence (f) is used to reveal the hidden texts (e) 65

Figure 4.3 Two multispectral images: (a) 495 nm (Green), (b) 800 nm (IR), (c) 1100 nm (IR) 66

Figure 4.4 Binarization of the green and various IR channels. First row : (a) u_3 (green) band; (b) u_5 ($u_{\text{IR-1}}$) band; (c) u_7 ($u_{\text{IR-3}}$) band. Second row: the corresponding binary versions 70

Figure 4.5 Correction of slight degradations. (a) slight-degraded document image, (b) most distinctive and informative IR band of (a), (c) corrected document image using the model expressed by Eq. (4.10) 72

Figure 4.6 An example of 1D-data correction 73

Figure 4.7 Histogram based comparison between slight degradation correction models; (a) histogram distribution of the original data; (b) histogram of the corrected data by the model 4.10; (c) histogram of the corrected data by the model 4.9 73

Figure 4.8 λ mask estimation. (a) selected IR channel, $\hat{u}_{\text{IR-}\hat{p}}$; (b) binary mask λ..... 74

Figure 4.9 Mask estimation procedure from text/degradation overlap: (a) degraded image, (b), $\hat{u}_{\text{IR-}\hat{p}}$ channel; (c) binary mask λ, generated from (b); (d) inpainted image using λ; (e) overlapped text pixels are shown in yellow; (f) more similar overlapped pixels are shown in white; (g) mask λ', resulted from pixel-to-pixel addition of (c) and (f); (h) inpainted image using the mask (g) 74

Figure 4.10 Comparison of the end-member spectra of different classes to the reference mean spectrum of the class ink..... 76

Figure 4.11 Regions to be inpainted localisation. (a) color channel image; (b) estimated binary mask λ; (c) projection of λ on the image of Fig.(a). The regions to be inpainted are shown in black color..... 76

Figure 4.12 The overall restoration procedure; step 1: slight degradation correction; step 2: binary mask generation; step 3: projecting the binary mask onto the corrected color channel; step 4: inpainting processes . 78

Figure 4.13 Graphs showing the impact of the restoration step on the result of different binarization algorithms. The continuous red curves correspond to the results after restoration while the dotted blue curves correspond to the results of the binarization before restoration..... 83

Figure 4.14 Effect of the restoration step on the binarization process (the image shown in Fig. 4.17 (4a) is chosen as an example). (a-d), binarization outputs of the algorithms A1, A3, A6 and A9 before restoration. (e-h), binarization outputs of the same algorithms after restoration 84

Figure 4.15 Example of thin size degradations and the corresponding inpainting result; (a,c) degraded images; (b,d) inpainted images 85

Figure 4.16 Influence of the binarization transform on the inpainting process: (a) original image with main text and degradations (stamp and annotations); (b) example in which the mask is not well extracted; (c) inpainted image using the mask shown in (b); (d) inpainted image when the mask is well extracted (not shown here)..... 85

Figure 4.17 Result of the proposed algorithm. $[Xa]$, original image, $[Xb]$, enhanced image. $[X]$ is the number of the image 87

Figure 4.18 Result of the proposed algorithm. $[Xa]$, original image, $[Xb]$, enhanced image. $[X]$ is the number of the image 88

Figure 4.19 Result of the proposed algorithm. $[Xa]$, original image, $[Xb]$, enhanced image. $[X]$ is the number of the image 89

Figure 5.1 Visual contour-based comparison. (a) original image, (b) mapping between the contours of G_m and the the contours of the original image, (c) mapping between the contours of G_e and the the contours of the original image 104

Figure 5.2 Correlation between F-measure on contours and inter-classes divergence measures. (a) majority voting method; (b) proposed method 105

Figure 5.3 Domain definition. (a) visible band; (b) average ground-truth where the white color (1) means accurately-labeled text, black (0) means accurately-labeled background, and between 0 and 1 means partially-labeled text or background 105

Figure 5.4 Influence of β on the reference data estimation process (from left to right: $\beta = 0.1, 0.5, 1, 5, 10$) 106

LIST OF ABBREVIATIONS

ADU	Analogue to digital unit
AOTF	Acousto-optical tunable filters
BAnQ	Bibliothèque et Archives Nationales. du Québec
CCD	Charge-coupled device
DIBCO	Document image binarization contest
EM	Electromagnetic
ET	Exposer time
FoV	Field of view
HDI	Historical document image
HS	Hyperspectral
IR	Infrared
IRR	Infrared reflectography
LCTF	Liquid crystal tunable filters
LED	Light Emitting Diodes
ML	Maximum likelihood
MRF	Markov random fields
MS	Multispectral
MV	Majority voting
NRM	Negative rate metric
OCR	Optical Character recognition
QE	Quantum efficiency
RD	Reference data
RGB	Red Green Blue
ROI	Regions of interest
SNR	Signal to noise ratio

TV	Total variation
UV	Ultraviolet
UVF	UV fluorescence
UVR	UV reflectography
VCA	Vertices component analysis

INTRODUCTION

0.1 Context of the thesis

Since ancient times, man has felt the need to record his knowledge on physical media that can be exchanged among individuals. Once he discovered writing, many issues related to his daily life were tuned and codified in response to common public relations. Over time, the physical medium used for writing has greatly evolved, from a rigid base (wood, stone, clay tablet) to a flexible substrate (papyrus, parchment, paper). The year 105 AD is often cited as the year in which paper-making was invented in China. Since then, paper has played an important role in the development of civilizations and of cultures handed down from generation to generation, and it has become indispensable in our daily lives. More recently, our source of paper, the paper industry, is incorporating the technological and the societal aspects of life through the invention of tools adapted to the needs of consumers. As a consequence of our use of paper, millions of documents have been collected and are stored on library shelves. It is of paramount importance to create an environment in which these documents can be preserved, so that they are available for consultation for as long as possible. The sheer volume of existing documents and the ongoing production of new ones urgently require the development of new storage tools. Numerical discs and servers are the tools that allow the material to be saved in a computationally exploitable format. Several digitization companies have been established around the world to meet the requirements of archival centers for the preservation of their collections, to reduce the volume of documents they have stored, and to arrange them in readily accessible digital form.

Digitization services offer many advantages:

- They protect documents against future deterioration caused by inept handling or poor storage conditions. A single disc can store hundreds of digital documents and doesn't require a large space for doing so, which reduces the risk of damage. Moreover, multiple digital copies can be made of every document and stored in different places, so that even if files are lost or damaged, other intact copies are still available.
- They have the ability to create digital resources that can be exploited remotely by researchers, experts, and even the public at large via computer networks, mobile phones, etc. This facilitates knowledge transfer among individuals and groups everywhere. As well, because the information is in digital form, file exchange can be easily and rapidly managed in real time.

- They provide the option of document conversion to digital format at a time when the available space for storing physical documents is becoming scarce and increasingly expensive. Discs provide a simple and cheap means for storing thousands of documents, as well as generating savings on paper costs.
- They are moving progressively towards electronic document management. Many companies now offer digital document management services, including digital document archiving, metadata association, adding annotations, deletion, document retrieval, etc.

Digitization as a tool to serve humanity is a major technological challenge in today's information society, and gives rise to many historical, social, and political issues. The task of digitizing documents on a massive scale alone requires the development of new areas of tool research, in order to ensure quick and efficient access, as well as enable recognition and interpretation, of these documents. Document image analysis is one such research domain, and is aimed at the automatic processing of the overall physical structure of documents, recognition of the text and graphical components in document images, etc. (Nagy *et al.*, 1992; O'Gorman, 1997; Nagy, 2000). Document image analysis deals with two main categories of data: textual (i.e. characters, words, lines, etc.), and graphical (images, logos, stamps, diagrams, etc.). The data obtained from document image processing are the result of effective cooperation at all levels between the computer scientists who develop the algorithms that provide access to the image pixels, and the document specialists (historians, archivists, etc.) who define the need and analyze the results. Gaining access to the pixels calls upon all the methodologies and computational techniques available for analyzing, enhancing, compressing, and reconstructing the images that will be interpreted and analyzed by document experts.

Historical documents, which are the focus of this study, are of interest to researchers because of their cultural and scientific significance to society as a whole. Their studies are therefore to highlight the importance of the document in terms of understanding what happened in the past. The essential difference between a modern document and a historical document is that the latter is unique, i.e., it does not have multiple copies, and contains specific difficulties impeding access to its content, e.g. the presence of physical degradation caused by environmental conditions, dust, dirt, etc. Today, the digital archiving of historical documents is growing in the areas of heritage studies and preservation (Couasnon *et al.*, 2004; Antonacopoulos and Karatzas, 2004; Antonacopoulos and Downton, 2007; Manso and Carvalho, 2009; Joosten, 2008). This task requires that archived images be enhanced, restored, and stored in physical devices, in order to facilitate access to the valuable information they contain. As the ultimate objective of research in historical document analysis, either physical or logical, aims to provide

physical and logical analysis of historical documents, with the ultimate objective of their understanding (Shiel, 2010). This is considered an important aid to scholars interested in dating historical documents, reading old historical writing they contain and establishing their origins. The digital image processing of historical documents is one of the important low level tasks that provide a better quality of data that can be easily interpreted and understood by high level analysis tasks applied subsequently. Specialized conferences and workshops (ICDAR, ICFHR, DIAL, DAS, HIP, etc.), as well as contests and journals (Elsevier IJDAR, Elsevier PR, IEEE IP, IEEE PAMI), on the study of cultural heritage have helped to define various research directions related to historical document image (HDI) enhancement and restoration issues. As a result, hundreds of algorithms and frameworks have been developed, competitively and continuously, to solve the problems of text extraction, degradation attenuation, data reconstruction, and so on. A set of tools has also been developed for HDI, to provide easy access to their content and to ensure their preservation for future generations. Digital image processing in the field of historical document study encompasses a wide variety of topics, such as text extraction, character recognition, line segmentation, digital enhancement, restoration, etc. In fact, enhancement and restoration, which are addressed in this thesis, can significantly improve the quality of HDI and increase the efficiency of many of the high level analysis tasks. Moreover, they have paved the way for the introduction of new and interesting applications, such as document authentication, manuscript transliteration, etc., and have led to significant progress over time in our understanding and interpretation of cultural heritage.

0.2 Problem statement

Historical documents constitute a significant portion of cultural heritage, which plays a fundamental role in the economic and social development of nations. They are an essential feature of peoples and communities worldwide, and a record of their culture and civilization. Protecting them not only helps preserve the heritage itself, but also the underlying culture. Unfortunately, these documents are unique, and there is a serious risk of losing them, perhaps irretrievably. At the very least, they suffer from many forms of deterioration and degradation, owing to a combination of many factors: adverse environmental conditions, tears, dust, dirt, rusty staples, discoloration, etc. Such phenomena continue to harm these precious objects, and so there is an urgent need for a method of preserving them and providing broader access to them. Digital archiving is a standard way to meet this need. This process involves converting the document into a digital representation (format) and storing it in a computer to be accessed locally or remotely via the Internet. Digital document imaging is the most appropriate solution, because of its flexibility and ease of access. With it, huge amounts of data can be stored, duplicated,

managed, and manipulated, thanks to the availability of digital communications systems. The document images can be visually interpreted and provide the basis for data analysis. Two main concerns for any document imaging analysis method are quality and size. High quality digitization often requires expensive equipment, as well as more sophisticated verification procedures, which are also more time-consuming. Faced with the high cost of digitization, smaller firms favor quantity over quality. In fact, they prefer to scan with at low resolution to prepare documents for reading on screens, but not for analysis (Drira, 2007). The resulting images are imperfect, and contain physical degradations that make further treatment difficult. These images also show degradation caused by less advanced acquisition tools. Below, we discuss important sources of degradation and common problems encountered in low level document image processing, and, finally, we discuss the issue of the lack of standard HDI datasets.

Sources of degradations

Degradation in historical documents is acquired over time, and can make these documents difficult, if not impossible, to read. The sources of degradation are too numerous to count. Some of it is caused by a physical phenomenon (chemical, biological, or human), and some of it is caused by inadequate acquisition tools (Drira, 2007).

1. **Physical sources of degradation**

   - **Chemical**: This type of degradation of paper is largely caused by manufacturing processes. A sheet of paper is made from cellulose fibers of plant or animal origin. The glucosidal links in cellulose are stable in a neutral and slightly alkaline environment. On the one hand, however, these links are rapidly hydrolyzed in the presence of a strong acid or strong base, which results in a decrease, to a greater or lesser degree, in polymerization, and, consequently, in molecular weight. The hydrolysis of cellulose is also enhanced by the presence of oxide groups (aldehydes, carboxyls). On the other hand, lignin and rosin may cause the formation of peroxides, which are oxidizing agents that react very strongly and directly on the various chemical groups in the cellulose. Sulfur dioxide, nitrogen oxide and ozone caused by pollution also promote the hydrolysis and oxidation of cellulose.
   - **Biological**: Agents, such as mold, insects, and rodents, are primary contributors to the degradation of documents and manuscripts, as they feed on any organic substrates they find. Their proliferation is favored by a lack of ventilation and light, as

well as by high temperatures and high levels of humidity. These agents can destroy paper, binding, and ink.

- **Human-caused**: Improper storage and careless handling of the documents can cause tears, folds in the paper, etc., and some human interventions, such as chemical restoration, the addition of notes for archiving, etc., can hide portions of the writing. Frequent consultation by readers also accelerates the destruction of paper.

2. **External degradation**: Various defects may be generated during the document acquisition process, by, for example: the physics of the apparatus used; the coursing that results from finite spatial sampling; geometric deformation, such as non zero "skew" (rotation); defocusing, which is the result of a lens deviating from accurate focus; thermal sensor noise; pixel sensor sensitivity variation, etc.

Difficulties in historical document image processing

Historical document image enhancement and restoration are common low level problems in document image processing (and analysis). The purpose of document image enhancement and restoration (image clean-up) is mainly to remove unwanted artifacts, in order to make the image legible. This is especially important in form processing applications, OCR-based applications, word spotting, information retrieval, layout analysis, document classification, etc. These are some of the problems that arise related to low level degraded document image processing:

- **Uneven illumination:** This is caused by a non uniform trajectory of light between the camera and the document to be acquired, which results in non uniform contrast along the surface of the document image (see Fig. 0.1(f)), making it difficult to differentiate the foreground from the background.
- **Bleed-through**: This is a very difficult problem to address in HDI processing (see Fig. 0.1(d)). It occurs when ink seeps from the recto side of the document to the verso side (and vice versa) over a long period of time in storage. It becomes serious when the ink has seeped completely through the page, so that it appears similar to, or even darker than, the foreground.
- **Show-through:** This occurs with non opaque paper when the image on the back of the leaf is detected, resulting in a scan of the image of the other side (see Fig. 0.1(e)).
- **Ink fading:** This is a sign of poor storage conditions (high humidity, water leakage, high temperatures, etc.), which can accelerate the degradation of historical documents. The

ink can disappears completely, becoming invisible to the human eye (see Fig 0.1(a)). In some harsh conditions, the properties of the ink itself may change. Even a chemically neutral solvent (e.g. water) used in a chemical restoration procedure can affect the ink adversely.

- **Noise and blur:** These effects mostly affect recorded HDIs. One of the well known noises is the so-called dark current noise, which emanates from imperfections or impurities in the depleted bulk silicon, or at the silicon-silicon dioxide interface (Cortés, 2003). This noise is caused by thermally generated electrons that build up in the pixels of all CCD (charge-coupled device), whether they are exposed to light or not. The blur in document images can be caused by a number of factors: long exposure time, movement during the document image capture, defocusing, use of a wide-angle lens, insufficient gain, etc.
- **Others:** Many other defects contribute to the degradation of historical documents, making on-screen readability very poor. Among these defects are stains, holes caused by tears, shadowing caused by paper folds, etc. (see Fig. 0.1(b, c)).

With respect to tackling the problem of degradation in historical documents, we can state the following: i) HDI contain complex structures (e.g. text strokes, holes, etc.) located in different areas of the document, which makes them difficult to treat with static data structures and features; ii) The loss of some pixels during treatment may significantly affect subsequent recognition tasks; iii) Missing pixels cannot be recovered easily by simple regularization techniques, and so advanced a priori information about the document image content (e.g. baseline height, average stroke width, etc.) must be investigated. In some situations, consultation with a historical document expert on the type of text or meaning of words is required, which is time-consuming; iv) Where there is degradation, the statistical distributions of the intensity of the various document constituents (background, text, graphics) often overlap, which makes separating these constituents very difficult; v) Document image degradation is random and non linear, and can rarely be modeled or predicted. So, in spite of the large number of document image enhancement and restoration methods available today, dealing with the degradation phenomena mentioned above remains a challenge, and they often fail to effectively deal with multiple types of degradation. As a result, methods designed to handle some specific situations may not be appropriate for others. Efforts are therefore focused on developing individual solutions to specific problems. Fig. 0.2 summarizes the overall correspondence between high-level issues (e.g. restoring and preserving cultural heritage) and problems related to the low-level processing of digital data.

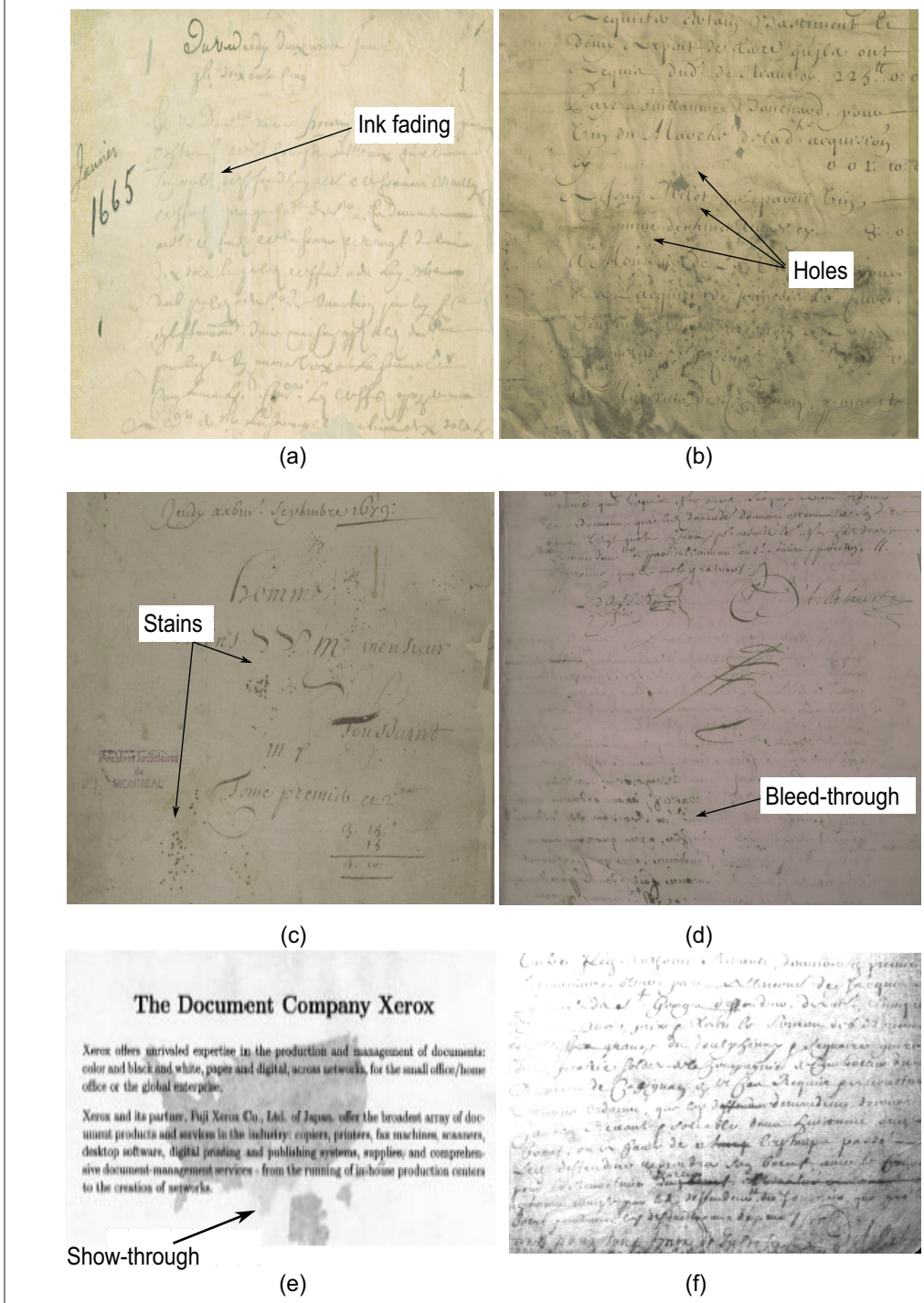


Figure 0.1 Examples of degradation types

Problem of the lack of benchmarks for HDI processing

One of the major problems in the document image processing field, and more specifically for historical document image processing, is the lack of real benchmarks or standard datasets. A standard dataset is mainly used to compare different algorithms, tracking their progress over time towards emulating human performance, and even beyond. A standard dataset should be accompanied a reference data (RD) generated by a human (i.e. manual RD), which is believed to accurately reflect the true target information to be processed. Given an evaluation metric, the performance of any algorithm is usually validated by comparing its output with the RD. Standard datasets are very helpful for developing new algorithms and provide a way of evaluating their performance. In spite of efforts made by researchers to produce datasets (Gatos *et al.*, 2009b; Pratikakis *et al.*, 2010, 2011), there are still two major problems:

1. These datasets cover only a small set of documents and show only simple types of degradation. Therefore, even though the various binarization methods have achieved very encouraging performance scores on these datasets, there is no guarantee that these methods can be generalized. In addition, a dataset of finite size of samples can easily contaminate the design and evaluation of a proposed system (Raudys and Jain, 1991). and have a negative impact on the search for the most effective features for designing optimal systems, owing to the introduction of bias and variance into system performance (Way *et al.*, 2010).

2. The process of manual generation of RD remains a subjective notion, which means that it can be viewed in a dramatically different way by different humans, at different times, and under different RD generation conditions.

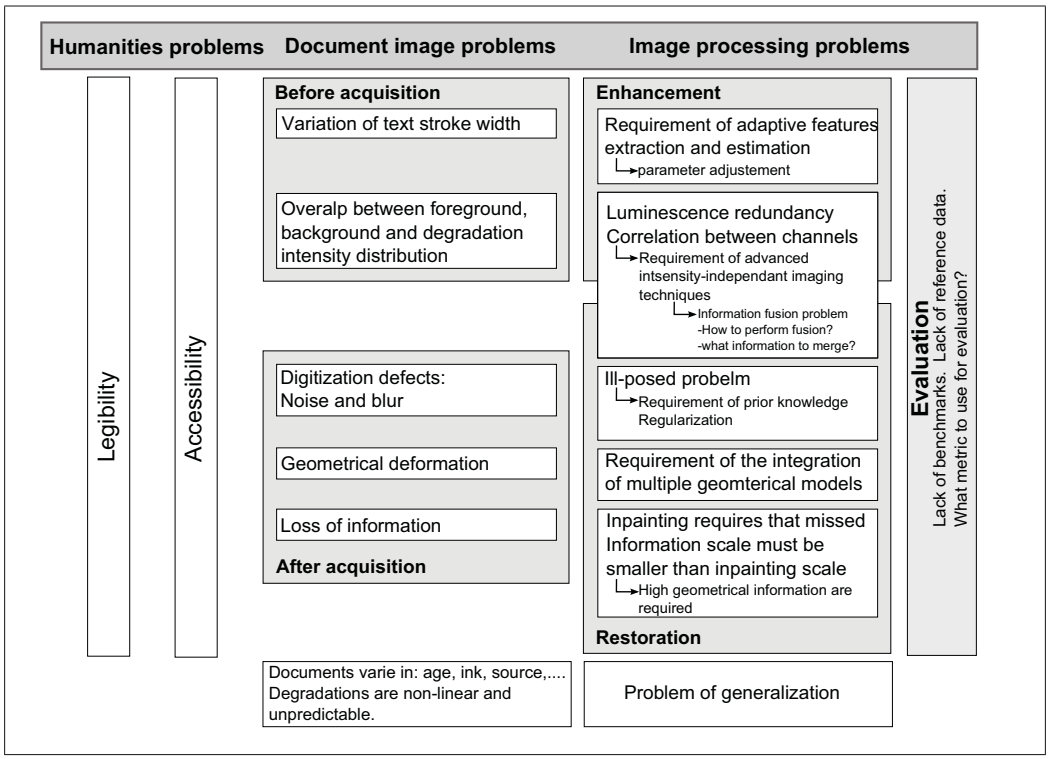


Figure 0.2 Summary of our research problem

0.3 Objectives of the thesis

The overall objective of this thesis is to define an advanced image processing framework for the enhancement and restoration of HDI, which constitute a very important part of cultural and scientific heritage. This framework should make HDI legible and more accessible. In this thesis, improving legibility means reproducing HDI with better visual quality by restoring them to reduce or eliminate the effects of degradation that are generated over time, or by the digitization process. It is usual for libraries and archival centers to try to provide the end user of their collections with better visual quality. Improving accessibility means making HDI ready for high-level analysis (information retrieval, character recognition, text understanding, etc.) by providing them with more accurate binary (bi-level) information easy to manipulate at a lower computational cost. The positioning of our objectives within the HDI analysis framework is illustrated in Fig. 0.3 and situates the ultimate goals within the analysis HDI framework).

Specifically, access to binary images by high level analysis methods depends primarily on the accuracy of the binarization outputs. Where the degradation is distributed over the entire document image (i.e. shadows, uneven illumination, etc.), local binarization methods are required

to adapt to the degradation and stroke width changes. However, these methods are ineffective when the extent of the degradation is less than the average line height in the document (i.e. the average distance between baselines). This means that degradation and changes in intensity which are local and confined to small parts of text strokes cannot be captured by traditional local and adaptive binarization methods. In this case, the appearance of holes and discontinuities within the binary text strokes obtained leads to a reduction in the performance of OCR engines. We can hypothesize that a more flexible, local, soft-thresholding model could carefully handle the non uniformity of document image intensity, and preserve weak connections between text strokes to provide more accurate binarization results. This hypothesis led us to formulate the first specific objective of this thesis:

Specific Objective 1

- To define a new classification model that is robust enough to handle strong intensity variations, and capable of preserving weak connections between text strokes, in order to achieve more accurate binarization results (Fig. 0.3).

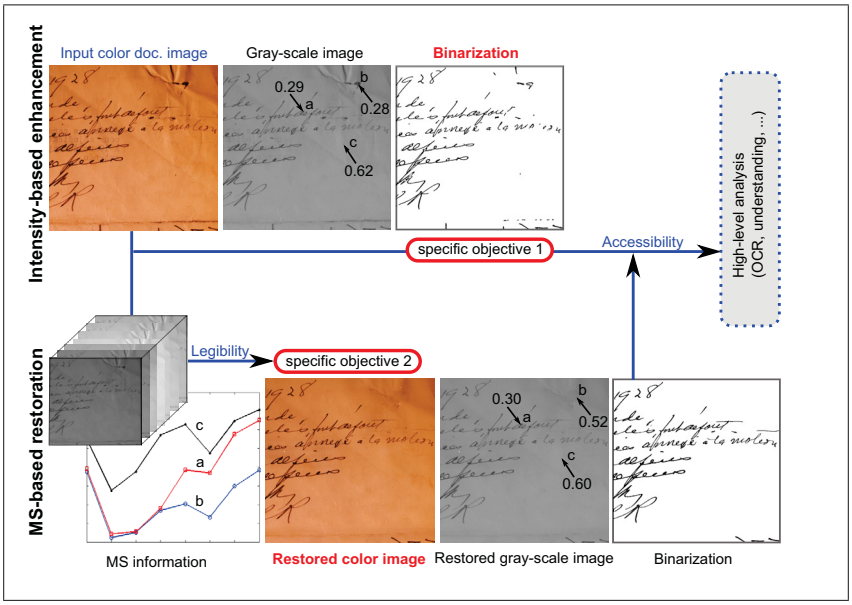


Figure 0.3 The two first specific objectives of our research problem

The legibility of color HDI can be improved by means of image restoration techniques. Unfortunately, since there is a strong correlation between the various channels when a typical HDI is represented in an RGB color space, the restoration process fails, especially with severely

degraded HDI, where many different objects appear to be similar. The gray-scale image in Figure 0.3 shows a degradation (pixel **b**) of almost the same intensity as the text (pixel **a**)(i.e. 0.29 and 0.28), and this obviously affects the binarization result. To improve the restoration performance, the HDI should be acquired by one of the new imaging technologies, which are more efficient in terms of providing other, less strongly correlated channels in addition to the RGB channels. In this thesis, an MS imaging system is used to produce spectral images in the invisible light range (IR and UV), which will help us build robust and reliable restoration models. In Figure 0.3, pixels **a** and **b**, which appear similar in an intensity-based representation, have different spectral signatures in the MS representation space. This observation led us to formulate the second specific objective of this thesis:

Specific Objective 2

- To propose an automatic approach for the visual restoration of color HDI. This approach will provide the end user (scholar or librarian) with an acceptable view of an HDI, in which only the original text is retained without any significant changes to the texture of the background. The restored color and gray-scale images in Figure 0.3 show that only the intensity of the degradation (pixel **b**) is changed, while text and background still have almost the same intensity.

Any new enhancement or restoration algorithm must be evaluated and compared to the state-of-the-art based on human RD. Otherwise; it is not possible to track its progress over time towards emulating human performance. A common way to generate RD is to involve an expert who is capable of manually labeling HDI pixels, both foreground and background. Unfortunately, even an expert RD is subjective in nature, and liable to include mislabeling and judgment errors, owing to the influence of degradation and noise in the data. As a result, the idea of combining multiple RD generated by several experts seems to be a promising approach to estimating new, more accurate RD. These data are then validated and refined based on an MS representation space in a way that cannot be achieved in an intensity-based representation space (gray-scale or color). This realization led us to formulate our third specific objective of this thesis:

Specific Objective 3

- To devise a robust and reliable statistical model for the estimation of new RD that is more accurate than the RD of any expert. This model will be able to quantify both mislabeling errors and judgment errors, and also estimate a new, more accurate RD from that of multiple experts (see Figure 0.4).

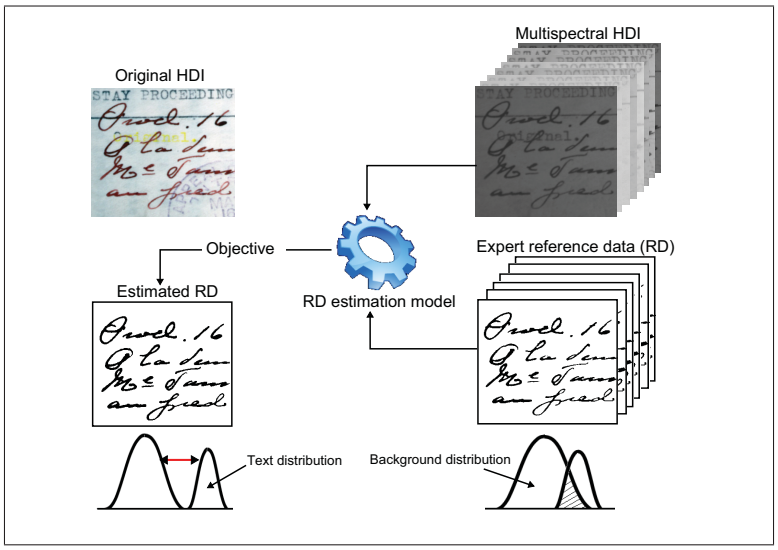


Figure 0.4 Objective 3

These three specific objectives offer a framework comprising a complete HDI enhancement and restoration chain, from intensity-based to MS image processing. They will serve as cornerstones of a comprehensive analysis system designed to meet the needs of end users, whether historians, librarians, or the general public. Achieving these objectives will not be as easy as one might imagine, and so specific considerations and complementary efforts are needed to achieve each of them.

0.4 Outline of the thesis

This thesis is organized as follows: The introduction explains the context of our work and presents the problem statement and our intended objectives. Chapter 1 describes the relevant state-of-the-art methods related to the proposed methodology. Chapter 2 describes the general methodology designed to achieve the objectives of the thesis. Chapter 3 describes our new method of intensity-based document image binarization based on local statistics and maximum likelihood classification. Chapter 4 presents our new method of historical document image enhancement in a multispectral representation space. Chapter 5 describes our new reference data estimation method for document image binarization. A general discussion is given in chapter 6. Finally, in our general conclusion, we summarize the work accomplished during the thesis and prospects for the future.

CHAPTER 1

LITERATURE REVIEW

1.1 Intensity based document image enhancement

Recently, document image enhancement methods based on intensity information have attracted a great deal of attention and interest, as they play an important role in other automatic analysis tasks (OCR, document recognition, etc.) and enhance the readability of documents for the benefit of experts (historians and librarians). Hundreds of methods have been proposed over the years, which can be classified in two main categories: those that require access to both the recto and verso sides of the document simultaneously (double-sided enhancement methods); and those that process each side independently (one-sided enhancement methods).

1.1.1 One-sided document image enhancement methods

One-sided enhancement methods attempt to eliminate interfering patterns using thresholding or classification techniques. Thresholding techniques are aimed at finding an optimal threshold (gray-level) which separates the document image pixels into two classes, foreground and background, and classification-based techniques are aimed at classifying the document image pixels into two or more classes, foreground and background, and potentially a fuzzy class.

With thresholding-based enhancement techniques, a pixel is considered to be foreground if its value is above the threshold, and background otherwise. There are two categories of thresholding techniques: global and adaptive. Global thresholding techniques (Otsu, 1979; Kapur *et al.*, 1985; Abutaleb, 1989a) are designed to find an single optimal threshold for all the document image pixels. Unfortunately, in the presence of a high level of degradation, such as severe bleed-through, simple thresholding techniques are inadequate for the task of image enhancement, because the intensity of interfering patterns or degraded background can be very similar to that of the foreground. Global thresholding cannot eliminate the degradation in such cases either, or, if it succeeds in doing so, it may eliminate parts of the main text as well. As a result, adaptive thresholding techniques (Niblack, 1986; Bernsen, 1986; Mardia and Hainsworth, 1988; Sauvola and Pietikainen, 2000; Gatos *et al.*, 2008; Moghaddam and Cheriet, 2012) have been developed and adopted. These techniques calculate a single threshold for each document image pixel in a specific window size. Other authors have proposed performing the classification task in the gray-level or color space using k-means clustering (Leydier *et al.*, 2004;

Drira *et al.*, 2006), MRF (Markov random fields) (Wolf, 2009; Hedjam *et al.*, 2010) or entropy information (LIoud, 1985; Kittler and Illingworth, 1985; Boussellaa *et al.*, 2007). Structure-based (Wang *et al.*, 2003; Lu *et al.*, 2010), multiscale-based (Leedham *et al.*, 2002; Nishida and Suzuki, 2002), inpainting-based (Zhang *et al.*, 2009), and variational-based segmentation techniques have also been proposed (Rivest-Hénault *et al.*, 2012; Bar-Yosef *et al.*, 2009). Although these latter methods have been more successful than local and global thresholding methods, they have more complexity, owing to the learning phases their parameters must undergo. Also, they only exploit a limited amount of information from one side of the document. Double-sided enhancement methods, which apply to both the recto and verso sides, are adopted to overcome this disadvantage.

1.1.2 Double-sided document image enhancement methods

Double-sided enhancement methods are based mainly on the registration of both the recto and verso sides of the document. They perform the classification task in two stages. The first step is to register and align the two sides of the document. The second step is to eliminate interfering patterns originating on the verso side of each registered image using matching (Wang and Tan, 2001; Dubois and Pathak, 2001; Dubois and Dano, 2005), non linear inverse diffusion (Cheriet and Farrahi Moghaddam, 2008a,b,c; Farrahi Moghaddam and Cheriet, 2010a), classification/segmentation (Knox, 1997; Tan *et al.*, 2002; Moghaddam *et al.*, 2009; Rowley-Brooke and Kokaram, 2012), or source-separation techniques (Tonazzini *et al.*, 2004; Ophir and Malah, 2007; Su and Mohammad-Djafari, 2007; Tonazzini *et al.*, 2010; Martinelli *et al.*, 2012).

Double-sided enhancement methods are an improvement over the one-sided enhancement methods, in terms of thresholding results, owing to the advantage of using other available data, generated by the simultaneous treatment of both sides of the document. Nevertheless, additional computational time is required for the registration/alignment step, which may not be a simple task because of possible geometrical deformations (rotation, skew, wrapping, etc.) and the potentially different resolution of the recto and verso images.

1.1.3 Criticism

In spite of the large number of image enhancement/restoration algorithms in the literature, there are no generic algorithms that can be used to handle a number of types of document image degradation. Most algorithms are trained on a finite set of document images, and then tested on another set of document images in the same category, i.e. having similar characteris-

tics (Cheriet *et al.*, 2012). It seems that the time has not yet come to design generic frameworks for the document image enhancement problem that can at least handle a large set of degraded documents that belong to a single culture or to a specific time period. There are two main reasons for this. One reason is the nature of non linear degradations, which is that they are not predictable. This makes it difficult to develop robust and reliable enhancement/restoration models. Researchers are particularly interested in designing specific models that incorporate information gathered from the available data, in order to regularize the results of their algorithms. Unfortunately, they fail to consider (intentionally or unintentionally) how degradations occur. The second reason is that intensity-based information is not suitable for designing good discriminant features, especially in the case of severely degraded document images. This is mainly due to the physical and mechanical limitations of the conventional tools used for document image acquisition. The cameras and scanners typically used to capture these images provide a subset of information that is made available to us by combining the responses of the visible radiation into three spectral images or less (color or gray-scale). Although the RGB color space is the most common choice for computer graphics, it is not very efficient in dealing with real-world images, because the RGB channels contain redundant luminance information. This information is highly correlated, as all of it includes a representation of brightness (de Campos, 2006). So, if the acquisition is based on color information only, or on gray-level information only, the various document image constituents may appear similar to the human eye, which makes the process of separation difficult, or even impossible. Multispectral (MS) imaging systems seem to be a good alternative, as they offer detailed quantitative measurement of the spectral responses of the document image constituents. These systems are the subject of the next chapter.

1.2 Multispectral Imaging based historical document image restoration

1.2.1 Electromagnetic radiation and optical proprieties of objects

Light or electromagnetic (EM) radiation can be viewed as a continuous series of particles moving in the absence of a medium and at the speed of light ($c = 3 \times 10^8 m/s$). These particles, which carry energy, are called *photons*. EM radiation can be also seen as wave with the proprieties of frequency f and wavelength λ. The velocity of the wave is computed as $v = f \times \lambda$. The wave is considered as a spectrum of light. The energy of a photon is equal to $E = h \times f$, where $h = 6.63 \times 10^{-34} Joule/s$ is Planck's constant. Consequently, the intensity of light is related to the number of photons involved (i.e. the more photons there are, the brighter the light, and vice versa), and the energy of each photon depends only on its frequency:

$$E = h \times f = h \times \frac{c}{\lambda} \tag{1.1}$$

The waves in the entire spectrum of EM radiation vary from short (Gamma rays, X-rays, ultraviolet (UV) rays, and cosmic rays) to long (infrared (IR) rays, microwaves, and radio waves). Visible radiation, which is of medium wavelength, is the light that humans can see. Fig. 1.1 shows the various regions of the EM spectrum. When a light wave with a unique wavelength comes into contact with an object, it can either be absorbed or transmitted by the object, or reflected directly or diffused by that object. In the case of absorption, the energy is converted into heat. Reflection occurs when the incident light is partially or totally reflected. The amounts of light reflected over a range of wavelengths constitute what is known as the ***reflectance spectrum***. Transmittance occurs when light penetrates layers of objects. In this case, the light can be directly transmitted in one direction or diffused in different directions.

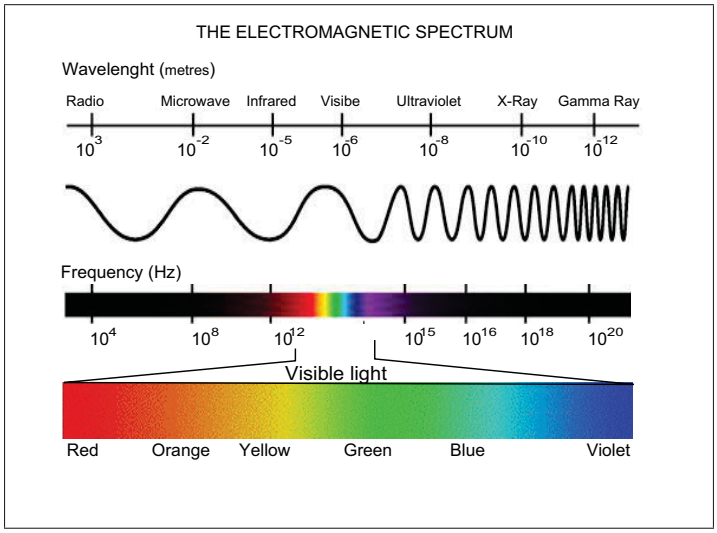


Figure 1.1 Electromagnetic spectrum ranges(http://www.robertdalsanto.com/cosmology.php)

1.2.2 Multispectral imaging

Multispectral (MS) imaging is used mostly to record spectral images in the visible light range and in the invisible light range (i.e. UV and IR). Thanks to the use of UV and IR sensors, MS imaging can extract information that the human eye cannot capture with its receptors for red, green and blue. Light that is visible (to the human eye) has wavelengths in the range of about 380 nm to 740 nm. A spectral image is reproduced as a grey-scale image or an RGB color

image. Visible light is situated between UV light, which has short wavelengths – in the 10 nm to 400 nm range, and near-IR light, which has long wavelengths – in the 700 nm to 1 mm range. IR spectral images can be combined into a grey-scale image, and three of them can be used to create pseudo color RGB images.

The principle underlying MS imaging systems is the concept of the spectral signature. The main idea is that all materials emit, transmit, or absorb EM radiation based on the inherent physical structure and chemical composition of the material, and the wavelength of the radiation. Every material transmits, absorbs, or emits an amount of EM radiation commensurate with the wavelength and intensity of the radiation impinging on the material. The ratio of reflected to emitted radiation from the surface of an object varies with the frequency of the wavelength and the angle of incidence of the radiation. The combination of emitted, reflected, and absorbed EM radiation across a range of wavelengths produces what we call a spectral signature, which is unique to that material (see Figure I-7 in Appendix I: MS imaging system, set-up and acquisition). It is therefore possible to differentiate between objects based on differences in their spectral signatures. There are a number of applications for MS imaging. The technique was first developed for remote sensing and geo-observational applications (Landgrebe, 1980; Kerekes and Landgrebe, 1991; Chang, 2003; Kalacska and Sanchez-Azofeifa, 2008). However, MS imaging has recently been applied in new applications, in fields as diverse as biology (Backer *et al.*, 2007) and the military (Ifarraguerri and Chang, 1999). IR wavelengths penetrate the layers of a material; potentially deeply enough to reveal information that is invisible to the human eye. IR is used, for example, as a night vision aid when visibility is poor, and in thermography to remotely determine the temperature of objects. UV light reflected from the surface of a scene can also lead to the detection of trace evidence in a forensic investigation (Tahtouh *et al.*, 2007; Chalmers *et al.*, 2012). In another application, artifacts integrated into paper currency, such as scratches or stripes, can be revealed by the florescent light that they emit when exposed to UV radiation. Artwork has also benefited from MS imaging (Melessanaki *et al.*, 2001; Casini, 2002; Ribes *et al.*, 2008; Pelagotti *et al.*, 2008a), as has historical document transcription (Walvoord and Easton, 2008; Easton *et al.*, 2003; et al., 2003; Klein *et al.*, 2006).

1.2.3 MS Images

An MS image is a stack of spectral images (or bands) of the same scene, each image having been captured at a specific wavelength band range. The main difference between Hyperspectral (HS) and MS is in the number of bands they produce. MS imaging produces from 4 to 10 bands

at discrete EM wavelengths over the range of light, from UV light, to visible light, to IR light, while HS imaging produces more than 10 and up to 200 bands over a continuous range of wavelengths in several narrow light spectrum bands. In good conditions of illumination and imaging system setup, MS images are recorded using various strategies:

- A monochrome camera and various color filters are manually placed on the top of the camera's sensor. This strategy is simple and inexpensive, but rarely used in real situations. A monochromator is a device used to select narrow band of wavelengths of light or other radiation chosen from a wider range of available wavelengths (Bei *et al.*, 2004). Light Emitting Diodes (LED) are another source of a narrow band of wavelengths in all light regions, from UV to IR.
- Tunable filters are used to gather a number of monochromatic images, one for each chosen narrow rang of wavelength band. The choice of tunable filter involves a compromise between narrow-band filters, which provide specific information in a spectral region, and broad-band filters, which transmit sufficient light intensity for the application (Pelagotti *et al.*, 2008b). The filters are installed in a wheel and controlled automatically, with the aim of improving the temporal resolution of the measurement relative to the manual strategy. Two types of filter are used more frequently than others: liquid crystal tunable filters (LCTF) (Brettel *et al.*, 1999), and acousto-optical tunable filters (AOTF) (Bei *et al.*, 2004; Calpe-Maravilla *et al.*, 2004). LCTF use liquid crystal elements to transmit a specific or selected wavelength and exclude others. High spectral resolution can be obtained by combining several electronically tunable stages in series. An AOTF is a diffraction-based optical band-pass filter that can be rapidly tuned to enable various wavelengths of light to pass through by varying the frequency of an acoustic wave propagating through an anisotropic crystal medium (Stratis *et al.*, 2001). The choice of filter type is application-dependent.
- A triangular prism is placed in front of the camera (Du *et al.*, 2009). Based on the refractive index of each wavelength, the prism disperses each incoming light ray into a spectrum of its constituent colors. This technology consists of low-cost off-the-shelf components and is easy to setup, and provides true MS measurements of an imaged scene location at a given time.

1.2.4 Historical document image analysis

MS and HS imaging are very important tools which enable conservators and art historians to obtain valuable information on artworks and ancient documents without causing physical

damage to the materials. This technique is widely accepted as a non invasive method of investigation. Thanks to its simultaneous use of UV, IR, and visible light, it is possible to reveal newly painted areas or overwritten text, to distinguish and recognize the chemicals composing the ink, to enhance the visibility of latent patterns in a palimpsest, to detect signs of degradation in historical documents, etc. It also improves the readability of old documents and provides information on alterations and degradation phenomena that may have been added over the years by conservators, or others with less well intentioned motives. It can also help to extract information from cultural heritage patterns which cannot be extracted using conventional color photography (Klein *et al.*, 2008). Several non destructive analysis methods using MS or HS imaging can be found in the literature:

1. **IR reflectography**(IRR): This technique records portions of absorbed and reflected IR light, which passes through the document layers (i.e. varnish and overwriting) to interact with the underwritten portions of the document. It can provide a document historian with very important information about the types of ink used and the document constituents, all of which help him assess the condition of the document under study (see Figure 1.2). Many IRR applications and projects have been undertaken, and many works on IRR have been published in the literature. Examples are: digital character recognition for digital transcription of the Archimedes Palimpsest (Walvoord and Easton, 2008); highlighting of the various classes of the Archimedes text (Easton *et al.*, 2003); extraction of the faint and highly degraded underwritten text in Archimedes' manuscripts (Salerno *et al.*, 2007); improvement of the visual quality of text-based documents corrupted with unwanted artifacts (Kim *et al.*, 2010); systematic investigation of old writing inks (Senvaitiene *et al.*, 2005); paper characterization and conservation (Manso and Carvalho, 2009; Maino, 2007); interpretation of the photometric properties of inks and classification of the various types of ink used in Byzantine-era manuscripts (Kokla *et al.*, 2006); identification of the elemental composition of the pigments and characterization of the decomposition phenomena that affect writing and print substrates (Faubel *et al.*, 2007); study of the aging of ink, of the biological and physical damage to documents, and of the legibility enhancement of deteriorated documents, etc. (Klein *et al.*, 2008, 2006; Goltz *et al.*, 2010); and study of the aging effects of iron-gall ink (Havermans *et al.*, 2003b). The experimental setup for IRR is to illuminate the document uniformly using a source of IR radiation (a tungsten lamp, for example), and then install an IR pass filter in the front of a digital camera. This results in a monochromatic image of the IR light reflected by the document (see Figure III-1 in Appendix III: Experimental set-up for **IRR**, **UVR** and **UVF** imaging techniques).

2. **UV reflectography** (UVR): This technique records portions of absorbed and reflected UV light. UV light is an effective tool that can be used to detect newly touched up areas and later restorations that are not visible to the human eye. The experimental setup for UVR involves illuminating the document under study using UV lamps (usually referred to as *black light*) and installing a UV pass filter in the front of the acquisition camera to exclude the reflected visible light and allow only reflected UV light to pass through (see Figure III-1 in Appendix III: Experimental set-up for **IRR**, **UVR** and **UVF** imaging techniques). The result is a gray-scale (monochromatic) image of the UV light reflected from that document. UVR is also a very useful tool for investigating ancient manuscripts.

3. **UV fluorescence** (UVF): Some materials absorb invisible UV radiation and use the energy to give out extra visible light. The combination of the extra light and the normally reflected visible light makes some particles appear brighter than others. This phenomenon, which we call *fluorescence* (see Figure 1.3), can occur with particles that become excited when exposed to UV radiation. The experimental setup for UVF is similar to that for UVR, except that the UV pass filter is replaced with a UV cut filter, in order to exclude reflected UV radiation and allow only visible light to pass through. This technique is useful for many historical document analysis applications. It is often used to reveal degraded text that cannot be seen with visible light (andJ. Bartl and Jacko, 2003; Comelli *et al.*, 2008; Casini, 2002; Lettner *et al.*, 2008; Salerno *et al.*, 2007) (see Figure III-1 in Appendix III: Experimental set-up for **IRR**, **UVR** and **UVF** imaging techniques), as well as in bank security and counterfeit currency verification (Hardin, 1999; Chen *et al.*), for example.

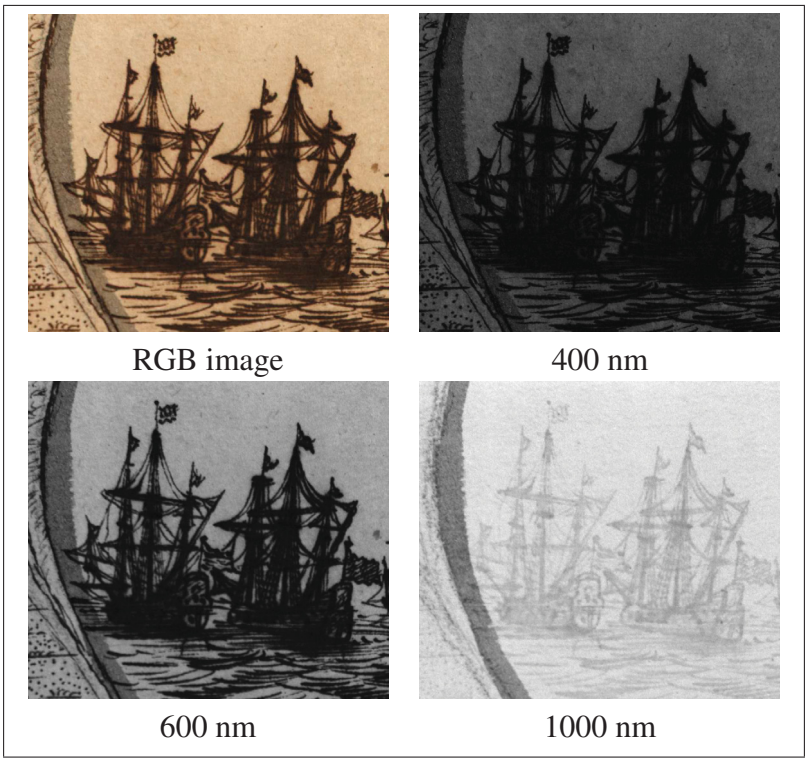


Figure 1.2 Spectral bands at different wavelengths from VIS to IR. Image from (Klein *et al.*, 2008)

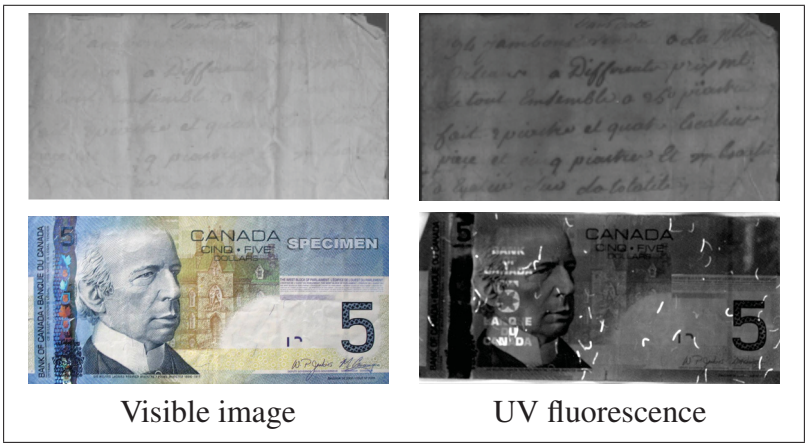


Figure 1.3 The use of UV fluorescence to reveal the faded ink and highlight integrated strips

CHAPTER 2

METHODOLOGY AND CONTRIBUTIONS

The general methodology of this thesis is directly linked to the objectives defined in chapter 1, and consists of three principal themes: 1) intensity-based adaptive binarization of historical document images (HDI); 2) MS-based restoration of HDI; and 3) reference data (RD) estimation for HDI binarization. These themes differ, both theoretically and technically, but together they form the production chain that we have designed to meet the ultimate goal of document legibility and accessibility, once the necessary datasets have been acquired. In this thesis, we study the role of each theme of the methodology, which traces the path of document image analysis from simple intensity information to MS information, and on to more complex notions, such as data labeling and expert judgment. Through this process, we show the effectiveness of the local estimation of classifier parameters in the design of a robust method of soft thresholding, which can exceed the parameter estimation limits of hard thresholding. We then study the problem of HDI restoration in an MS representation space. More information can be analyzed simultaneously in this space, as it affords greater power of classification. In addition, because information processing is based on the photometric properties of objects rather than on their intensities, the relationship between objects can be better understood. MS information is also beneficial for various image processing problems, such as the estimation of more accurate RD. The methodology proposed here for RD estimation requires that the RD of a number of experts be available, along with MS information about each pixel, and that all the information be used concurrently to estimate the class label for the information in a Bayesian framework. These methods are briefly described here, and then in detail in a chapter devoted to each method.

2.1 Intensity-based binarization of historical document images

The binarization of HDI characterized by spatially variant degradation is considered a major problem, because it has a direct effect on optical character recognition (OCR) engines, and consequently on the automatic document understanding. We propose a method to achieve spatially adaptive binarization which is robust to the presence of local degradation, and capable of recovering weak connections between text strokes, a task that is directly linked to meeting the first objective of this thesis. During our research, we studied the effectiveness of adaptive binarization and its ability to separate text from a document image suffering from spatially variant degradation. We found that more advanced adaptive binarization methods are required not only to address the degradation issue, but also to preserve weak connections between text

strokes, so that OCR performance is not compromised. As a result, we changed the approach of our research problem to one of classification, which has allowed us to address intensity variation and changes in text stroke width.

As stated in the first objective, we focus on methods that have the potential to treat very local variations in stroke intensity. Typically, the intensity of some text strokes is lower than that of the degradation, and they are so narrow that they are nearly invisible and very difficult to recover using a simple thresholding technique. Several binarization methods have been proposed for this problem, in an attempt to adapt the threshold to local intensity variations. They do this by using document image features, such as contour sharpness, contour direction, contour gradient magnitude, local minimum-maximum variance, intensity distribution, entropy, frequency information, etc., to separate text from degradation. Unfortunately, these methods are not very effective and the results with severely degraded document images are modest at best, and they generate noise and artifacts in the outputs. Statistical methods, based on MRF (Markov random fields), have also been tried. Although they appear to be successful, thanks to the introduction of the spatial relationship between a pixel and its neighbors, they are also inclined to fail in challenging situations, the loss of small looped characters being one of the more disastrous consequences.

In this thesis, we propose a simple and efficient soft thresholding method for binarization, designed to remove degradation while preserving the connections between weak strokes that are narrow and of low intensity, and to avoid the complexity of previous methods. Our method is based on the following assumption: Since there is at least one subset of the most probable text pixels (also called the region of interest – ROI), it becomes possible to detect the neighboring low probability text pixels by means of local classification. Low probability text pixels are all those with low intensity or which belong to thin strokes. First, a binary ROI map is roughly estimated by a simple binarization technique, and then this map is used as a mask to locally estimate the mean and variance of each class. These parameters are interpolated over the unmasked document image pixels, and a maximum likelihood (ML) classifier, based on the assumption of a Gaussian noise model of document image intensity in each class (i.e. text and background), is applied locally to each pixel, to determine whether it is text or background, based on its class membership. In order to eliminate a highly degraded background while preserving weak stroke connections, low intensity text pixels are considered in the variance estimation. This helps to improve the extraction of those connections. A simple morphological operator is applied on the ROI, followed by a simple Otsu thresholding process, to take into account the original low intensity text pixels and separate them from the background pixels.

The high- and low-intensity pixels are then used to estimate the variance. In addition, the variance on each pixel is weighted, based on its distance from the main text.

The work described above meets the objective of document image binarization. The local soft thresholding technique preserves weak connections and results in smooth, continuous text strokes. The main contribution of this work is to define a new ML-based framework for the adaptive thresholding of HDI. We applied this method to the binarization of HDI that suffer from various types of degradation, and tested it on the DIBCO'09 dataset. Its results are promising, compared to the state of the art, and remains among the best methods of document image binarization proposed to date. However, like other intensity-based binarization methods, this method has some drawbacks. It is, unfortunately, sensitive to the accuracy of the initial binary map (ROI). If the latter is not accurate (it contains pixels of interest that should not be present), the statistical parameters will not be accurately estimated. Further details can be found in chapter 3.

2.2 Multispectral restoration of historical document images

HDI restoration based solely on intensity information is considered a very difficult challenge, particularly where there is major degradation owing to the presence of objects with an intensity close to that of the main text, and in many cases appearing darker. This level of degradation dramatically affects the appearance and readability of HDI. Directly addressing such problems using conventional intensity-based methods may lead to results that are very modest at best, or totally unsatisfactory, because it is virtually impossible to properly separate two objects with intensity distributions that overlap. It is also virtually impossible for the machine to learn from non linear and unpredictable degradation. To cope with these limitations, independent intensity restoration methods are required. We propose such a method here, which is based on MS imagery, a task that is linked to the second objective of this thesis. We studied the effectiveness of MS imaging in quantitative HDI analysis, and found it very useful for the restoration of degraded document images. It is used successfully in art work and forensic manuscript investigations. This technology provides us with additional information available in the invisible light range, like IR and UV, which make it possible to see objects that cannot be seen by the human eye.

To meet the second objective of this thesis, we focus on restoration methods that can address degradation while preserving the original color and texture of HDI. To the best of our knowledge, no similar work has been published. There is one work on enhancement (Kim *et al.*, 2011) which is somewhat like ours. The aim in that work was to use IR images, as they capture

more detail than visible images, with a view to improving the contrast of the document image. In this thesis, we propose a new restoration method based on the variational approach because of its elegant formulation, which is well known in the computer vision literature. The many hours that we have devoted to examining the spectral images in our collection have helped us understand the physical phenomena involved in degradation. It is very interesting to note that objects of the same color behave differently in visible light from the way they behave in invisible light. We have found that iron-gall-based ink, which was used to create the handwritten documents in our collection, is more noticeable in visible spectral images (color) than it is in the first IR spectral image, after which it begins to disappear. However, degradation that is noticeable in visible spectral images is still visible in IR spectral images. This key feature led us to develop the mathematical formula that addresses our restoration problem. We consider degradation to be data that are missing from visible spectral images, and base our restoration model on a TV (total variation) inpainting framework aimed at isolating the degradation in IR spectral images, which we then inpaint into the visible spectral images. The restored visible spectral images are then used to reconstruct the color HDI to be displayed.

The main contribution of this work is the definition of a new TV inpainting-based restoration model that is capable of removing degradation from HDI while preserving their original colors and texture. A limitation of this and other TV inpainting-based restoration approaches is that they do not seem to work well in cases where large amounts of data are missing (larger than the inpainting scale), because they interpolate the level lines by taking into account only the initial geometric information using the shortest possible connection. As a result, instead of original missing textures being reproduced, smooth areas may be generated. For this reason, high order geometric information (i.e. length) is mandatory to accurately reproduce large amounts of missing data. Overall, the restoration model that we propose is capable of removing degradation while preserving the original appearance of the document, and the quality of the enhanced document images is for the most part satisfactory, according to the feedback provided by our collaborator. Further details on this work can be found in chapter 4.

2.3 Reference data estimation for historical document image binarization

RD are very important for comparing binarization algorithms and validating them. We now define a new method for estimating them. In the literature, there are several methods for generating RD, and researchers are trying to develop algorithms that generate them automatically or semiautomatically, in order to avoid manual generation, which is an extremely tedious task. However, the automatic and semiautomatic algorithms available have not yet reached the level

of precision of manual generation. Usually, it is necessary to involve an expert to ensure that the generated RD are correct, or to manually generate another, more accurate set. RD generation is a process aimed at assigning labels to data samples. Each label represents a particular class of data. Unfortunately, because human generation is a subjective endeavor, influenced by the personal preferences of the expert and his expertise, samples can be labeled incorrectly due to judgment errors (referred to as mislabeling), especially when the data are degraded. This obviously affects the quality of the RD. Originally introduced in the medical imaging field, the idea of estimating accurate RD based on the label assignment of multiple experts has been adopted in the HDI field to overcome this limitation. One of the well methods is the majority voting (MV). A label can be assigned to a sample if more than half the experts agree that it is the correct choice. This strategy cannot be guaranteed to be free of errors of judgment, however, because more than half the experts could vote for a label that is not the correct one, especially where there are confusing pixels (e.g. on the borders of text strokes). The process becomes even more challenging when the experts are working in a restricted representation space, such as in gray-scale or color images. To confirm the presence of errors in RD estimated by the majority voting (MV) rule and in RD generated individually by experts, the following hypothesis must be validated: Mislabeling in training data leads to a decrease in inter-class variability and a corresponding decrease in the performance of the subsequent classification tasks. For this, we define a new classification model into which we input some accurately labeled samples for training, and then we assign new labels for the samples. The output of the classification (estimated RD) is compared with the RD estimated by MV, as well as by individual experts, in terms of divergence between classes (inter-class variability). The greater the divergence, the better the estimated RD. Our results show that the RD estimated by the proposed method is more accurate than the RD estimated by other methods.

As we stated in the third objective of this thesis, we focus here on the statistical classification methods, which can estimate more accurate RD when multiple references of experts are considered. The idea is to combine the labeling uncertainty of the experts and the MS data likelihood in a unique Bayesian framework. It is interesting to note that not only does the expert judgment count, but also the probability that the samples belong to the appropriate class. Therefore, the full uncertainty about expert voting can be included in the form of a priori information, which will be updated based on observational data in a Bayesian framework, with the goal of estimating the a posteriori probabilities of pixel labels. To the best of our knowledge, no similar work has been published in the literature. The proposed model for estimating more accurate RD is a simple one. The main contribution of this work is the definition of a new,

simple classification model for RD estimation that takes into account multiple expert reference data. Further details on this work are reported in chapter 5.

CHAPTER 3

ARTICLE I: A SPATIALLY ADAPTIVE STATISTICAL METHOD FOR HISTORICAL DOCUMENT IMAGE BINARIZATION

Rachid Hedjam[1] and Reza Farahi-Moghaddam[1] and Mohamed Cheriet[1],

[1] Département de génie de la production automatisée, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Published in Elsevier Pattern Recognition journal,
Volume 44, Issue 9, September 2011, Pages 2184–2196

Abstract

In this paper, we present an adaptive method for the binarization of historical manuscripts and degraded document images. The proposed approach is based on maximum likelihood (ML) classification and uses *a priori* information and the spatial relationship on the image domain. In contrast with conventional methods that use a decision based on thresholding, the proposed method performs a soft decision based on a probabilistic model. The main idea is that, from an initialization map (under-binarization) containing only the darkest part of the text, the method is able to recover the main text in the document image, including low-intensity and weak strokes. To do so, fast and robust local estimation of text and background features is obtained using grid-based modeling and inpainting techniques; then, the ML classification is performed to classify pixels into black and white classes. The advantage of the proposed method is that it preserves weak connections and provides smooth and continuous strokes, thanks to its correlation-based nature. Performance is evaluated both subjectively and objectively against standard databases. The proposed method outperforms the state-of-the-art methods presented in the DIBCO'09 binarization contest, although those other methods provide performance close to it.

Keywords

Historical and degraded documents, Document Images Binarization, Adaptive local document image classification.

3.1 Introduction

Digital archiving of ancient and historical documents, some examples of which are shown in Figure 3.1, is an expanding trend in heritage study and preservation (Antonacopoulos and

Downton, 2007; Manso and Carvalho, 2009; Joosten, 2008). It requires that the archived images be enhanced and restored, regardless of the quality of the real acquired images. The enhancement techniques are usually based on the separation of text and background using various binarization algorithms. However, there are many challenges facing the binarization algorithms. Because of new needs in terms of accurate searching and "reading" through huge volumes of priceless historical documents which have been neglected for a long time, mainly because of the enormous labor associated with studying them in the traditional manual ways, binarization methods require that new criteria be met, such as topology preservation which is called for in word spotting (Farrahi Moghaddam and Cheriet, 2009c; Nakayama *et al.*, 1993; Steinherz *et al.*, 2000) especially in skeleton-based techniques. For historical manuscripts, the performance of OCR techniques is still low. Therefore, word spotting can be used for key word indexing.

The degradation on historical document images is, in general, physical, and can take different forms, such as the fading of ink, the presence of interfering patterns (ink bleed-through, show-through, etc.), and denotation of the cellulose structure, among others. Therefore, before any process is carried out, like feature extraction, suitable preprocessing, such as enhancement or bleed-through removal, is essential in order to correct the degradation. Continuity of the strokes and smoothness of the edges are two main factors of the binarization process on which the skeletonization step strongly depends.

In this work, a maximum likelihood model is introduced to extract text from document images. The model is locally adaptive to the input image, and uses a sub-binarization map to estimate the parameters. This enables the method to recover the missing parts of text that have been lost in the sub-binarization map (see Figure 3.2(b)). In other words, the proposed method adapts to different locations of the document image by considering spatial relations on the input image. For this purpose, the model parameters are interpolated and corrected across the image domain. The details of the method are presented in the subsequent sections.

The structure of the paper is as follows. In section 3.2, related work on the binarization of document images is discussed. In section 3.3, the problem statement is set out. The details of the model are given in sections 3.5. In section 3.6, the experimental results and evaluations are presented. Finally, in section 3.7, we provide our conclusions and prospects for future extensions.

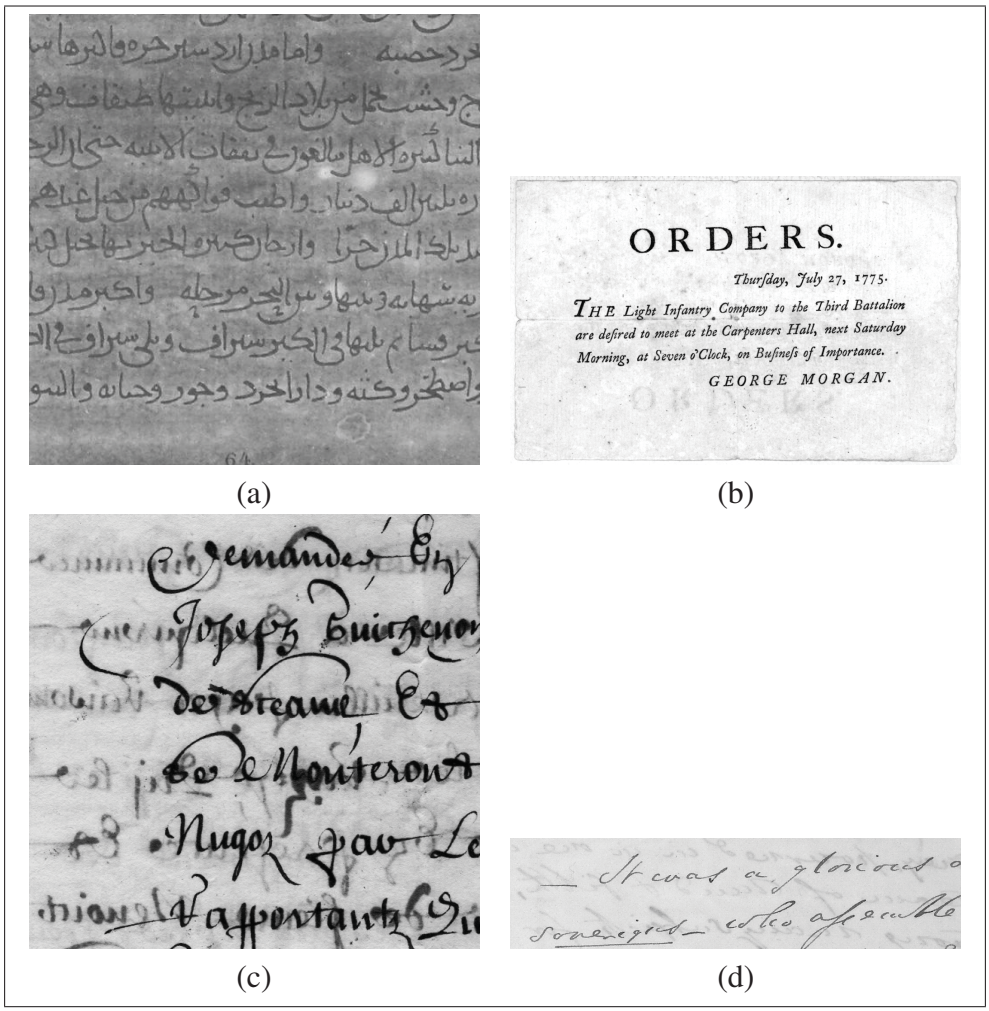
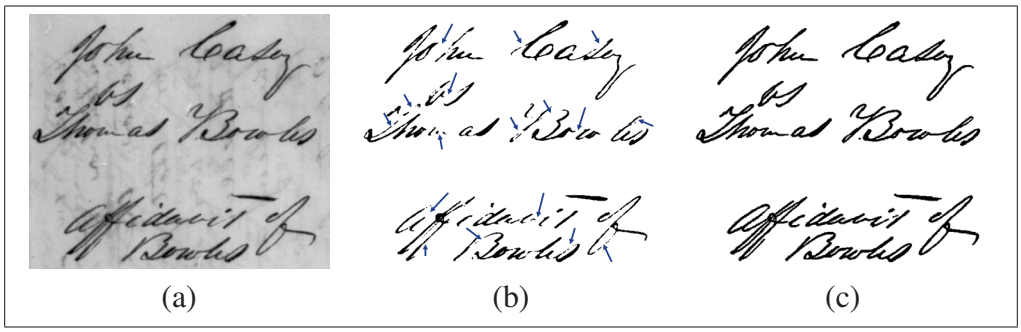


Figure 3.1 A sample set of some historical document images. (a) the original image (jum), which suffers from low-contrast degradation, (b) the original image (Gatos *et al.*, 2009a), which suffers from a degraded background, (c) the original image (Drira, 2007), which suffers from bleed-through degradation, (d) the original image (Gatos *et al.*, 2009a), which suffers from bleed-through and show-through degradation

3.2 Related work

Many methods, such as global thresholding (Otsu, 1979), local thresholding (Sauvola and Pietikainen, 2000; Bernsen, 1986; Gatos *et al.*, 2006; Farrahi Moghaddam and Cheriet, 2010b), the statistical approaches (Tonazzini *et al.*, 2007; Wolf, 2009), the entropy-based method (da Silva *et al.*, 2008), feature-based methods, such as the edge-based methods (Tan *et al.*, 2000), and multi-level classifiers (Farrahi Moghaddam and Cheriet, 2009b), have been used for the enhancement and binarization of document images. Because of the presence of gray-level degra-



(a) (b) (c)

Figure 3.2 An example of rough initialization. (a) original image, (b) initialization map (Sauvola binarization), the arrows showing the lost parts of the text, (c) ground truth map

dation over the image (i.e. shadows, non-uniform illumination, defects in some areas of the document), local thresholding methods are required to adapt to the degradation and text changes. However, these methods are ineffective when the extent of the degradation is less than the document average line height, the latter being the average distance between baselines (Farrahi Moghaddam and Cheriet, 2010b). Therefore, degradation and changes in intensity, which are very local and confined to a small part of the strokes, cannot be captured by traditional local and adaptive methods. They usually appear as holes or discontinuities on the final output. Therefore, to achieve a suitable binarization, a robust method that is capable of preserving very weak connections and strokes is needed.

One of the successful approaches to this problem is Bayesian-Markovian classification (Wolf, 2009; Geman and Geman, 1984). Some of the works associated with these methods are discussed in section 3.2. In Bayesian-Markovian classification, the Markov model controls the class of a pixel based on its spatial relationship to its neighbors. Although, this approach is very effective and intuitive, its performance in the case of highly degraded document images is limited. Even with an accurate initialization (which can be a rough binarization map), the final factor that determines the output is the complexity of the Markov model which is lower than the complexity of the degradation. Therefore, the Markov-based methods may not, for example, be able to preserve the small holes and loops while correcting the degraded text. In order to maximize the use of the *a priori* information and to avoid the complexity of the Markov model (as in the case of *a posteriori* parameter estimation), the spatial relationship is transfered to the proposed model via an initial binary map, including certain a priori information such as the average stroke width and the average line height. This information is useful for estimating the local intensity average and variance, which are at the core of the clustering problem. In fact, our model uses a simple Maximum Likelihood (ML) approach based on an assumption of

Gaussian models for the distribution of data pixels in each class. The ML, compared to other classification methods like neural networks (Marinai *et al.*, 2005) and support vector machines (Kim *et al.*, 2002), remains a popular classification tool. However, it requires a large on sample set with a normal distribution (Duda *et al.*, 2001; Jia and Richards, 1994). In addition, a major consideration in applying the ML rule is its approximation of the normal distribution and sample variances that can be used to generate confidence bounds and hypothesis tests for the parameters.

Recently, an international contest on binarization methods (DIBCO'09 (Gatos *et al.*, 2009a)) was held at the ICDAR'09 conference. The database used in the contest is available, and we use it here as a reference to compare our method with other methods. Among the binarization methods, Sauvola's is a standard document image binarization method. Its threshold value is locally adaptive, and is calculated based on a local gray level mean and standard deviation. The method has the advantage of being able to detect background regions and prevent noise pixels from appearing on the output. Although the original method is time-consuming, various computationally low cost implementations, such as an integral image method (Shafait *et al.*, 2008) and a grid-based method (Farrahi Moghaddam and Cheriet, 2010b), are available. An example of local and adaptive thresholding, which is based on the detection of edges and using information around boundaries, is presented in (Chen *et al.*, 2008). Recently, in (Gatos *et al.*, 2008), a new adaptive approach has been proposed based on the combination of several state-of-the-art binarization methods, which considers the edge information as well.

Bayesian methods have also been used for document image enhancement and binarization (Su and Mohammad-Djafari, 2007; Kuk *et al.*, 2008; Lelore and Bouchara, 2009; Tonazzini *et al.*, 2007; Wolf, 2009; Lettner and Sablatnig, 2010). In another work (Farrahi Moghaddam and Cheriet, 2009b), a Bayesian model for the binarization of historical and degraded documents is presented, which uses the output of a PDE-based enhancement method developed in the same work as the training data. Then, the trained model is used to binarize other document images. This model has been successful, especially with historical manuscripts consisting of many pages in a similar state of degradation. Its drawbacks are its ignorance of spatial correlation (however, the PDE-based part considers these data implicitly) and its inability to adapt to individual images. In this work, our model adapts to each input image and considers the spatial relation explicitly. In (Su and Mohammad-Djafari, 2007), a Bayesian model with a hidden Markov model has been proposed to binarize double-sided document images suffering from bleed-through degradation. The method is limited to linear mixtures, however, and suffers from noisy pixels on the output because of its global nature. In another work (Kuk *et al.*, 2008),

an MAP-MRF model has been used to binarize nonuniform illuminated document images. The model assumes a Gaussian distribution for the Bayes model with a fixed variance value, as well as using the same variance value for both text and background, which limits the ability of the method to adapt to different behaviors of text and background distributions. Also, it is not applicable to cases with strongly interfering patterns, such as bleed-through degradation. The main application of the model is to binarize non-uniformly illuminated document images. Another model based on MRF has been introduced in (Lelore and Bouchara, 2009). This one has limited performance in terms of preserving thin and weak strokes. A linear-mixture model has been presented in (Tonazzini *et al.*, 2007) which again suffers from noisy artifacts on the output because of its global and linear nature. In (Wolf, 2009), another MRF model is presented that requires double-sided document images, which are not always available. The method has been successful on 18^{th} century documents, although its performance decreases for large document images. Also, the small neighborhood used in the method limits the extent of its applicability. The disadvantage of the MRF-based models is that they use hand-crafted clique potentials and small neighborhood systems (4 or 8 connected systems), which is not enough to describe the complete spatial relations in the model (Lettner and Sablatnig, 2010). Variations in text and background intensity over the document image domain can be very local, and therefore a simple MRF model cannot handle them. In this work, instead of using an MRF model, we locally adapt the model parameters in the spatial domain for each pixel in order to host the spatial relations within the model. The details of this adaptation are provided in section 3.5.

3.3 Problem statement

A degraded gray value document image, suffering from various degradation phenomena, such as bleed-through, dark background, or weak strokes, is given: $u(x)$, where $x = [i, j]^T \in \Omega \subset \mathbb{R}^2$. The domain Ω is an open rectangle. We follow the BW01 representation (Farrahi Moghaddam and Cheriet, 2010b): the image values range from 0 to 1, where a 0 value means that the pixel is black, and a value of 1 means that the pixel is white. Also, we assume that as *a priori* information, a rough binarization map of the image, $u_{BW,0}$, is available. Although this approximate map may suffer from a large error, it is assumed that its recall value against the ground-truth binarization map is high. A high recall value is chosen to reduce the presence of interfering patterns (such as degraded background, bleed-through, and show-through) on the map. Although some low-intensity text will be lost on this map (as shown in Figure 3.2(b)), these parts of the text will be recovered later in the succeeding steps using a local ML classification, which will be discussed in section 3.4.

A set of parameters, such as the average stroke width and the average line height (distance between two successive baselines), is considered as *a priori* information (Farrahi Moghaddam and Cheriet, 2010b). Also, information about the presence of bleed-through on the document image is considered to be known. The goal is to binarize the true text regions on the image, and separate them from the background and possible interfering patterns. In other words, a binarized map of u, u_{BW}, will be the final output. Preserving weak connections and strokes is the main objective, as they are very important in the subsequent processing steps.

3.4 Formulation

The problem of text binarization can be considered as a binary hypothesis problem with two hypotheses, H_0 and H_1. The main hypothesis, H_0, assigns a pixel to the text, while H_1 rejects this assignment and labels the pixel with "background" (including degradation and interfering pattern pixels). Usually, in order to make a decision between H_0 and H_1, posterior probabilities calculated based on the data and *a priori* information are used. Let us denote these probabilities as $P\left(H_0|u, u_{BW,0}\right)$ and $P\left(H_1|u, u_{BW,0}\right)$ respectively. To make the model spatially adaptive, it is assumed that each pixel on the image has its own binary problem. This will be discussed later in section 3.5.2.

There are several decision rules that can be used in Bayesian models: maximum likelihood (ML), minimum probability of error, maximum a posteriori, and Bayes risk decision (Jiang and Mahadevan, 2007). Although ML does not consider the *a priori* information on the classes, this cannot be considered a drawback. Determination of the models for classes is a very difficult task because of the complexity and variability of text and background on the historical documents. At the same time, the local nature of the method enables it to adapt to possible variations on the input image. In particular, variations in text and background intensity over the document image domain can be very local, and therefore a simple model for $P\left(H_i\right)$ is not able to handle them.

Considering the well-known behavior of text and background data in the features space, two simple, basic models can be use to model each of them: histogram-based model (Mezghani *et al.*, 2008) and Gaussian models. In histogram-based models, the model of each class is estimated according to its probability density function (Farrahi Moghaddam *et al.*, 2009):

$$PDF_{\omega}(k) = \frac{1}{\sqrt{2\pi}hn_{\omega}} \sum_{j}^{n_{\omega}} n_{\omega,j} \exp\left\{\frac{-(k-j)^2}{h}\right\}$$

where $\omega \in \{t, b\}$, n_ω is the total number of pixels in class ω computed based on $u_{BW,0}$, j represents the bins of the histogram, $n_{\omega,j}$ is the number of pixels in class ω having the intensity in bin j, and h determines the level of smoothing. For example, $PDF_t(u(x))$ can be used as $P\left(u(x)|H_0\right)$.

In document image processing, a single-mode distribution for text and a multi-mode for background are usually used (Otsu, 1979; Cheriet *et al.*, 1998). This is because of the complex nature of the background on degraded document images. Normally, a Gaussian model is used for each mode in these distributions (Otsu, 1979; Cheriet *et al.*, 1998; Sezgin and Sankur, 2004). For the case of the background distribution, we skip the assumption of having a multi-mode distribution, because the proposed method adapts a separate model to each pixel on the document image. Therefore, the distributions are highly local, and so in most cases the background distribution of a pixel is a single-mode one. This allows us to assume two Gaussian models, one for text and one for background:

$$P\left(u(x)|H_0\right) = P\left(u(x)|\text{text}\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(\frac{-1}{2}\frac{\left(u(x)-\mu_t\right)^2}{\sigma_t^2}\right) \tag{3.1}$$

and

$$P\left(u(x)|H_1\right) = P\left(u(x)|\text{bkgd}\right) = \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(\frac{-1}{2}\frac{\left(u(x)-\mu_b\right)^2}{\sigma_b^2}\right) \tag{3.2}$$

where μ_ω and σ_ω are the mean and standard deviation for $\omega \in \{t, b\}$ and should be determined, and $u(x)$ represents the gray value of a pixel x. Applying the decision rule, the classification problem is equivalent to the following minimization problem:

$$u_{BW}(x) = \arg\min_{w\in\{t,b\}} \left\{ \frac{\left(u(x)-\mu_w\right)^2}{2\sigma_w^2} \right\} \tag{3.3}$$

The estimation of μ_t, μ_b, and σ_b will be discussed in the next section. As a special case, the parameter σ_t will be discussed in subsection 3.5.4.

3.5 Methodology

In most documents (degraded or not), the text is blacker than the degradation, but, at the same time, it may contain some strokes that are degraded and have the same intensity as the degradation, or less (see Figure 3.3(a)). The proposed method is composed of two steps. The first step consists of under-binarization of the document image by eliminating, to the maximum extent possible, the degradation information and keeping only the blacker part of the text. In this case, interfering patterns, noise, and even weak and low intensity strokes will be erased (see Figure

3.2(b)). In the first step, the grid-based Sauvola binarization method (Farrahi Moghaddam and Cheriet, 2010b) is used, because it can produce an under-binarized image when high k values are used. The k value should be close to one in the case of document images with severe degradation like bleed-through. The second step recovers the missing text strokes (lost in the previous step) using a local binarization based on ML classification. The details are provided in the subsequent subsections.

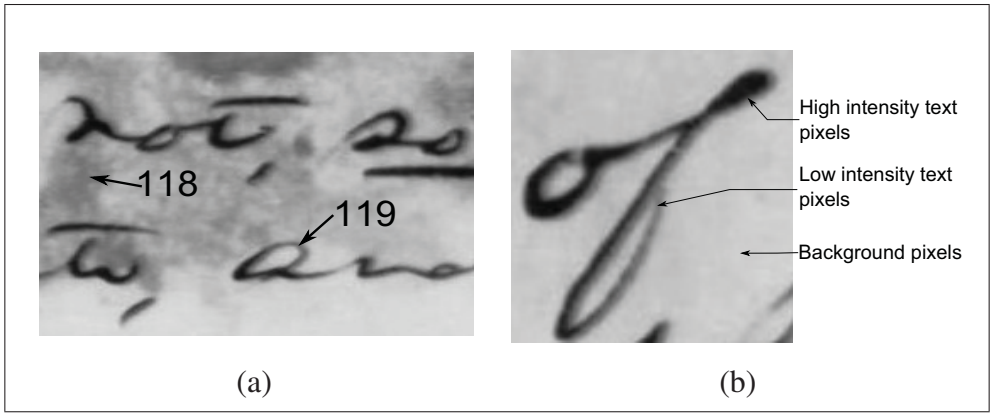


Figure 3.3 (a) An example of degraded text. This is a part of one image of the data set used in this work. It is zoomed to display the local variation in intensity where some text pixels are brighter than the degraded background. (b) An example of text containing low-intensity pixels. For better visualization only a part of the input image (Figure 3.14) is shown

3.5.1 Sauvola binarization algorithm

As the first step, the input image u is binarized with a simple binarization method; in our case, we used Sauvola's method which processes each pixel locally and independently of the rest of the pixels. It is claimed that it gives good results in the binarization of a document image in which the background contains large variations and non-uniform illumination. This algorithm calculates a threshold for each pixel according to the information contained in its local neighborhood. The idea of the method is to vary the threshold in the image based on the values of local mean and local standard deviation in a window centered on each pixel:

$$T_{\mathrm{G,Sau},u}(x) = m_G(x)\left(1 + k\left(\frac{\sigma_G(x)}{R} - 1\right)\right) \tag{3.4}$$

where m and σ are the mean and standard deviation of the whole local window, R is the maximum value of the standard deviation ($R = 0.5$ for a gray-scale document image), and k is a

parameter that takes positive values. The subscript G refers to grid-based modeling (Farrahi Moghaddam and Cheriet, 2010b). The threshold is adapted using a mean and a standard deviation according to the local contrast. If the contrast is low in some regions ($\sigma \ll R$), it means that the threshold drops below the mean value, which allows the relatively dark regions of the background to be removed. The parameter k controls the threshold in a local neighborhood. If k is high (towards 1), then only the darkest objects remain.

3.5.2 Spatially adaptive model

As discussed in the introduction, a large proportion of the data of a document image is in the form of the spatial relations among pixel data. Unfortunately, the spatial relations behave differently on different scales (or distances). Therefore, the task of modifying the feature vector to include these data would be very different in the case of small scales relative to large scales. In a typical processing patch, both small and large scales are presented, which requires a *multi-scale* representation of spatial relations in the feature vector, which is a difficult task. This is why, the spatial data are ignored in many approaches, particularly in the statistical ones. In brief, in our model, μ_ω and σ_ω are replaced with their localized equivalents $\mu_\omega(x)$ and $\sigma_\omega(x)$, which are in turn computed in a neighboring system using a sliding window (patch). Now, the problem, (3.3), can be rewritten as follows:

$$u_{BW} = \arg \min_{w(x)\in\{t,b\}} \sum_x \left\{ \frac{\big(u(x) - \mu_{w(x)}(x)\big)^2}{2\sigma_{w(x)}(x)^2} \right\} \tag{3.5}$$

The exact definitions of the model parameters are as follows: $\mu_t(x)$, the gray mean value of class text at pixel x, $\sigma_t(x)$ the gray standard deviation value of the class text at pixel x, $\mu_b(x)$, the gray mean value of the class background at pixel x, and $\sigma_b(x)$ the gray standard deviation value of the class background at pixel x. The neighborhood size used to compute these parameters will be discussed in subsequent subsections.

3.5.3 Computing the fields of μ_t, μ_b, and σ_b

In the first step, the input image u is binarized with a simple binarization method. In our case, we used the grid-based Sauvola method (Farrahi Moghaddam and Cheriet, 2010b) which processes each pixel locally and independently of the rest of the pixels (Figure 3.4(b)). Also, a few postprocessing steps, such as the removal of noise pixels, are applied to the output of the method. *A priori* information, such as the average stroke width, is used to make the postprocessing steps parameterless. The result is then called the initial map, $u_{BW,0}$, (see Figure 3.4(b)),

and is used as a mask to generate two new maps: one containing only the text data u_t (Figure 3.4(c)) and the second, u_b containing only the background data (Figure 3.4(d)). The two maps are governed by the following equation:

$$u_\omega = \begin{cases} u, & \text{if } u_{BW,0} = \omega; \quad \omega \in \{t, b\}; \\ \text{NaN}, & \text{otherwise.} \end{cases} \tag{3.6}$$

For statistical purposes, the incomplete data were assigned as the notation NaN, which means *not a number* (Solit, 2006).

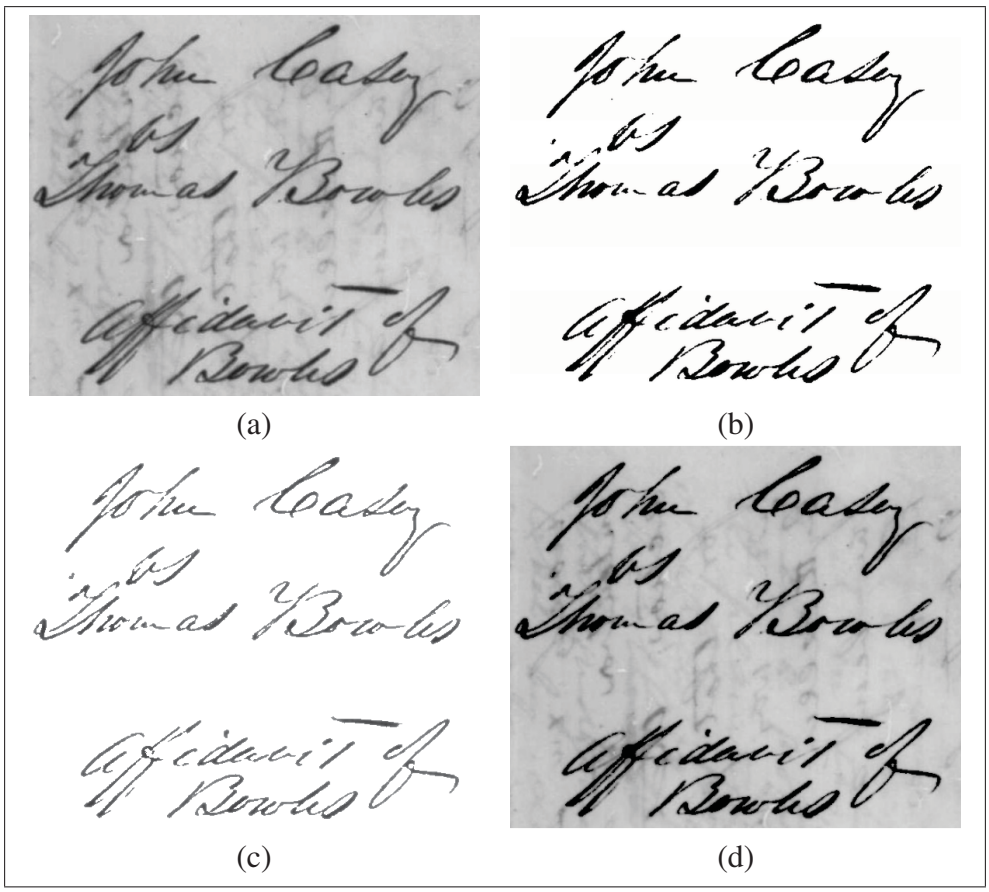


Figure 3.4 An example of how we produce the masked data (text and background). (a) original image, (b) initialization map (Sauvola binarization), (c) masked text data, the white area containing NaN values, this image is used to compute μ_t and σ_t. (d) masked background data, the black areas containing NaN values: this image is used to compute μ_b and σ_b

Now, we consider the problem of estimating the local parameters (local average and variance of intensity). Given the area delimited by the patch J in Figure 3.5(a), as an example of background data, we estimate $\mu_b(x_j)$ and $\sigma_b(x_j)$ by including only the information belonging to the gray pixels in that patch (but not to the black ones, which contain NaN values), as in the following equations:

$$\mu_w(x) = \frac{1}{|N_w(x)|} \sum_{\substack{x_k \in N_w(x) \\ u_w(x_k) \neq \text{NaN}}} u(x_k); \quad w \in \{t, b\} \tag{3.7}$$

$$\sigma_w^2(x) = \frac{1}{|N_w(x)|} \sum_{\substack{x_k \in N_w(x) \\ u_w(x_k) \neq \text{NaN}}} \left[u(x_k) - \mu_w(x)\right]^2; \quad w \in \{b\} \tag{3.8}$$

where $|N_w(x)|$ represents the number of the neighborhood pixels x_k around the main pixel x, except those having NaN values. In another case of the same type, for the pixel x_i in Figure 3.5(a), the estimation of $\mu_b(x_i)$ and $\sigma_b(x_i)$ is not possible, because the patch I centered on pixel x_i (called a singular pixel) does not contain any background information. In the next paragraph, we explain how to estimate the local parameters of a singular pixel.

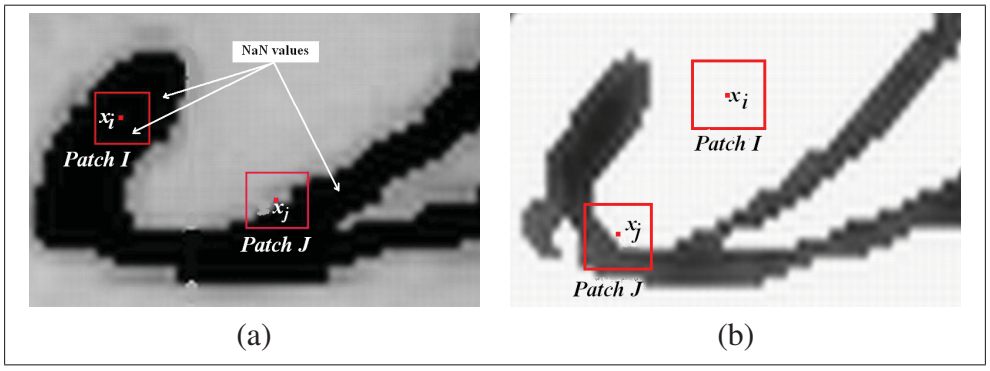


Figure 3.5 An example showing how to compute the local parameters. (a) computing $\mu_b(x_j)$ and $\sigma_b(x_j)$ for the background data. In practice, the black area contains NaN values which are not considered for the calculation of the parameters. The parameter of the pixel x_j in patch J is computed by using only the gray text pixels (not the black ones) belonging to the corresponding patch. (b) the same applies to estimating the text parameters: for example, to estimate the $\mu_t(x_j)$ in patch J, we take only the gray pixels in that patch and ignore the white pixels, which contain NaN values (painted in white to distinguish them from the gray text)

Since the μ_t, μ_b, and σ_b fields contain a large number of unknown NaN values, an inpainting technique using the averaging method (Bertalmio *et al.*, 2000) is applied to fill in the empty areas (NaN values or singular pixels). The inpainting process permits propagation of the values on the empty area with control of the variation on the gray values while avoiding their amplification, as shown in Figure 3.6. In order to ensure a stable output, we first replace all the border pixels of μ_t by the average value of its known values, and then we apply the inpainting process, as in the following equation:

$$\mu_t \left(x \mid x \in \partial\Omega \right) = M \left(\{ \mu_t(x) \mid \forall x \in \Omega; u_t(x) \neq \text{NaN} \} \right) \tag{3.9}$$

where $M(\{z\})$ is a function computing the mean value of the set $\{z\}$, and $\partial\Omega$ is the border of the domain Ω. For the other fields (μ_b and σ_b), a 0 value is used on the border. By applying the inpainting procedure, the smooth and complete fields become available. A main parameter to be set is the patch size for calculating μ_w and σ_w, $w \in \{t, b\}$. As mentioned before, to reduce the computational cost, we use the grid-based approach (Farrahi Moghaddam and Cheriet, 2010b), in which all functions and variables are represented by their equivalent variables on a grid of size d_c. In this approach, the patch size is represented by $2d_c+1$. We set d_c as a factor of W_s: $d_c = W_s$, where W_s is the average stroke width (Farrahi Moghaddam and Cheriet, 2010b), one of the a priori parameters. Therefore, a patch containing enough information from both the text and the background classes will be used to estimate the local mean and standard deviation.

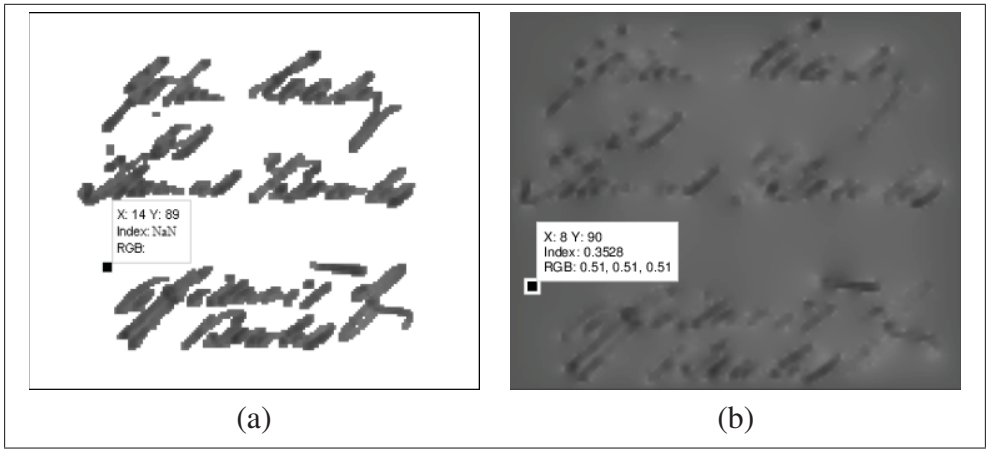


Figure 3.6 An example of how to inpaint the data. (a) μ_t only for masked text data, the white area containing NaN values, (b) inpainted mean text data values μ_t

3.5.4 Estimation of the σ_t field

To estimate σ_t, a global estimation of the standard deviation of the text intensities is first calculated. Then, this value is mapped onto the image domain using a spatial Euclidean distance. In this way, very local and sharp fluctuations in the σ_t field are avoided.

3.5.4.1 Estimation of the global σ_t: S_t

Local text variations are very small, due to the constant local intensity of the text. Therefore, any estimation of text variance based on u_t will result in a small value, and will prevent the model from capturing the degraded parts of text strokes with low intensities in the subsequent classification process. This phenomenon leads to cuts and discontinuities in the strokes, which, as discussed above, have a severe impact on the performance of the subsequent process of understanding. To resolve this problem, we may use another map instead of $u_{BW,0}$, which is able to catch the degraded parts of strokes and give a more accurate estimation of σ_t. This is achieved by dilating the text region in the initial map with a structuring element of radius $r = W_s$, in order to include low intensity text pixels of the original image u in the calculation of the new variance of text S_t^2. This will increase the estimated text variance, because of the presence of many low intensity pixels. Let us call the dilated map $u^D_{BW,0}$ (Mukhopadhyay and Chanda, 2003). The data in the new masked text field u_t^D, which is obtained using $u^D_{BW,0}$, consists of both text and background data. In order to remove the background data, Otsu's method is applied. Let us consider P as the set of pixels extracted from the image u by using $u^D_{BW,0}$ as a mask:

$$P = \{x \mid x \in \Omega, \quad u^D_{BW,0}(x) = 0\} \tag{3.10}$$

P contains a subset of high intensity text pixels (H), a subset of degraded low intensity text pixels (L), and a subset of background pixels (B) (see Figure 3.3(b)):

$$P = H \cup L \cup B \tag{3.11}$$

Since the values in L are more similar to those of H, a simple Otsu classification allows the division of P into two subsets, T and B (see Figure 3.7(b)), as follows:

$$P = T \cup B \tag{3.12}$$

where $T = H \cup L$. The remaining data, $u_t^{D'}$, cover both strong and degraded text intensities, and are therefore suitable for calculating the standard deviation.

$$u_t^{D'}(x) = \begin{cases} u(x) & \text{if } x \in T \\ \text{NaN} & \text{otherwise} \end{cases} \tag{3.13}$$

The standard deviation value of the dilated text is given by the STD function:

$$S_t = \text{STD}\left(\left\{u_t^{D'}(x) \mid \quad x \in \Omega, \quad u_t^{D'}(x) \neq \text{NaN}\right\}\right) \tag{3.14}$$

Figure (3.7(b)) shows how Otsu's thresholding separates P into two subsets. The new variance S_t^2 is then greater than σ_t^2. The standard deviation values of the map σ_t are substituted by S_t, following the spatial correction described in the next section. It is worth noting that, although the intensity variation of the text pixels across the image domain may result in the loss of some of the text pixels, the remaining pixels in T could provide an approximation of the text intensity variation induced by degradation. Because of the local nature of the method, an approximation of the text variance is enough to recover the degraded text pixels.

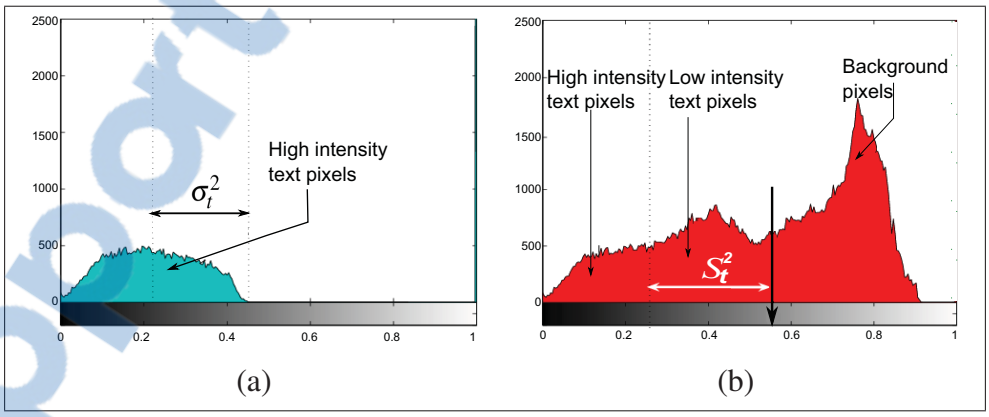


Figure 3.7 Estimation of the global S_t. **(a)** histogram of the pixels of Fig. 3.8(c), **(b)** histogram of the pixels of Fig. 3.8(d). We can clearly see that $S_t^2 > \sigma_t^2$, and this is because of the low intensity text pixels

3.5.4.2 Spatial adaptation of σ_t

On some images, parts of the document background are very dark, and the intensities on these areas are very close to the text intensity. These pixels can appear as wrong patterns on the

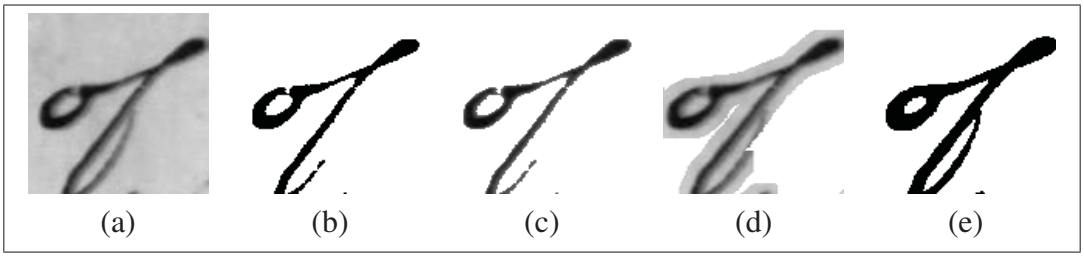


Figure 3.8 An example of how low intensity text pixels are recovered. (a) part of the original image shown in Figure 3.14, (b) the output of Sauvola's method (initialization map), (c) text stroke masked by (b), (d) text pixels masked by the output dilation of (b), (e) the result of the binarization method. The output of the whole image is presented in Figure 3.14

output because of misclassification. In order to prevent this effect, we define

$$\sigma_t(x) = S_t \exp^{-d(x)} \tag{3.15}$$

where $d(x)$ is the Euclidean distance (Fabbri *et al.*, 2008) of a background (white) pixel x from the black text region computed on the initial map $u_{BW,0}$. When a degraded pixel x is far from the main text, $\exp^{-d(x)}$ becomes very small (depending on its distance d), and therefore, the degraded pixels will automatically be labeled as background in the classification step.

3.5.5 Estimation of u_{BW}

Having all four text and background features ready on the image domain, the decision rule can be applied to the minimization problem (3.5) to estimate the final binarization. The classification problem can be described as follows:

$$u_{BW}(x) = \begin{cases} 0 & \text{if } \frac{(u(x)-\mu_t(x))^2}{2\sigma_t^2(x)} < \frac{(u(x)-\mu_b(x))^2}{2\sigma_b^2(x)} \\ 1 & \text{otherwise} \end{cases} \tag{3.16}$$

Although direct calculation can be used to compute $\mu_\omega(x)$ and $\sigma_\omega(x)$, grid-based modeling is used (Farrahi Moghaddam and Cheriet, 2010b) in order to reduce the computational cost and also to smooth these functions. In this approach, all functions and variables are calculated from their equivalent variables on a grid of size d_c. In this way, while the pixel-wise nature of the model is preserved, functions and variables can be transferred easily to any suitable scale. Figure 3.9, shows the grid operation (Farrahi Moghaddam and Cheriet, 2010b) for estimating

the parameters in a few nodes and then for interpolating the remaining parameters values. A schematic algorithm of the proposed method is provided in Algorithm 1.

Algorithm 1: Principle of the proposed algorithm

Input: The original image u;
Under-binarization of u;
Calculate the under-binarized map $u_{BW,0}$ using Sauvola's algorithm; c.f.; Figure3.2(b);
Get the masked text u_t and masked background u_b maps c.f.; Figure (3.4);
Parameter estimation;
Estimate μ_w , $w = \{t, b\}$ and σ_b; eq. (3.8);
Use the inpainting technique to propagate the estimated values in the empty areas;
Global estimation of σ_t;
Dilate the text of $u_{BW,0}$ to get $u^D_{BW,0}$;
Get the masked dilated text P from u using $u^D_{BW,0}$, eq. (3.10);
Separate P into two subsets: T and B. T is the part of P that is associated with text;
Compute the standard deviation of T, S, eq. (3.14);
Spatial adaptation of σ_t;
$\sigma_t(x) = S_t \exp^{-d(x)}$; eq. (3.15);
Calculate the binarized image of u, eq. (3.16);
Output: The binarized image u_{BW};

3.6 Experimental results and discussion

The performance of the proposed method is evaluated on different datasets. The first dataset consists of samples from the Google Book Search dataset (Google, 2007) which have suffered degradation. The Google Book Search dataset contains the scanned, resized, and flattened images of several books. The DVD of the dataset contains data for 68 books. The second dataset is from the Juma Al Majid Center (Dubai) (jum), and contains 85 images (about 160 document pages) of very old and degraded text. The document images were acquired by the center using camera imaging (a 10.3 megapixel CCD camera), at an acquisition resolution of 150 DPI. The content varies from normal handwritten text to very complex maps. The third dataset is from the Institute of Islamic Studies[1] (IIS) at McGill University, and consists of 280 images corresponding to a 260-page manuscript. The document images were acquired using camera imaging (a 21 megapixel CCD camera), at an acquisition resolution of 300 DPI. Finally, the fourth dataset is a Latin and Arabic manuscript dataset (Deriche and Faugeras, 1996), which contains a large number of ancient documents, including Arabic manuscripts

[1] http://www.mcgill.ca/islamicstudies

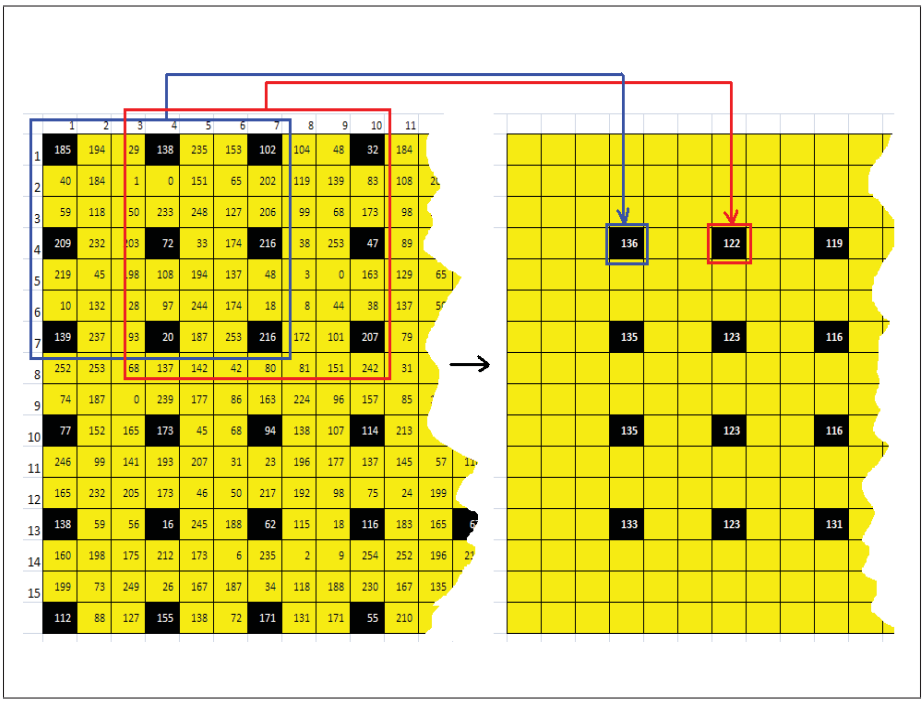


Figure 3.9 Example of how the grid works. Left, a sample image, the black squares representing the nodes of the grid, and the distance between two successive square equal to d_c. Right, the values in the black squares represent the calculated values of the grid. Each of them is computed by the function performed on all values belonging to the windows shown on the left. In this example, the input function is the mean function

from Tombouctou, Mali. Below, the performance of the method is evaluated both subjectively and objectively.

3.6.1 Subjective evaluation

Figure 3.10 shows the performance of the method for a subimage from the Google dataset. The results of two standard methods are also shown in the figure for comparison purposes: Otsu's method (Otsu, 1979), which is a global thresholding method, and the local-thresholding method of Sauvola (Sauvola and Pietikainen, 2000) are used to binarize the input image. Because of the variations in text intensity, none of the global or adaptive methods is able to preserve weak connections and provides smooth, continuous output. As can be seen from the figure, the outputs of both methods suffer from cuts and false holes, which have a highly negative effect on the performance of subsequent skeletonization processes.

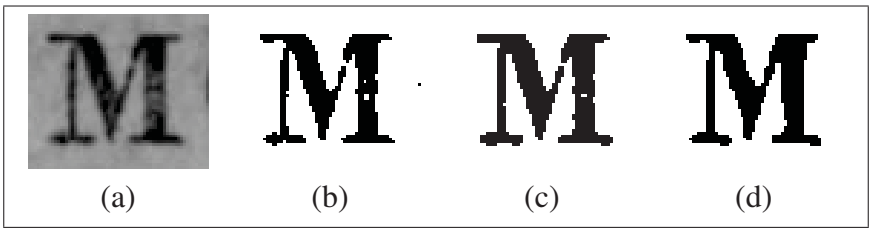


Figure 3.10 An example of degraded shapes with variable intensities. (a) part of a degraded image (selected for better visualization), (b) the output of Otsu's method, (c) the output of Sauvola's method, (d) the output of the proposed method, which is continuous and smooth

It is worth noting that holes and loops are very important in the understanding steps, such as word spotting, and should be preserved. Because of the presence of loops and holes in the correct shapes, morphological opening/closing operators cannot be applied in many cases. At the same time, the output of the proposed method is continuous, and very weak connections are preserved thanks to the very local and correlation-based nature of the method. In Figure 3.11, the same behavior can be observed for a sample from the McGill dataset. Not only do the global and local thresholding methods suffer from false discontinuities and holes, some parts of the background appear on the outputs. Again, the proposed method provides a very smooth, continuous output.

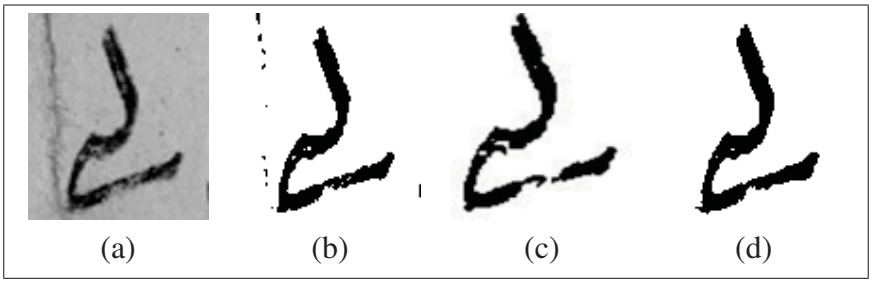


Figure 3.11 A second example of degraded shapes with variable intensities. (a) the original image, (b) the output of Otsu's method, (c) the output of Sauvola's method, (d) the output of the proposed method which is continuous and smooth

The final example, shown in Figure 3.12, suffers from the bleed-through effect. As can be seen from the input image (Figure 3.12(a)), the degree of the bleed-through effect is high, and at some regions the intensity of the interfering patterns is very close to the true text intensity. Also, it is worth noting that only the image of one side of the document is available and the double-sided bleed-through removal techniques (Tonazzini *et al.*, 2007; Farrahi Moghad-

dam and Cheriet, 2009a) cannot be applied in this case. The outputs of the Otsu and Sauvola methods are shown in Figures 3.12(b) and 3.12(c). None of them is successful in removing the bleed-through, and the segmented texts suffer from a high degree of pattern interference. The output of the proposed method is shown in Figure 3.12(e). The method is not only able to remove the bleed-through interfering patterns, but also provides continuous strokes which are ready for the subsequent processes. Because of suitable initialization using the recursive Sauvola method (Farrahi Moghaddam and Cheriet, 2010b) and the spatially adaptive classification of the proposed method, the interfering patterns are largely separated from the true text, even though their intensity is very close to the text intensity. This example confirms the discussion in the introduction on the importance of the a priori information. The initial binarized map obtained using the recursive Sauvola method is the key to separating the bleed-through patterns. In the case of a simple Markov model, the degradation pixels can easily be added to the output. We are currently working on a combined model which contains a Markov part adapted to the a priori information.

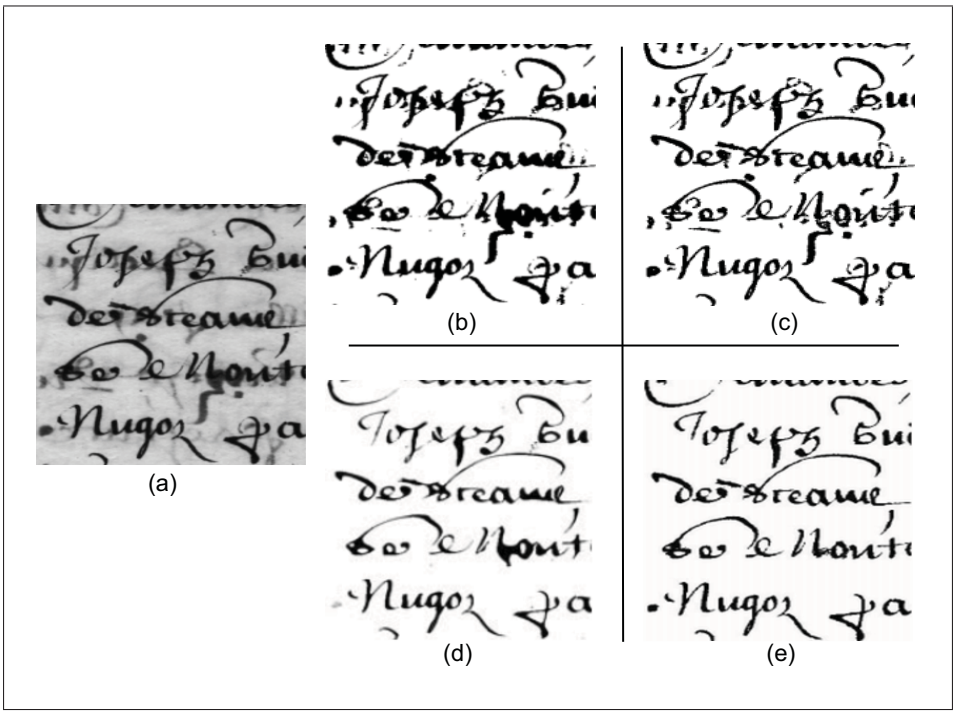


Figure 3.12 A third example of degraded shapes with a strong interference of bleed-through. (a) the original image, (b) the output of Otsu's method, (c) the output of Sauvola's method, (d) the output of Drira's method (Drira, 2007), (e) the output of the proposed method

In Figure 3.1, a set of sample degraded document images from various datasets is provided. The documents suffer from degraded background and bleed-through. For example, Figure 3.1(a) has both degraded background and degraded text. The outputs of the proposed method for this image are shown in Figure 3.13(a). The background is clean and weak strokes are recovered. The outputs of the proposed method for the other samples in Figure 3.1 are also shown in Figure 3.13. The method is able to remove strong bleed-through interfering patterns. It is worth noting that for Figure 3.1(c) the recursive Sauvola method (Farrahi Moghaddam and Cheriet, 2010b) has been used to obtain $u_{BW,0}$.

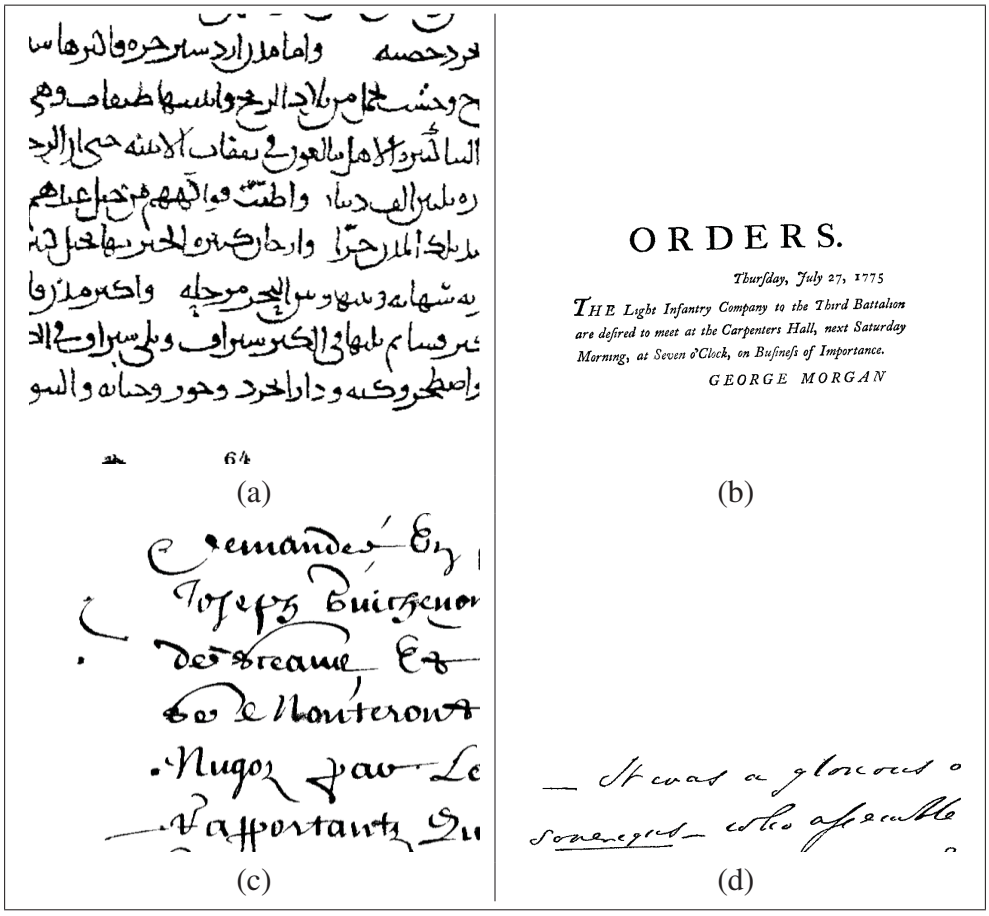


Figure 3.13 Binarization results of the images in Figure 3.1

Another subjective comparison with the-state-of-the-art is shown in Figure 3.14. Figure 3.14(a), which is from the DIBCO'09 dataset, contains a very degraded background and text. Figure 3.14(b) is the result of Sauvola's thresholding method (Sauvola and Pietikainen, 2000), which is able to recover the text, but it cannot remove the strong background degradation. Also, the

output suffers from noise. Figure 3.14(c) , which is taken from (Su *et al.*, 2010), performs well in extracting text from background, but its ability to recover the degraded text strokes is limited. The result of the proposed method is shown in Figure 3.14(d): the degraded strokes of the text are recovered, while the degraded background is eliminated. However, the proposed method fails to recover very small loops in characters, especially when they belong to a highly degraded background. We are working to generalize the model to cover this case of degradation as well.

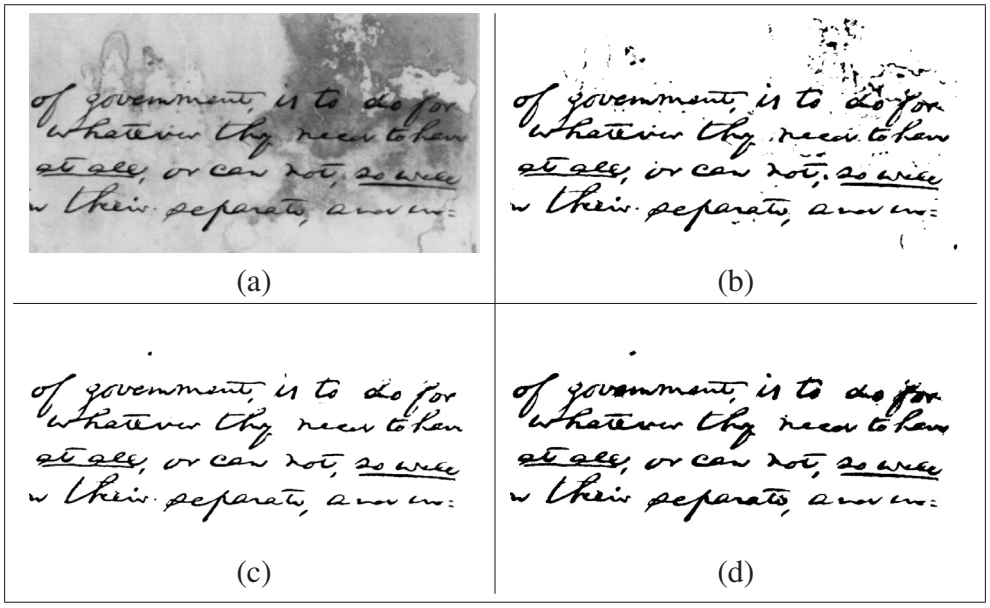


Figure 3.14 Subjective comparison with the state of the art, (a) original image, (b) Sauvola thresholding result, (c) Su and Tan (Su *et al.*, 2010) result, (d) the proposed method result

3.6.2 Objective evaluation against DIBCO'09 (Gatos *et al.*, 2009a)

The objective evaluation has been performed using the DIBCO'09 dataset. The results of the proposed method are compared to the top three methods of the DIBCO contest (Gatos *et al.*, 2009a), involving 43 participants from 22 laboratories worldwide. It is worth noting that the dataset was not available to the participants in the contest. We used the available input images simply to determine their class: low contrast images, images with a degraded background, and images with the bleed-through problem, as will be defined in section 3.6.2.1. Before conducting an objective evaluation, a brief description of a few binarization methods used for this purpose is provided. The first algorithm, which was proposed by S. Lu and

C.L. Tan (Gatos *et al.*, 2009a) and won first place in the competition, consists of four parts, dealing with document background extraction, stroke edge detection, local thresholding, and postprocessing. The local threshold is estimated by averaging the detected edge pixels within a local neighborhood window. The second algorithm considered was proposed by J. Fabrizio and B. Marcotegui (Gatos *et al.*, 2009a), and is based on the toggle mapping operator. The image is first mapped onto the corresponding morphological erosion and dilation. Then, if the pixel value is closer to the erosion, it is marked as background, otherwise it is marked as foreground. To avoid salt-and-pepper noise, pixels with erosion and dilation that are too close, are excluded from the analysis. Pixels are then classified into three classes: foreground, background, and homogeneous regions. Finally, the homogeneous regions are assigned to the foreground or background class, according to the class of their boundaries. The third algorithm, proposed by D. Rivest-hénault, R. Farrahi Moghaddam and M. Cheriet. (Gatos *et al.*, 2009a), uses a curve evolution approach implemented in the level-set framework to erode a stroke map (Farrahi Moghaddam and Cheriet, 2009b) estimated from some statistical information obtained from the content of the input image. The evolution of the curve is based on a term of energy measuring the fitness of the stroke pixels with respect to the stroke's gray level image. The fourth algorithm used in the comparison task is the well-known Sauvola method, which is based on a local estimation of variance and mean to obtain the threshold by which the pixel will be classified as either text or background (Sauvola and Pietikainen, 2000).

3.6.2.1 Evaluation setup

In all the experiments, we have considered the following internal parameters for our image binarization model. The first step consists of creating a black and white initialization map, $u_{BW,0}$. The Sauvola binarization algorithm integrated on a grid function (Farrahi Moghaddam and Cheriet, 2010b) is used to generate this map by empirically fixing the line height parameter h_l to 100 pixels, and the parameter k is 0.2, 0.4, 0.5, and 0.9 for the low contrast, faded text, degraded background, and bleed-through degradation types respectively. The size of the grid cell d_c and the radius of the structure element r are proportional to the stroke width: $d_c = \lambda_{d_c} W_s$ and $r = \lambda_r W_s$, where W_s is the stroke width, and λ_{d_c} and λ_r are proportional factors. In our experiments, $\lambda_{d_c} = 2$ and $\lambda_r = 1/4$. The stroke width is estimated based on a statistical method described in (Farrahi Moghaddam and Cheriet, 2010b).

3.6.2.2 Performance measures

We measure the similarity between the output of the proposed algorithm and the ground-truth made available by the DIBCO'09 organizers. The performance measures used to evaluate the proposed algorithm are the well-known *recall* (R), *precision* (P), and *F-measure*:

$$\text{F-Measure} = \frac{2 \times R \times P}{R + P}; \quad R = \frac{\text{TP}}{\text{TP} + \text{FN}}; \quad P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.17}$$

where TP, FP, TN, and FN denote the true positive, false positive, true negative, and false negative values, respectively.

We have replicated the scenario used in the evaluation of the state-of-the-art binarization methods described in the DIBCO contest (Gatos *et al.*, 2009a). The dataset consists of five machine printed images (referred to as PR images and labeled as P0X images, where X is from 1 to 5) and five handwritten images (referred to as HW images and labeled as H0X images, where X is from 1 to 5), resulting in a total of ten images for which the associated ground-truths are ready. One of the images is shown in Figure (3.16(a)). These images suffer from a different kind of degradation, one frequently found in old documents (e.g. variable background intensity, shadows, smearing, smudging, low contrast, bleed-through, and show-through).

Table 3.1 Performance in terms of precision, recall, and *F-measure* of the proposed algorithm

image No.	Image	Recall	Precision	F-measure
1	H01	94.873	88.571	91.614
2	H02	92.574	92.349	92.461
3	H03	89.143	91.164	90.142
4	H04	95.997	81.150	87.952
5	H05	85.924	83.670	84.782
6	P01	87.883	95.709	91.630
7	P02	94.247	97.833	96.007
8	P03	95.298	98.854	**97.043**
9	P04	95.847	87.800	91.647
10	P05	89.213	91.343	90.266
	Average	92.100	90.725	91.354

Table 3.1 shows the *recall*, *precision* and *F-measure* results obtained by the proposed algorithm, and Table 3.2 compares the average *F-measures* generated by the various methods. In Table 3.1, we observe that the precision is higher than the recall in the case of the HW images (see Table 3.1). This means that the algorithm is able to recover the weak strokes by attaching the discontinuous parts. These gaps are mostly created because of low-intensity text caused by ink shortage on some parts due to rapid movement of pen on paper during writing. It is worth noting that the dilation operator of the algorithm may result in the addition of some undesirable pixels, especially on the border of a stroke. In the case of HW images, we observe that the recall measure is higher than the precision. This means that the algorithm is more effective at filling the gap. This is due to the way in which the ink flows onto the paper. Generally, writers move the pen horizontally over the paper, so the amount of ink is reduced if the writer does not lift the pen from the paper to move to the next piece (see Figure 3.15(a)). Therefore, in the gap areas, the intensity is closer to the local neighbor text than to the local neighbor background. This helps the proposed algorithm to classify these gaps as text. In contrast, printing and typesetting devices inject ink vertically onto the paper, so the discontinuities of the text are sharp, and the missed parts of strokes are more similar to the background (see Figure 3.15(b)). This is why the algorithm classifies these areas as background, although they are more likely to be text. From the same table, we found that the average *F-measure* of the PR images (93.318%) is greater than the average *F-measure* of the HW images (87.958%). The explanation for this is that most of the HW images are more degraded, and many parts of the text have a low intensity, close to the intensity of the background. The smallest *F-measure* (84.782%) is given by image H05, which is a very degraded document where many parts of strokes are quite close to the background in terms of intensity. For this type of document images, because some strokes may be missed on the initialization map, achieving a good initialization map for the Bayesian classification process is not easy. This is why their *precision* measures are mostly lower for the initialization. For, the HW images, H02 has the best *F-measure* (92.461%), because the text in this image is more regular and does not contain low intensity pixels. Also, we have *a priori* information about the existence of bleed-through, so the initialization map (obtained using the recursive Sauvola method (Farrahi Moghaddam and Cheriet, 2010b)) is more stable and helps the Bayesian classification to separate the text from the interfering objects accurately. For the PR images, the highest *F-measure* corresponds to image P03, because it has less degradation, Thus, the classification problem is more efficient. Image P05 contains many interfering objects and many parts of the text are interrupted, which reduces its *F-measure* in comparison to that of the other PR images.

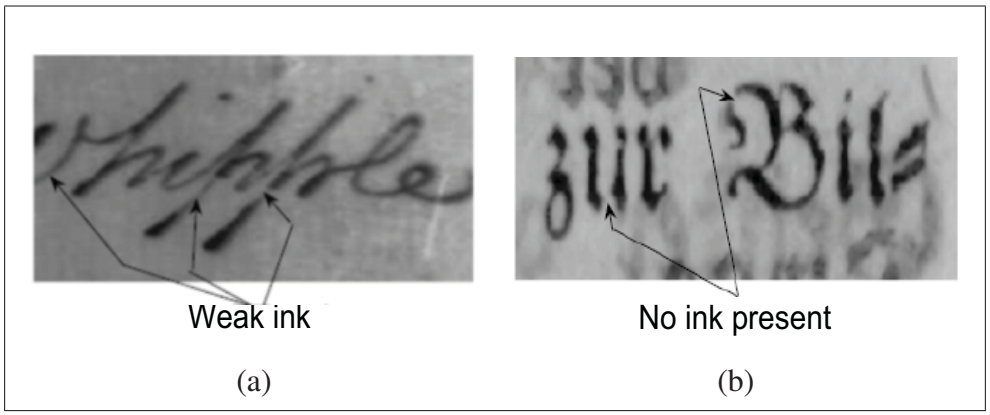


Figure 3.15 Printed and handwritten ink degradation, **(a)** handwritten document, **(b)** printed document. The arrows indicate the missing ink or areas with little ink

3.6.2.3 Comparison with the state of the art

As explained above, we have evaluated the proposed algorithm on the DIBCO'09 dataset (Gatos *et al.*, 2009a), which consists of 10 ancient document images (5 are handwritten and 5 are machine-printed). The images are of different sizes, 9 of them are gray, and one is in color. The proposed algorithm is compared against the four algorithms discussed previously in subsection 3.6.2. These four algorithms have been chosen because of the availability of their implementation code (the third and fourth algorithms) and the availability of their performance measures (the first and second ones) (Gatos *et al.*, 2009a). Table 3.2 provides the average of the *F-measure* values of various methods (Gatos *et al.*, 2009a). The proposed method outperforms the others, although the results are close. The proposed method also has a better performance in terms of subjective evaluation (see Figures 3.16). Also, because of its ability to preserve weak connections and recover important holes, the method is also more suitable for the skeletonization problem. Sauvola's algorithm gives a good result, but it also suffers from some noise in the output image, especially in the presence of strong interference (Figure 3.16(b)). It is worth noting that D. Rivest-Hénault *et al.'s* algorithm is also competitive. However, it suffers from interfering patterns and fails to separate some strokes.

In order to obtain a more specific analysis of the performance of the proposed method, its scores are compared with those of other methods in Table 3.3. It is worth noting that we have access to the performance of just three methods, because the results of other works are not available to us. From Table 3.3, we can conclude that the global thresholding method (for example, Otsu's method) works well in the case of uniform document images (images P03, H01) and also in

Table 3.2 Comparison of the performance of the proposed algorithm and the others against DIBCO'09. For Sauvola's method, the performances reported in (Farrahi Moghaddam and Cheriet, 2010b) have been used

Algorithms	Method N^o in DIBCO'9	F-measure
The proposed algorithm	–	**91.354**
Lu and Tan(Gatos *et al.*, 2009a)	26	91.24
Fabrizio and Marcotegui(Gatos *et al.*, 2009a)	14	90.06
Rivest-Hénault, Farrahi Moghaddam, and Cheriet(Gatos *et al.*, 2009a)	24	89.34
Sauvola's(Sauvola and Pietikainen, 2000)	–	87.27
Multi-scale grid-based Sauvola(Farrahi Moghaddam and Cheriet, 2010b)	–	89.26

the case of light bleed-through (image P02) (Farrahi Moghaddam and Cheriet, 2010b). The local methods, such as Sauvola's, work well in the case of non-uniform background (images H04, H05). In the case of bleed-through, for example image H02, these methods fail, while the proposed method with a good initialization achieves a good result for this type of document image. The multi-scale grid-based Sauvola method (Farrahi Moghaddam and Cheriet, 2010b) performs well compared to the Otsu and Sauvola methods in almost all cases. Only in the case of bleed-through in image H02 does this method fail to extract the true text. In contrast, the proposed method achieves a good performance in all cases.

3.6.3 Computational cost and complexity of the method

The proposed binarization algorithm, implemented in Matlab 7.0, takes 36 seconds on average for all images in the dataset on an AMD Athlon (TM) 64×2 Dual Core Processor 6000+, 3.01 GHz, 3.25GB of RAM with Windows XP. Figure 3.17 concludes that the time cost of the algorithm is linearly proportional to the linear size of the images. The complexity of the algorithm can be expressed as $O(\frac{n^2 \times m^2}{m^2})$=$O(n^2)$, where n is the size of the images and $m = 2d_c$ is the size of a grid patch.

3.7 Conclusions and future prospects

In this paper, we have introduced a new approach to the binarization of degraded ancient document images. The main idea is to spatially adapt a two-class (text and non-text) ML classifier to the pixels by a local estimation of the parameters of each class. The parameters of a class for

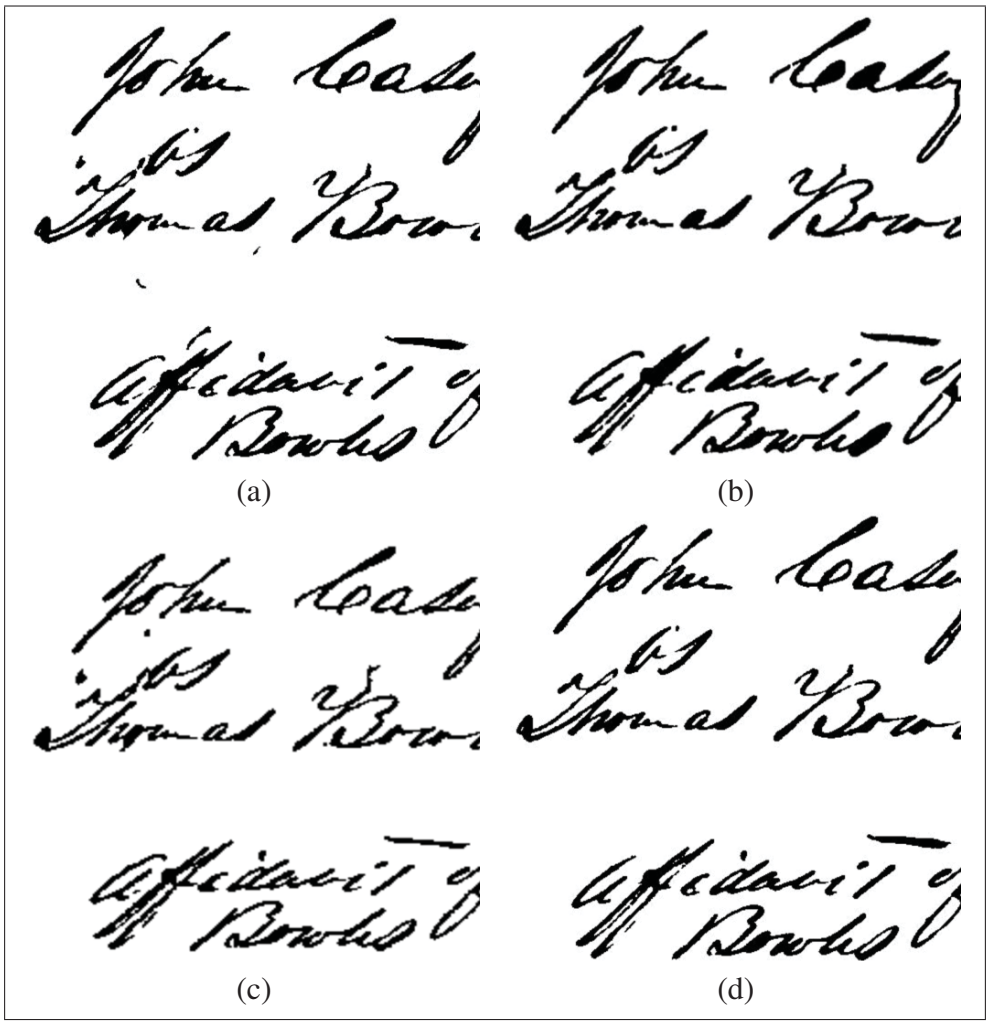


Figure 3.16 Visual comparison of the binarization algorithms on the image shown in 3.2(a). **(a)** Sauvola's binarization, **(b)** Lu and Tan's binarization, **(c)** Rivest-hénault, Farrahi and Cheriet's binarization, **(d)** the proposed binarization

each pixel are computed from the gray-level distribution of its neighbors. In order to eliminate the strongly interfering background while preserving the weak parts of low-intensity text, low intensity text pixels are considered in estimating the variance. This helps to improve the extraction of these weak parts. A simple morphological operator is applied on an initialization map, followed by a simple Otsu thresholding process to include the original low-intensity text pixels and separate them from the background pixels. The high-intensity and low-intensity pixels are then used to estimate the variance. In addition, the variance on each pixel is weighted based on its distance from the main text.

Table 3.3 Performance in terms of precision, recall, and *F-measure* of the proposed algorithm. For the Otsu (Otsu, 1979) and Sauvola (Sauvola and Pietikainen, 2000) methods, the performances reported in (Farrahi Moghaddam and Cheriet, 2010b) have been used

Image	Otsu	Sauvola	Multi-scale grid-based Sauvola	Our method
H01	90.84953	82.02617	91.40965	91.614
H02	86.14536	76.91948	77.62213	92.461
H03	84.11402	88.19474	86.55118	90.142
H04	40.55702	88.94607	85.67288	87.952
H05	28.03838	84.32496	83.13418	84.782
P01	89.80882	90.15921	91.80884	91.630
P02	96.54794	94.57987	95.92604	96.007
P03	96.77968	89.58904	96.99918	97.043
P04	82.71352	91.88228	94.06863	91.647
P05	88.36841	86.07259	89.42637	90.266
Average	78.39227	87.26944	89.26191	91.354

The main advantage of the proposed algorithm is that it preserves weak connections and provides smooth and continuous text strokes, thanks to its local and correlation-based nature. Furthermore, an under-binarization map is used as the initialization, as well as local ML classification help, on one hand to remove noise from the output images, and on the other hand to recover text with very low intensity strokes in order to preserve the topology of the text. This concept leads to an increase in the recall and precision of the output results, which is very important in subsequent processing. Another advantage of the proposed algorithm is that it is a soft decision method that is parameterless for thresholding. In addition, it is simple to implement and of low complexity. However, the proposed method fails to recover very small loops in characters, especially when they belong to a highly degraded background. We are working to generalize the model to cover this case of degradation as well. The proposed algorithm has been tested on the DIBCO'09 dataset. The results show that the proposed algorithm outperforms the other algorithms developed in the literature, in terms of *F-measure* and subjective evaluation. Also, it has been noted that there are different types of degradation depending in particular on whether the document is handwritten or printed. This information can be helpful in future work where we will investigate the integration of the Markov random field in order to smooth the edges and preserve the connectivity between the broken parts of the text using both the direction and intensity gradient of image pixels.

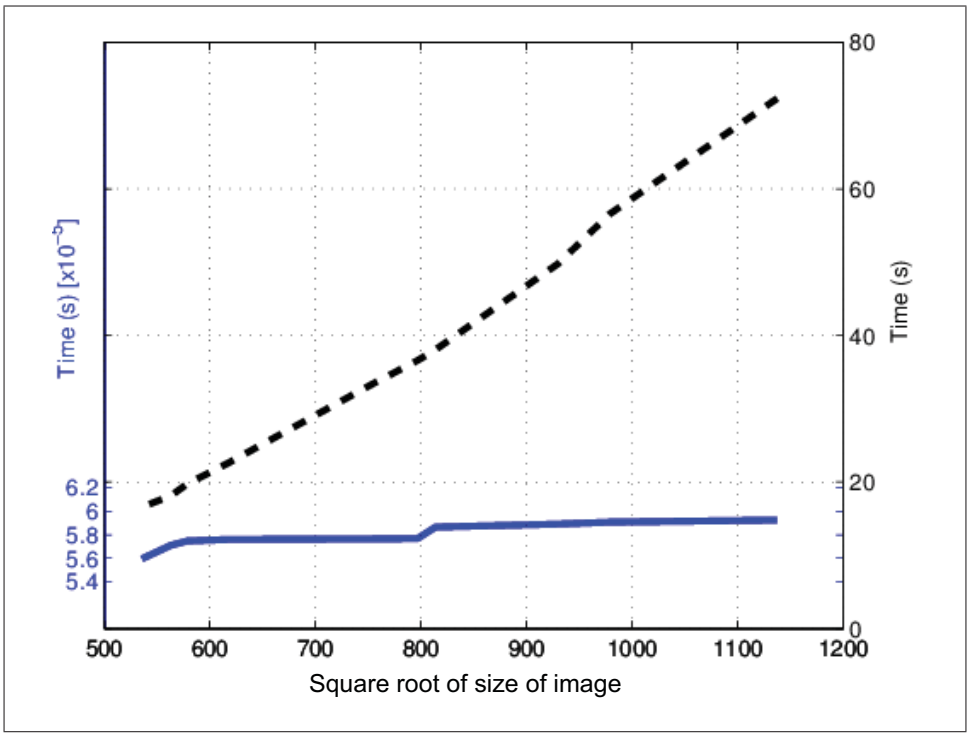


Figure 3.17 Time cost of the proposed algorithm. The continuous line indicates the processing time per pixel (the left axes), while the dashed line shows the image processing time

Acknowledgment

The authors would like to thank the NSERC of Canada for their financial support.

CHAPTER 4

ARTICLE II: DOCUMENT IMAGE RESTORATION USING MULTISPECTRAL IMAGING SYSTEM

Rachid Hedjam[1] and Mohamed Cheriet[1],

[1] Département de génie de la production automatisée, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Published in Elsevier Pattern Recognition journal,
Volume 46, Issue 8, February 2013, Pages 2297–2312

Abstract

Thousands of valuable historical documents stored on the shelves of national libraries throughout the world, are waiting to be scanned in order to facilitate access to the information they contain. The first major problem faced is degradation, which renders the visual quality of the document very poor, and in most cases, difficult to decipher. This work is part of our collaboration with the BAnQ (*Bibliothèque et Archive Nationales de Québec*), which aims to propose a new approach to provide the end user (historian, scholars, researchers, etc.) with an acceptable visualization of these images. To that end, we have adopted a multispectral imaging system capable of producing images in invisible lighting, such as infrared lights. In fact, in addition to visible (color) images, the additional information provided by the infrared spectrum as well as the physical properties of the ink (used on these historical documents) will be further incorporated into a mathematical model, transforming the degraded image into its new clean version suitable for visualization. Depending on the degree of degradation, the problem of cleaning them could be resolved by image enhancement and restoration, whereby the degradation could be isolated in the infrared spectrum, and then eliminated in the visible spectrum. The final color image is then reconstructed from the enhanced visible spectra (red, green and blue). The first experimental results are promising and our aim in collaboration with the BAnQ, is to give this documentary heritage to the public and build an intelligent engine for accessing the documents.

keyword

Historical document images restoration, Ferro_ gallic ink, Multispectral document Imaging, Multispectral image in-painting.

4.1 Introduction

Cultural heritage is considered an important source of national / international patrimony, and plays a fundamental role in economic and social development. It is an essential feature of peoples and communities and a reference to their culture and civilization. Protecting it means not only protecting the heritage itself, but also the culture behind this heritage. At present, the world's visual cultural heritage is essentially composed of ancient documents and old manuscripts. Usually, ancient manuscripts, which can constitute a significant portion of cultural heritage, are unique, with only single copies available. Unfortunately, these copies incur a serious risk of loss and extinction and, at the very least, suffer from many forms of deterioration and physical degradation phenomena thanks to a combination of factors: environmental conditions, tears, dust, dirt, rusty staples, poor and risky handling practices during storage, poor ways of sorting the drawers, effects of natural disasters or accidents, improper assembly, low quality paper, insects and rodents, darkness and non-conforming temperature levels, air pollution, humidity causing mold and discoloration of papers, etc. (Baird, 2000). Fig. 4.1 shows some samples of degraded document images.

Digital archiving of historical documents is growing in the areas of heritage studies and preservation (Antonacopoulos and Downton, 2007; Manso and Carvalho, 2009; Joosten, 2008). This task requires that archived images be enhanced, restored and stored in a logical manner in order to facilitate access and dissemination of valuable information. Indeed, the digital enhancement and restoration of a degraded historical document image can be seen as a transformation process aimed at restoring its original look. In addition, enhancement and restoration are required not only to improve the quality of documents, but also to improve the results of subsequent segmentation, recognition and word indexing operations.

4.1.1 Difficulty in analyzing degraded document images

A degraded historical document image, can be considered as being a combination of many superimposed layers of information (Moghaddam and Cheriet, 2009), including the background layer, the main text layer, and the degradation layers. The latter are assumed to have been added to the original document by a diffusion process (Moghaddam and Cheriet, 2009). In this work, we suggest that the different degradations be classified under two types: slight, and strong. Slight degradations, such as in the case of simple noise, small paper fluctuations, etc., can easily be treated and removed from the document image because of their weak intensity as compared to that of the main text. An example is shown in Fig. 4.5(a,b). Strong degradations are due to the presence of some objects whose color is close to that of the main text, and which

Figure 4.1 Some samples of degraded document images collected from the BAnQ

in many cases, appear darker. Stamps, annotations, ink bleed-through, tear, etc. are a few examples of these kinds of degradation. If the degradation overlaps the main text, it is considered strong (see Fig. 4.1(c,d,e)), if not, it is considered strong (see Fig. 4.1(a,b,f,g,h)). Degradations of the second type posed a significant problem in document image enhancement. They make it difficult to find an optimal thresholding strategy for the main text extraction process, and lead to unsatisfactory visualization results, because of the presence of interfering objects in the output images.

4.1.2 Objective of the paper

In this work, done in partnership with the BAnQ, we will be introducing the Multispectral imaging system (MSI), and we will be using its advantages to investigate a new way of enhancing the visual aspect of archived historical document images. The specific goal of this

study is to propose an automatic approach to enhance the degraded historical documents and to restore them, while maintaining their original appearance as much as possible. We intend to provide the end user with an acceptable view of the document image; a view in which only the original text is retained without any significant changes made to the texture of the background. We also intend to make this approach adaptive and to minimize the human intervention involved.

The remainder of this paper is broken down as follows: A brief introduction to the MSI, in the document analysis field, is given in section 4.2.2, while some related works are discussed in section 4.2. Section 4.3 describes data acquisition, and the proposed restoration model is detailed in section 4.4. The model parameter estimation and optimization is given in section 4.5. In section 4.6, experimental results are presented and, finally, section 4.7 provides our conclusions and prospects for future extensions.

4.2 Related work

4.2.1 Hyperspectral remote sensing image enhancement

Hyperspectral satellite images are usually corrupted by the loss of information due to many problems such as malfunctions of specific parts of the sensor, dramatic deviation from the initial characterization of spectro-radiometric response of the imaging system, etc. Some works (Mendez-Rial *et al.*, 2012; Shen and Zhang, 2009; Bouali and Ladjal) have been recently developed in order to enhance these kind of images by recovering the missing pixels, using an inpainting technique. The motivation behind these methods is the exploitation of the spatial and spectral correlations of information helping to interpolate neighboring data within the holes in spatial direction as well in spectral direction. To this end, in (Shen and Zhang, 2009), a Maximum a Posteriori (MAP)-based model has been proposed to restore the hyperspectral data with a likelihood, based on a linear image observation model, combined with a robust edge-preserving Huber–Markov model as image prior. Similarly, the authors of (Mendez-Rial *et al.*, 2012) are only interested in diffusing uncorrupted spectral signatures from the surroundings into the holes, preserving the extruded spatial structure characteristic of hypercubes and the spectral continuity of each pixel. Likewise, the paper (Bouali and Ladjal) describes a non local inpainting method into which a spectral similarity measure is incorporated, to restore missing data from a particular multispectral imagery.

4.2.2 Multispectral imaging in the area of document analysis

The problem of historical documents image processing has attracted the attention of many researchers over the last decade. In the context of this study, the existing methods proposed in the literature can be distinguished into two classes: those dealing with gray level document images (Lu *et al.*, 2010; Hedjam *et al.*, 2011a; Rivest-Hénault *et al.*, In Press; O'Gorman, 1994; Taxt *et al.*, 1989; Vonikakis *et al.*, 2011), and those that are based on multispectral imaging systems (Hedjam and Cheriet, 2011a,b; Klein *et al.*, 2008). Some of the former methods use thresholding techniques to extract the foreground text in the first stage, to interpolate the background in the second stage, and then to merge the extracted text with the interpolated background in order to recover the original aspect of the document; the other methods of the same class, for their part, require the registration of both sides of the document, in order to exploit the information from both sides, which can help to isolate the main text from the degradation. Unfortunately, when dealing with few spectral representation bands (i.e., gray level or RGB color), the text is not clearly contrasted from the degradation, as shown in Fig. 4.1, which makes it difficult to find an optimal thresholding strategy for separating text from the degradation. In the second class, when several spectral bands (> 3) are available, more information will be available about the document image content. In such a multispectral space, objects that are similar in color, may have different spectral reflectances.

In recent years, multispectral imaging (MSI) technology has successfully been used in artwork investigations (Ribes *et al.*, 2008) and in the transcription of historical manuscripts (Klein *et al.*, 2008; Joo Kim *et al.*, 2011) by detecting and measuring the portion of the light reflected by different zones of manuscripts to be investigated. MSI is a very important tool which enables conservators and art historians to obtain valuable information on artworks and ancient documents without causing any physical damages to the materials. This technique is widely known as a non-invasive method of investigation thanks to its simultaneous use of ultraviolet, infrared and visible radiations and makes it possible to reveal newly painted areas or overwritten text, to distinguish and recognize the chemical material composing the ink, to enhance the visibility of latent patterns in a palimpsest, and to detect signs of degradation in historical documents, etc. It also improves the readability of old documents and provides information on alterations and degradations that have just been added over the years as the materials were conserved. It can also help to extract information from cultural heritage patterns which cannot be extracted using conventional color photography (Klein *et al.*, 2008). The technique is based on the principle that the materials, making up different objects in the documents, react, absorb and emit electromagnetic radiations in particular manner, depending on the chemical composition of the

material. Fig. 4.2 shows how infrared and ultraviolet light can be used to enhance the contrast and reveal hidden text in degraded documents.

More recently, the work proposed in (Joo Kim *et al.*, 2011) aims to visualize the hyperspectral document image in its natural appearance, while simultaneously emphasizing various contents in the invisible range in order to enhance the readability of the data. To that end, the authors proposed a new scheme to maximize the amount of contextual details in the document image while maintaining the appearance of the document in the visible range. The above-mentioned work has been designed to enhance the contrast in the visible channels but not to remove strong degradations from them.

Our work aims at removing the unwanted artifacts from the visible color channels while preserving as much as possible the original visual aspect of the document image. Based on the hypothesis that the artifacts can be classified into two classes of degradation (SLIGHT and STRONG) as described in section 4.1.1, we propose to first perform a correction of slight degradations as a pre-process step, and then to rely on an unsupervised model which automatically decides if it is necessary to perform or not an inpainting technique to remove the (eventual) remaining strong degradations. Our work can be clearly distinguished from the inpainting-based methods used in remote sensing domain by the nature of the degradation and the particular type of multispectral document image where the presence of degradation is totally different. The detail will be given in section 4.4.

4.3 Multispectral Image Acquisition

The setup of most MSI systems, used in documents analysis, is shown schematically in Figure A 1-1 (a) in Appendix I: MS imaging system, set-up and acquisition, where the document under investigation is placed on an object support facing the camera. Two tunable lamps are used to illuminate the document, and are usually positioned on the upper-left and upper-right sides of the camera, under a 45^o angle, and close to the document. To ensure that this tool is non-destructive (Klein *et al.*, 2008), a low-intensity wavelength-tunable light source is generally used, with the temperature and relative humidity in the room kept constantly ambient, and the document not exposed for a long time. The MSI system uses a set of chromatic filters (which are considered as band-pass filters) to produce different high resolution spectral images corresponding to the different wavelengths in ultra-violet (UV), visible (VIS) and infrared (IR) wavelength. The MSI is entirely enclosed in a light-proof cabinet in order to avoid any stray light from external sources, which could disturb the measurements.

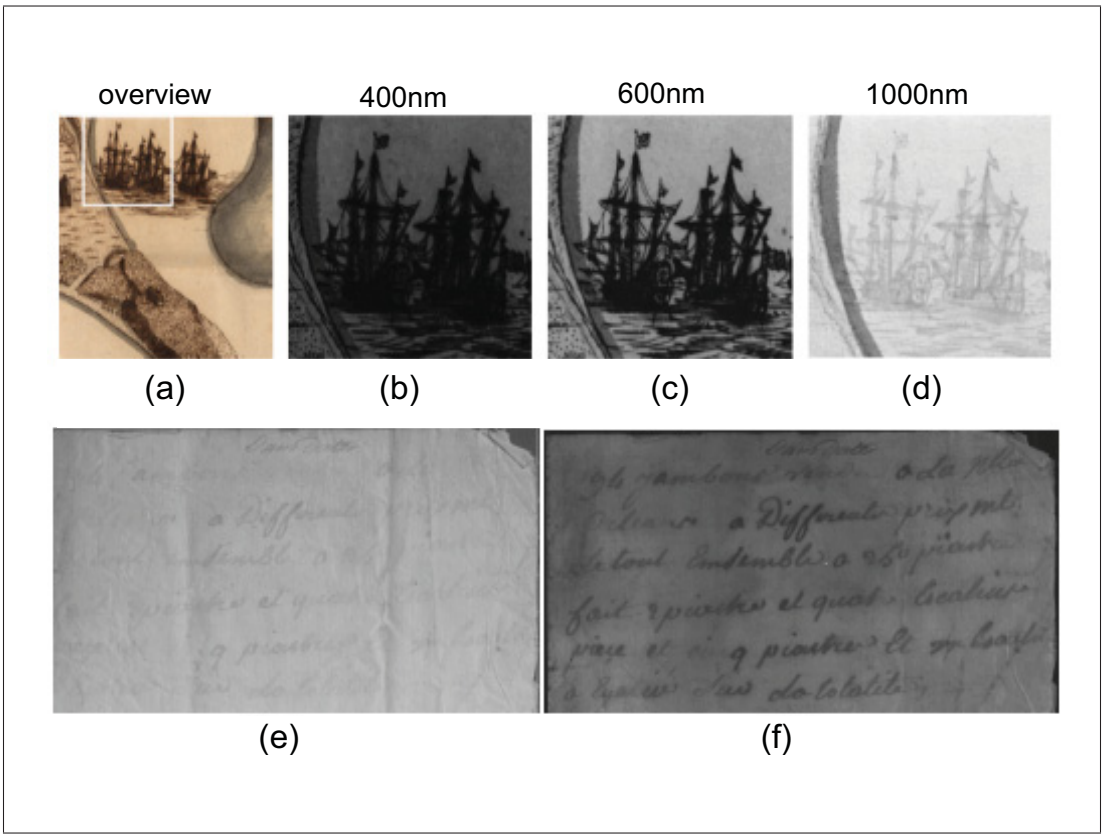
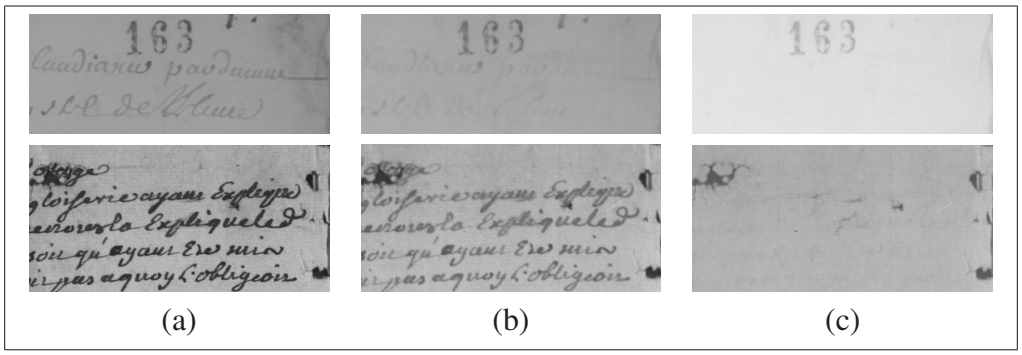


Figure 4.2 Document image enhancement using multispectral imaging system. (a) Color band. Due to the degradation of the ink, the details of the objects become invisible to the naked eye in the 400nm and 600nm band (b,c); but they can be distinguished in the infrared band at 1000nm (d). The images (a-d) are from (Klein *et al.*, 2008). The UV fluorescence (f) is used to reveal the hidden texts (e)

Our MS imaging system, is composed of a CCD camera, an objective in front of the camera, and a wheel containing 8 chromatic filters. The camera's sensor is a KAF 6303E (Kodak); it has high quantum efficiency of $1,100\ nm$, is front-side illuminated, and has a resolution of 3072×2048 pixels of 9×9 microns. The filter wheel is controlled by a software program and acts as a pass-band filter to generate 8 chromatic images (or channels): one in the UltraViolet (UV, 400 nm) wavelength, three in the VISible (500 nm, 600 nm and 700 nm) wavelength, and four in the infra-Red (IR) wavelength (800 nm, 900 nm, 1000 nm and 1100 nm). The set of collected channels constitute the so-called MS information cube (See Figure I-7 in Appendix I: MS imaging system, set-up and acquisition), which contains one spectral reflectance (or spectral signature) for each pixel (Klein *et al.*, 2008). It should be noted that the UV channel is not used in this work because it contains no useful information in terms of our objective.



(a) (b) (c)

Figure 4.3 Two multispectral images: (a) 495 nm (Green), (b) 800 nm (IR), (c) 1100 nm (IR)

4.3.1 Characteristics of the MS degraded document image

We collected and digitized over 110 MSI images from the BAnQ, based on historical documents, written between the 17th and 20th centuries. Most 20th century documents were printed by machines, and present less degradation, while the earliest ones, are handwritten with iron-gall (ferrogallic) based ink, and are highly degraded. The iron-gall based ink shows a particular characteristic when it is exposed to different light spectra (Havermans *et al.*, 2003a). Typically, this ink can be seen in the color (visible) channels, but disappears gradually in the IR ones. However, most degradations (annotations, paper fold and tears) continue to be visible even in the IR channels (see Figure 4.3). Another characteristic considered is the signal to noise ratio, which is high in the color channels and weak in the IR channels (i.e, this characteristic is related to our acquisition multispectral system). The question now is how to use these physical proprieties of the materials to separate the main text from the degradations.

4.4 Proposed restoration model

Let us define the notations used throughout this paper. We consider a multispectral (multichannel) image as a $N \times M \times B$ data cube where B is the number of spectral bands or channels. Each channel is assumed to be a $N \times M$ 2-D array or an image slice. We will use the term *spectral signature* to denote the B-dimensional vector related to each pixel s at coordinate $s = [i, j]^\top \in \mathbb{Z}^2$. More precisely, the spectral signature of a pixel s will be denoted as:

$$u(s) = \left[u_1(s), u_2(s), ..., u_B(s)\right]^\top \tag{4.1}$$

where $u_n(s)$, is the value (also called reflectance) of the pixel s at the channel corresponding to the $n-$th considered wavelength ($n \in [1, \ldots, B]$). In our application, $B = 7$, and u_1, u_2 and u_3 respectively denote the blue (B), green (G) and red (R) channels, and u_4, u_5, u_6 and u_7, denote the IR channels. To simplify the notations, the visible color channels are denoted by $u_{\text{COL}\in\{\text{R,G,B}\}}$ and the IR channels by $u_{\text{IR-p}}$ ($p \in [4, \ldots, 7]$).

Let us now consider a multispectral image u^0 (related, in our application, to a degraded historical document) in which each pixel s belongs to one of the three different classes of information taking its value in the set $\mathcal{L} = \{l_t, l_b, l_d\}$, namely; TEXT or INK ($l_t$), BACKGROUND ($l_b$) and (strong) DEGRADATION (l_d). Let $l(s) = l_k$, the class label associated to each pixel s. The restoration of the original color image u_{COL} from u_0 (with the image domain Ω) is done separately for each color channel $u_{\text{COL}\in\{\text{R,G,B}\}}$, by replacing the pixels belonging to the DEGRADATION class (l_d) in a way as convincing as possible, and more precisely regarding to some prior knowledge about the image and while preserving the information belonging to the class l_t or l_b. This problem can be resolved by the following unsupervised segmentation-restoration model:

$$\hat{u}_{\text{COL}} = \arg\min_{u_{\text{COL}}} \left\{ \int_\Omega \underbrace{\sum_{l_k \in \{l_t, l_b\}} \delta\big(l(s), l_k\big)}_{\lambda'(s)} \Big(u_{\text{COL}} - \Upsilon\big(u^0_{\text{COL}}\big)\Big)^2 ds + \alpha \int_\Omega |\nabla u_{\text{COL}}| ds \right\} \tag{4.2}$$

$\delta(a, b)$ is the Kronecker delta function (Kaplan, 2002) which is equal to 1 if $a = b$ and 0 otherwise. The second term of this equation is our regularization term and refers to the total variation of the restored image. α is a value controlling the contribution of these two terms and acts as a smoothness parameter. $\Upsilon(.)$ is the correction term of slight degradations (see Section 4.5.2) which is already performed on the original data according to our hypothesis which postulates that by default the historical document image contains at least slight degradations.

If the term $\lambda'(s) = \sum_{l_k \in \{l_t, l_b\}} \delta\big(l(s), l_k\big)$ is considered as a two-class segmentation map or a binary mask[1], then the segmentation-restoration model, expressed by Eq. (4.2), is modeled as an inpainting (image interpolation or completion) and denoising problem (Chan and Shen,

[1] i.e., a mask in which the value 0 represents the pixels belonging to the DEGRADATION class and the value 1 represents the useful information or the non-degraded pixels (belonging either to the TEXT or the BACKGROUND classes).

2002a) from the viewpoint of variational principles and exploiting the total variation criterion as an effective image prior model. Inpainting is a restoration process of reconstructing lost or deteriorated parts of images (herein belonging to the degradation class) by propagating the image information (color and texture) from the known region (i.e., the useful information belonging to the TEXT or BACKGROUND classes in our application) into the missing or degraded regions. It is worth noting that this restoration process is thus globally performed at the region level (for the inpainting process) and at the pixel level (for the denoising process) as opposed to the correction step of slight degradations (Υ) which is semi-local and performed at the pixel level.

It is also worth noting that our model is adaptive; in presence of strong degradation, a mask is automatically generated (see subsection 4.5.3.1 for more details) and the inpainting step will take place. Otherwise, if no strong degradation is detected, the mask will not be generated and the inpainting will not be performed. In a multispectral representation space, the different objects or classes of the document react differently from one another according to their reflectance properties, due to the chemical composition of their respective materials. Thus, ink (text) and degradations areas, for example, may appear differently in the infrared channels, although they will have a similar intensity in the visible color channels. This physical property, will be exploited in section 4.5.3.1 in order to automatically detect these strong degradation areas if they do exist.

4.5 Parameter estimation and model optimization

4.5.1 Unsupervised IR band selection

As already mentioned, the iron-gall ink, which constitutes in fact the text portion/class of the degraded historical document to be restored and degradations areas, for example, may appear differently in the infrared channels, although they will have a similar intensity in the visible color channels. More precisely, this purple-black or brown-black ink, made from iron salts and tannic acids from vegetable sources, can be seen in the color (visible) channels, but disappears gradually in the IR ones. However, most degradation (annotations, paper fold and tear) continue to be visible even in the IR channels (see Figure 4.3). This specific property, inherent to the chemical composition of this ink, can be efficiently used both to estimate accurately the strong degradation areas, which will be exploited in the restoration process at the region level (see Section 4.5.3.1), but also to estimate the most distinctive IR band used in the correction step of slight degradations (see Section 4.5.2).

In light of that and, as a preliminary step, our restoration model seeks to estimate the most distinctive and informative IR band, $\hat{u}_{\text{IR}}$, eliminating degradation in the color channels without introducing undesirable artifacts in the enhanced image. This being said, this most distinctive IR band is thus one that: (1) contains the minimum possible main text, and only shows the degradation that are visible in the color channel and also (2) includes less artefact and noise.

• The first condition can be met by computing the pixel-to-pixel distance between each IR channel and the green channel, u_{G} (GREEN), which usually have the highest signal to noise ratio according to our acquisition system. To render the estimation of this distance robust and less sensitive to the noise, we calculate the distance in a binary space of the image as follows

$$D_{\text{bin}}(u_{\text{G}}, u_{\text{IR-p}}) = \int_{\Omega} \|\Lambda(u_{\text{G}})(s) - \Lambda(u_{\text{IR-p}})(s)\| \, ds \tag{4.3}$$

where Ω represents the domain of the image, $\|a\|$ is the absolute value of a and, Λ is any model which transforms a gray image into its binary version. In this work, we adopted the grid-based Sauvola (GBS) method which is simple and fast algorithm introduced in (Farrahi Moghaddam and Cheriet, 2010b). GBS's method is an adaptive and threshold-based binarization that aims to transform the input image from the gray level space (i.e., $[0, \ldots, 1]$) to the binary space (i.e., $[0, 1]$) by assigning a threshold value $T(s)$ to each pixel s on the document image based on the statistics of the gray values around that pixel.

$$T(s) = \mu(s)\left[1 + k\left(\frac{\sigma(s)}{R} - 1\right)\right] \tag{4.4}$$

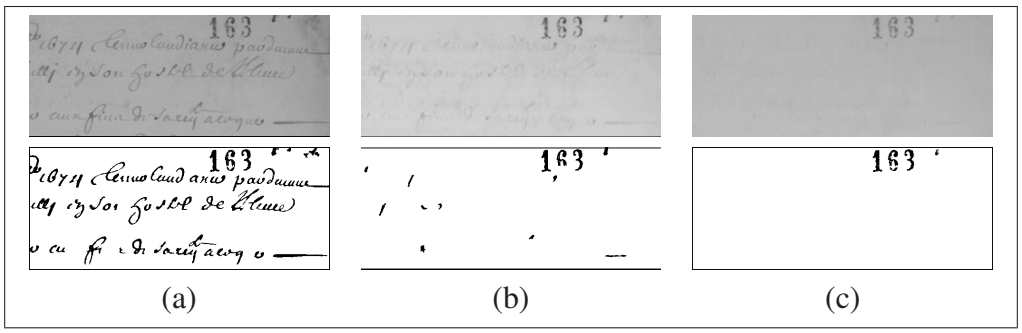
where $\mu(s)$ and $\sigma(s)$ are the mean and the standard deviation of the processed image values calculated on a neighborhood of $(2w + 1) \times (2w + 1)$ size around s. R is the maximum value of the standard deviation ($R = 0.5$ for a gray-scale document image), and k is a parameter that takes real positive values. Thus the binarization transform of a given gray image I is defined by:

$$\Lambda(I)(s) = \begin{cases} 0, & \text{if } I(s) < T(s) \\ 1, & \text{else} \end{cases} \tag{4.5}$$

where 0 represents a foreground pixel and 1 represents a background pixel. According to the first condition, the most distinctive IR band is the one that maximizes the distance D_{bin}, or that minimizes:

$$D'_{\text{bin}}(u_{\text{G}}, u_{\text{IR-p}}) = \int_{\Omega} \left(1 - \|\Lambda(u_{\text{G}})(s) - \Lambda(u_{\text{IR-p}})(s)\|\right) ds \tag{4.6}$$

From Figure 4.4, it is clear that, unlike the first IR ($u_{\text{IR-1}}$) band which shows some part of text, the third IR band ($u_{\text{IR-3}}$) does not contain any text, and thus, will be selected as the most distinctive and informative IR band.



(a) (b) (c)

Figure 4.4 Binarization of the green and various IR channels. First row : (a) u_3 (green) band; (b) u_5 ($u_{\text{IR-1}}$) band; (c) u_7 ($u_{\text{IR-3}}$) band. Second row: the corresponding binary versions

• The second condition can be formulated by the total variation measure over the image, which is small for images with less noise and artifact. The total variation can be computed as follows:

$$TV(u_{\text{IR-p}}) = \int_{\Omega} |\nabla u_{\text{IR-p}}(s)|\, ds \tag{4.7}$$

where $|\nabla u(s)|$ is the gradient operator of u at site or pixel s. Therefore, the model we propose for selecting the most distinctive p-th IR band ($u_{\text{IR-p}}$) combines the two last equations and can be written in the following form:

$$\begin{aligned} \hat{p} &= \arg\min_{p\,(\in[4,\ldots,7])} \Big\{ D'_{bin}(u_{\text{G}}, u_{\text{IR-p}}) + \beta TV(u_{\text{IR-p}}) \Big\} \\ &= \arg\min_{p\,(\in[4,\ldots,7])} \int_{\Omega} \Big[\Big(1 - \|\Lambda(u_{\text{G}})(s) - \Lambda(u_{\text{IR-p}})(s)\| \Big) \\ &\quad + \quad \beta\, |\nabla u_{\text{IR-p}}(s)| \Big]\, ds \end{aligned} \tag{4.8}$$

where β is a value controlling the contribution of these two terms.

4.5.2 Semi-local correction of slight degradations

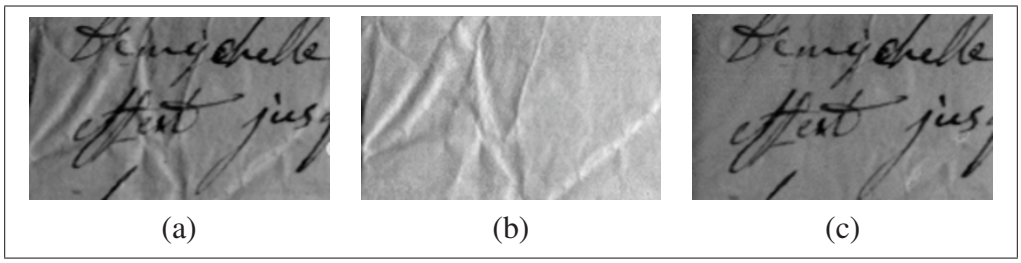
Some slight degradations in the color channel, u_{COL}, such as weak show-through effect, and simple paper fluctuation, can be corrected by the proposed correction model:

$$\Upsilon\big(u_{\text{COL}}(s)\big) = \mu_{\text{IR-}\hat{p}} \frac{u_{\text{COL}}(s)}{\hat{u}_{\text{IR-}\hat{p}}(s)} \tag{4.9}$$

where the correction is performed for every color channel and, $\mu_{\text{IR-}\hat{p}}$ is the global mean gray value of the most distinctive and informative IR band $\hat{u}_{\text{IR-}\hat{p}}$, defined in Eq. (4.8).
The factor $\mu_{\text{IR-}\hat{p}}/\hat{u}_{\text{IR-}\hat{p}}$ aims to reduce and mitigate the luminance of the bright pixels, if $\mu_{\text{IR-}\hat{p}} < \hat{u}_{\text{IR-}\hat{p}}(s)$ (i.e, $\frac{\mu_{\text{IR-}\hat{p}}}{\hat{u}_{\text{IR-}\hat{p}}(s)} < 1$), and enhances the dark pixels if $\mu_{\text{IR-}\hat{p}} > \hat{u}_{\text{IR-}\hat{p}}(s)$. This can lead to a decrease in the intensity variation. Figure 4.6 shows an example of 1D signal correction. The blue signal represents a horizontal pixel line extracted from the GREEN color channel, while the red signal represents the same pixel line extracted from the IR channel (i.e; $\hat{u}_{\text{IR-}\hat{p}}$), used to correct the blue signal. From this Figure, we can see that the model expressed by Eq. (4.9) is able to smooth the blue signal in locations with considerable intensity variation, as shown by the magenta signal between 20^{th} and 70^{th} x-coordinates pixels. Unfortunately, the original image perception is not always preserved with this model, because of the factor $\mu_{\text{IR-}\hat{p}}/\hat{u}_{\text{IR-}\hat{p}}(s)$. This can be explained by the change in the magnitude of the corrected signal.
To avoid this problem caused by the model 4.9, we propose to improve it with a semi-local adaptive correction strategy as follows:

$$\begin{aligned}\Upsilon\big(u_{\text{COL}}(s)\big) = {} & \chi(\sigma_{\text{IR-}\hat{p}}(s) \geq \tau) \cdot \mu_{\text{IR-}\hat{p}}(s) \cdot \frac{u_{\text{COL}}(s)}{\hat{u}_{\text{IR-}\hat{p}}(s)} \\ & + \chi(\sigma_{\text{IR-}\hat{p}}(s) < \tau) \cdot u_{\text{COL}}(s)\end{aligned} \tag{4.10}$$

where χ is the indicator function and $\sigma_{\text{IR-}\hat{p}}(s)$ is a local standard deviation calculated on a neighborhood of $(2w+1) \times (2w+1)$ size around a given pixel s. The model expressed by Eq. (4.10), ensures a correction in the areas where the $\hat{u}_{\text{IR-}\hat{p}}(s)$ channel presents variations ($\sigma_{\text{IR-}\hat{p}}$) higher than a given threshold τ. In other words, in areas where the intensity variation is very small (i.e; under τ), the first term on the right side of Eq. (4.10) is null and, thus $u_{\text{COL}}(s)$ will be equal to $u_{\text{COL}}(s)$ (i.e., $u_{\text{COL}}(s) = u_{\text{COL}}(s)$), which means that, the enhanced pixel preserves its value, as shown in Figure 4.5 (c). The threshold τ can be computed empirically or learned from the content of the degraded document image (as it is explained in Appendix II: Automatic finding of the threshold τ).



(a) (b) (c)

Figure 4.5 Correction of slight degradations. (a) slight-degraded document image, (b) most distinctive and informative IR band of (a), (c) corrected document image using the model expressed by Eq. (4.10)

By taking a look at Figure 4.6, we can clearly see that the blue signal is well corrected in locations where the intensity variance is high, as it is shown by the dark signal between 20^{th} and 70^{th} x-coordinates, and it is well preserved in locations where the intensity variance is small. Figure 4.7, shows also, the difference between the semi-local and adaptive semi-local corrections. It is clear that the histogram distribution of the data corrected by the correction model expressed by Eq. (4.10), is more similar to the histogram distribution of the original data compared to that of the data corrected by the model expressed by Eq. (4.9). Quantitatively speaking, the Bhattacharyya-based distance (Cha and Srihari, 2002) between the histograms of Figures 4.7(a) and 4.7(b) is equal to 0.10 while the distance between the histograms of Figures 4.7(a) and 4.7(c) is equal to 0.14. This is due to the fact that the adaptive semi-local correction model preserves the original intensity better than the semi-local correction model.

4.5.3 Correction of strong degradations

In the case of strong degradations, our problem is stated as an inpainting problem (Chan and Shen, 2002a), which follows the framework of the conventional interpolation model with the total variation based image prior model. We recall that this restoration step first relies on the selection of the most distinctive p-th IR band, $u_{\text{IR-}\hat{p}}$ (see Eq. (4.8)), i.e., the band that contains the least amount of text and for which most degradation (annotations, paper fold and tears) continue to be visible. This most distinctive IR band ($u_{\text{IR-}\hat{p}}$) will allow us to estimate the characteristic function λ' (see Eq. (4.2)), which concretely represents a two-class segmentation map or a binary mask in which the value $''0''$ represents the pixels belonging to the DEGRADATION class and the value $''1''$ represents the useful information or the non-degraded pixels (belonging either to the TEXT or the BACKGROUND classes).

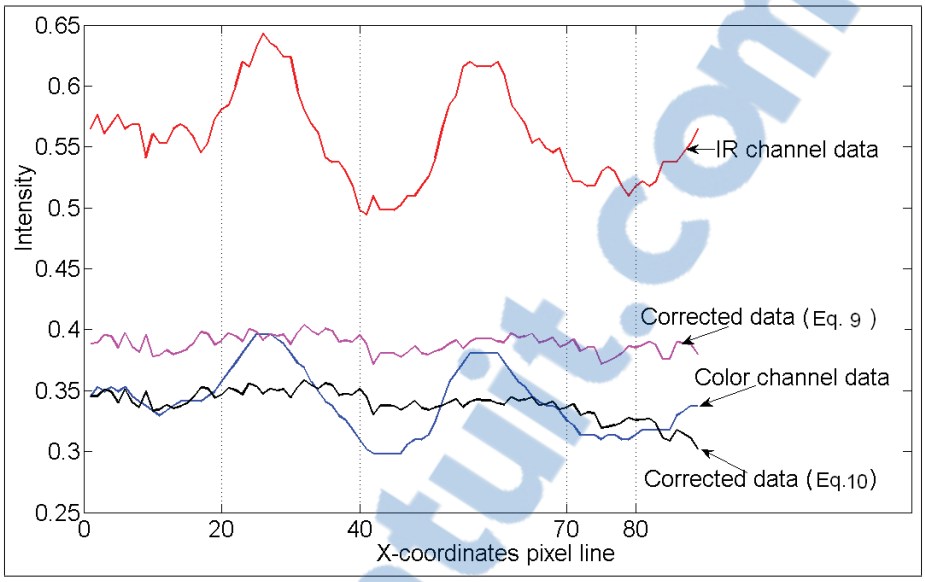


Figure 4.6 An example of 1D-data correction

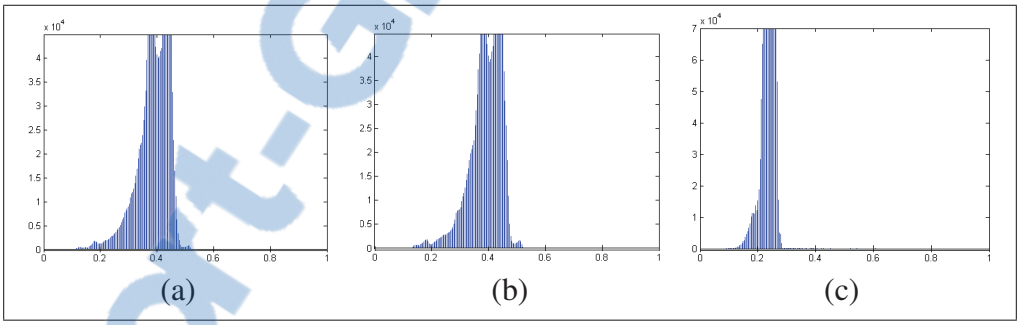


Figure 4.7 Histogram based comparison between slight degradation correction models; (a) histogram distribution of the original data; (b) histogram of the corrected data by the model 4.10; (c) histogram of the corrected data by the model 4.9

4.5.3.1 Estimation of the binary mask

Using the MSI system, the mask, λ is then extracted as shown in Fig. 4.8, by performing the transform Λ (see Eq. (4.5)) on the IR channel $\hat{u}_{\text{IR-}\hat{p}}$ (see Eq. 4.8):

$$\Lambda(\hat{u}_{\text{IR-}\hat{p}})(s) = \lambda(s) \quad \in \{0, 1\} \tag{4.11}$$

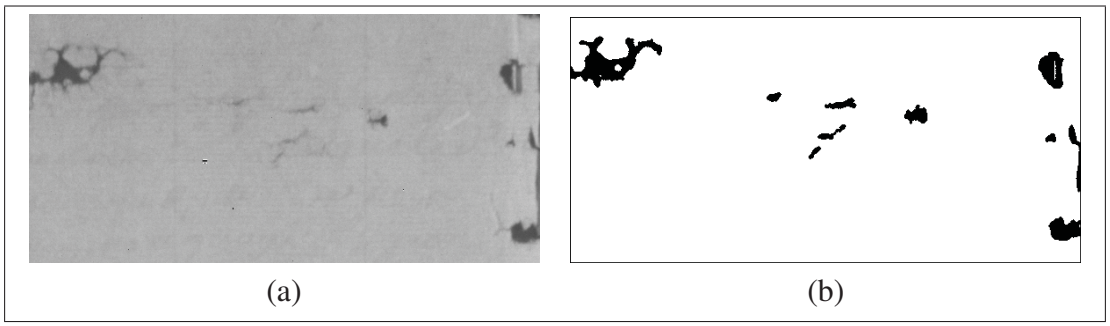


Figure 4.8 λ mask estimation. (a) selected IR channel, $\hat{u}_{\text{IR-}\hat{p}}$; (b) binary mask λ

By its random nature, some degraded areas may overlap with some parts of the text, as shown in Fig. 4.9(a) (in which the pixels overlapping the text and the degradation area are shown in yellow in Fig. 4.9(e)). Unfortunately these degraded areas, which may cover and/or overlap the text, will be considered as belonging to the degradation class in the mask λ (see Fig. 4.9(c)), and, therefore, will be eliminated with the useful text by the inpainting based restoration process (as shown in Fig. 4.9(d)). To avoid this undesirable situation, we propose to exclude from the mask λ, the pixels belonging to a mixture of these two classes or materials (DEGRADATION and TEXT) and thus to consider, as belonging to the DEGRADATION class, only the set of pixels belonging to the *pure* DEGRADATION class.

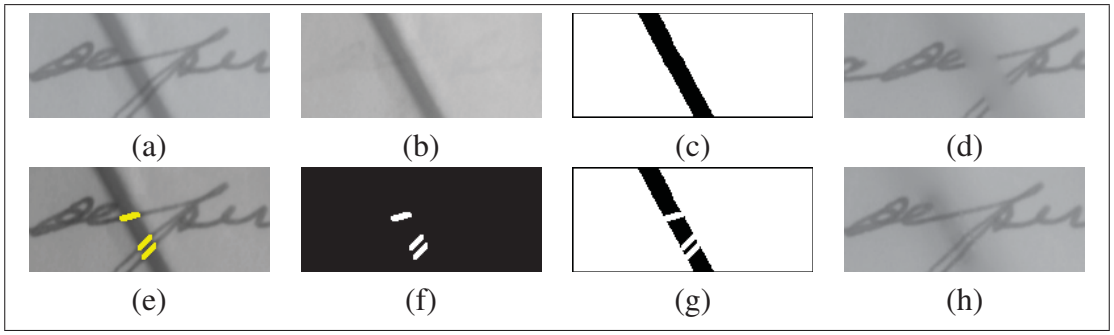


Figure 4.9 Mask estimation procedure from text/degradation overlap: (a) degraded image, (b), $\hat{u}_{\text{IR-}\hat{p}}$ channel; (c) binary mask λ, generated from (b); (d) inpainted image using λ; (e) overlapped text pixels are shown in yellow; (f) more similar overlapped pixels are shown in white; (g) mask λ', resulted from pixel-to-pixel addition of (c) and (f); (h) inpainted image using the mask (g)

To this end, we assume that each pixel of the MS document is either a *pure* pixel which contain a single material or class or either a *mixed* pixel which may contain a mixture (i.e., a linear

combination of) different materials or classes (e.g., for some pixels; a mixture of TEXT and DEGRADATION classes), each class possessing a different pure spectral signature. The extraction problem of these pure spectral signatures (also called end-members) along with their proportion for each pixel is called multispectral image unmixing (Nascimento and Dias, 2005).
In our application, the end-members of different classes (i.e; TEXT or INK, BACKGROUND and DEGRADATION) are extracted using the vertex component analysis (VCA) algorithm proposed in (Nascimento and Dias, 2005). This method assumes that some pure pixels are necessarily present in the MS document, which is a quite valid assumption in our case. Finally, the VCA algorithm works on the fact that the affine transformation of a simplex is another simplex which helps to find hidden (folded) verticies of the simplex. To do so, firstly, the image pixels labeled 0 by the mask λ, are compared to the (previously estimated) end-member of the class INK in the input MS document image. In our application, this pure spectral signature belonging to the class INK of the image to be processed, can be recognized by matching the spectral signatures of the end-members (extracted from the current image) to the reference spectral signature of an ink sample collected from a variety of other images documents. In Fig. 4.10, we can notice that this comparison can be easily done, since the spectrum of the pure INK class and the reference mean spectrum of the ink, collected from a variety of other images documents, are similar enough to be easily recognized, while being, however, slightly different. Indeed this end-member extraction and un-mixing estimation step (made on each MS document) allows us to adaptively estimate the pure spectral signature of the INK class, which varies a bit for each MS document, and to make robust the subsequent segmentation between the TEXT and the DEGRADATION classes.

After the unmixing step, the segmentation between the TEXT/INK and the DEGRADATION classes is then simply done by the following procedure; If the spectral distances between the vector of a given pixel belonging to the degradation areas (located by λ) and the pure spectral signature of the TEXT/INK class estimated on the current image is lower than a fixed threshold ε, we exclude this pixel from the mask λ. In our application, we use the *spectral information divergence* (SID) which is a commonly used mathematical distance for comparing spectral vectors (Chang, 2000). Finally, the set of pixels which are not excluded from λ gives the new mask λ' used in Eq. (4.2).

4.5.3.2 TV denoising and inpainting problem

The objective of the inpainting algorithm is to fill in a missing area (as shown in Fig. 4.11) based upon the image information available outside. In other words, the inpainting algorithm

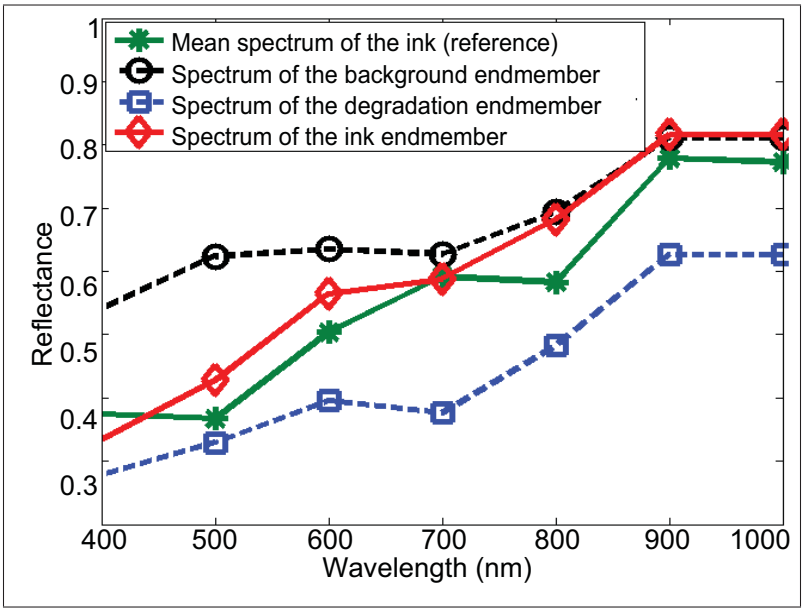


Figure 4.10 Comparison of the end-member spectra of different classes to the reference mean spectrum of the class ink

tends to replace a hole with the original data regarding to some image prior knowledge. Once the mask λ' is estimated, the TV denoising and inpainting based restoration process is modeled by the following minimization of the function E:

$$\hat{u}_{\text{COL}} = \underset{u_{\text{COL}}}{\text{argmin}} \underbrace{\left\{ \int_{\Omega} \lambda'(s) \left(u_{\text{COL}} - \Upsilon\left(u^0_{\text{COL}}\right) \right)^2 + \alpha |\nabla u_{\text{COL}}| \, ds \right\}}_{E} \tag{4.12}$$

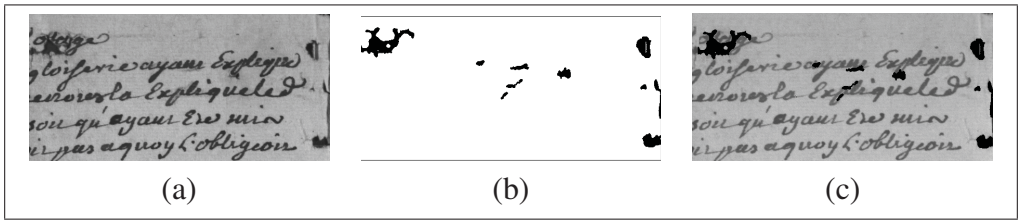


Figure 4.11 Regions to be inpainted localisation. (a) color channel image; (b) estimated binary mask λ; (c) projection of λ on the image of Fig.(a). The regions to be inpainted are shown in black color

To minimize the energy in Eq. (4.12), we solve the Euler-Lagrange equation. To simplify the notations in the next formulas, we replace u_{COL} by u and $\Upsilon\left(u^0_{\text{COL}}\right)$ by u_0, which gives:

$$F(u, u_x, u_y) = \lambda'\big(u - u_0\big)^2 + \alpha\,|\nabla u| \tag{4.13}$$

With u_x and u_y respectively the row and column derivative of u, the Euler-Lagrange equation for E is:

$$\frac{\partial E}{\partial u} = \frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\frac{\partial F}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial F}{\partial u_y}$$

$$\begin{aligned}
\text{where} \qquad & \frac{\partial F}{\partial u} = 2\lambda'(u - u_0); \\
& \frac{\partial F}{\partial u_x} = u_x/\sqrt{u_x^2 + u_y^2}; \quad \frac{\partial F}{\partial u_y} = u_y/\sqrt{u_x^2 + u_y^2} \\
& \frac{\partial}{\partial x}\frac{\partial F}{\partial u_x} = \frac{u_{xx}u_y^2 - u_x u_y u_{xy}}{(u_x^2 + u_y^2)^{3/2}}; \quad \frac{\partial}{\partial y}\frac{\partial F}{\partial u_y} = \frac{u_{yy}u_x^2 - u_x u_y u_{xy}}{(u_x^2 + u_y^2)^{3/2}}
\end{aligned}$$

and the final Euler-Lagrange equation is

$$\frac{\partial E}{\partial u} = 2\lambda'(u - u_0) - \alpha\frac{u_{xx}u_y^2 - 2u_x u_y u_{xy} + u_{yy}u_x^2}{(u_x^2 + u_y^2)^{3/2}} \tag{4.14}$$

So, to minimize the energy in Eq. (4.14), we solve the Euler-Lagrange differential equation $\partial E/\partial u = 0$ with a gradient-descent method

$$\frac{\partial u}{\partial t} = \alpha\frac{u_{xx}u_y^2 - 2u_x u_y u_{xy} + u_{yy}u_x^2}{(u_x^2 + u_y^2)^{3/2}} - 2\lambda'(u - u_0)$$

and with a discretization scheme using finite differences, we obtain the iterative update formula:

$$\begin{aligned}
& u_{ij}^{[n+1]} = \\
& u_{ij}^{[n]} + \triangle t\Bigg(\alpha\left[\frac{u_{xx}^n u_y^{[n]2} - 2u_x^{[n]}u_y^{[n]}u_{xy}^{[n]} + u_{yy}^{[n]}u_x^{[n]2}}{\left(u_x^{[n]2} + u_y^{[n]2}\right)^{3/2}}\right]_{ij} \\
& - 2\lambda'(u_{ij}^n - u_{0ij}^n)\Bigg)
\end{aligned} \tag{4.15}$$

where $\triangle t$ denotes the time step between temporal samples of u (any small constant makes the iteration stable). The reference (Niyobuhungiro, 2010) gives more details on the discretization scheme. A diagram of the proposed algorithm is given in Fig. 4.12.

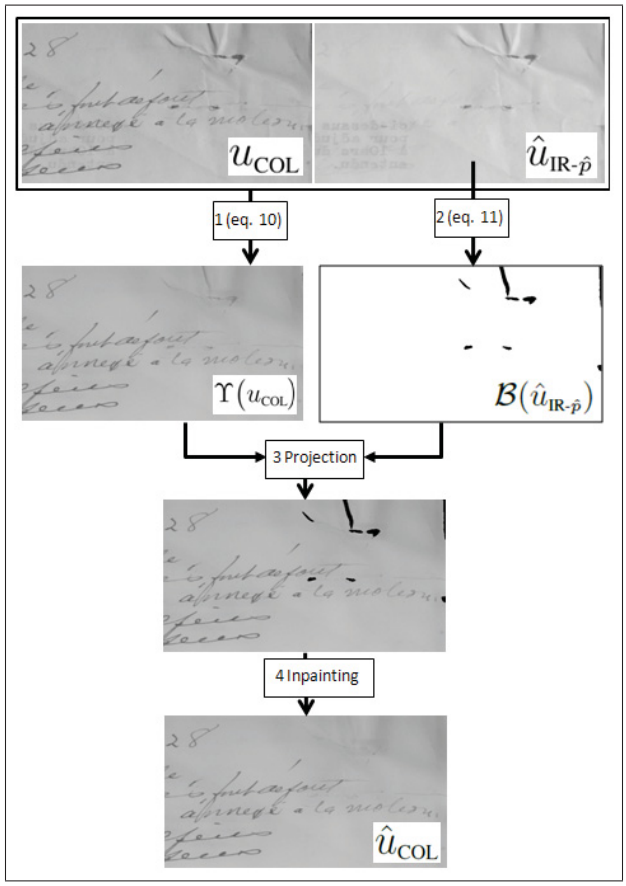


Figure 4.12 The overall restoration procedure; step 1: slight degradation correction; step 2: binary mask generation; step 3: projecting the binary mask onto the corrected color channel; step 4: inpainting processes

4.6 Experimental result

Before starting the discussion on the evaluation per say, we would like to clarify the following: our work is unique in the sense that there is no other similar works with which the evaluation can be performed. Recently, only one work (Joo Kim *et al.*, 2011) was published aiming to enhance of specific historical document images which are not for free use. The main difference between this work and ours, is that we do restoration. i.e, we look for the interfering patterns which are still visible in the IR bands and then clean them from the color ones while the for-

mer work aims to improve the contrast in the final color document image. For these reasons, we found it difficult to compare our work. Nevertheless, we thought to follow a reasonable methodology of evaluation in subjective and objective ways as we will explain in details in the subsequent paragraphs. In addition, we have the intention to make our data available for free in a way to facilitate the comparison of the future works. We have selected a subset of images from the whole dataset that we collected from the BAnQ and we have also started to create some ground truth images[1]. All the multispectral historical document images used for the experiment phase show different types of degradation such as ink fading, folding, tears, etc. They are of different sizes, 8 bits (256 gray-scales) in png format. Each of these multispectral images contains 8 bands, one in the ultraviolet (UV) wavelength (340nm), three bands in the visible wavelength (BLUE (500nm), GREEN (600nm), BLUE (700nm)) and four in the infrared (IR) wavelengths (800nm, 900nm, 1000nm, 1100nm). For each multispectral document image, a ground-truth is generated.

The subjective evaluation can be directly performed by visually checking the restored output against its degraded input. As for the objective evaluation, we proposed the following approach: since the image restoration problem may be used as a preprocessing for some subsequent tasks such as text extraction, word spotting, document binarization, etc, we thought to evaluate the performance of the proposed algorithm using nine algorithms, A1 (Lu *et al* (Lu *et al.*, 2010). algorithm), A2 (grid-based Sauvola's algorithm, GBS (Farrahi Moghaddam and Cheriet, 2010b)), A3 (Gatos *et al.* algorithm (Gatos *et al.*, 2004)), A4 (Wolf *et al.* algorithm (Wolf *et al.*, 2002)), A5 (AdOtsu algorithm (Farrahi Moghaddam and Cheriet, 2012)), A6 (R-Hénault *et al.* algorithm (Rivest-Hénault *et al.*, 2011)), A7 (Shading-based algorithm (Toennies, 2005, P.202)), A8 (Abutaleb algorithm (Abutaleb, 1989a)) and A9 O'Gorman algorithm(O'Gorman, 1994). The binarization task is performed on both degraded document image and the corresponding restored image and the F-measure is computed for each output. The restoration quality (Gain) can be evaluated according to the difference between the two obtained F-measure values. It is obvious that if the restoration is successfully realized, the binarization on the restored document image should be higher than that on the degraded one.

Some samples of the collected images are shown in Figures 4.17 and 4.18. The original and the enhanced images are labeled [Xa] and [Xb], respectively. [X] indicates the number of the image, [a] indicates the original color image, and [b] indicates the restored image. For the objective evaluation, 9 ground-truth images are generated for the document images [4a, 6a, 8a, 10a, 11a, 12a, 13a, 14a, 15a].

[1] http://www.synchromedia.ca/databases/HISTODOC1

4.6.1 Parameters setup

In all the experiments, we have considered the following internal parameters for our restoration model. The grid size w of the models (4.4) and (4.10) is set to 25. The parameter k of the model (4.4) is set to 0.025. The threshold ε used in the generation phase of the mask λ', varies in [0:015; 0:02] for the majority of the images. The parameter β in Eq. 4.8 is set to 1 and finally the parameter α in Eq. 4.12 is set to 10000.

4.6.2 Subjective and objective evaluation

Globally speaking, we can see from the results that the objective of this work is largely achieved. The different degradations are mostly removed, and the visual aspect of the original images is preserved. The color of the original document images is conserved in the majority of the enhanced images, except for the inpainted areas, which become a bit smoother because of the interpolation or propagation of information within them. Visually, the annotations are completely removed, and the corresponding inpainted areas are well merged with the background which gives the impression that nothing existed within the degradation areas, as shown in Fig. 4.17 ([2b], [3b], [15b]). Fluctuations in the document are generally less apparent than other strong degradations. An example is shown on the left and right areas of Fig. 4.17 ([[5a]) and 4.18 ([9a]). Correcting this problem may be feasible thanks to the models expressed by Equations (4.9) and (4.10), which balance the intensity variation and make the affected areas more homogeneous. Local correction has an effect only in areas with an apparent variance. Indeed, unaffected areas preserve their appearance in the enhanced images. For the areas where the paper is folded, as shown in Fig. 4.17 ([4a], [5a], [6a]) and 4.18 ([8a], [10a], [13a], [15a]), the proposed inpainting model is relatively effective in removing this effect, and greatly enhances the visual quality of the enhanced document image (see Fig. 4.17 ([4b], [5b], [6b]) and 4.18 ([8b], [15b])). The proposed algorithm is also able to remove stamp and show-through effects from the document image (see Fig. 4.17[7b]). The show-through effect, by its nature, is less dark than the foreground text, and is visible both in the color and IR channels. The reduction of this effect can be achieved by the correction model of slight degradation expressed by Eq. (4.10), while the stamps whose color is close to that of the text, need to be removed by inpainting. Other problems which can be resolved by the proposed algorithm are the tear and holes in paper substrate (Fig. 4.17 ([1a,b], [3a,b], [4a,b])).

Objectively speaking, under the assumption that a better restoration result leads to a better result of further processing, such as binarization, we seek here to compare the binary output of the degraded document image and that of its restored version against the corresponding

available ground-truth image based on F-measure index. Nine algorithms of document image binarization are used for the evaluation, A1 (Lu *et al* (Lu *et al.*, 2010). algorithm), A2 (grid-based Sauvola's algorithm, GBS (Farrahi Moghaddam and Cheriet, 2010b)), A3 (Gatos *et al.* algorithm (Gatos *et al.*, 2004)), A4 (Wolf *et al.* algorithm (Wolf *et al.*, 2002)), A5 (AdOtsu algorithm (Farrahi Moghaddam and Cheriet, 2012)), A6 (R-Hénault *et al.* algorithm (Rivest-Hénault *et al.*, 2011)), A7 (Shading-based algorithm (Toennies, 2005, P.202)), A8 (Abutaleb algorithm (Abutaleb, 1989a)) and A9 (O'Gorman algorithm (O'Gorman, 1994)). The parameters of each algorithm are optimally set up. The F-measure results are shown in Table 4.1 and plotted graphically in Figure 4.13. We note here that the binarization is performed on the gray images (i.e, $\frac{\text{RED+GREEN+BLUE}}{3}$) of the original and restored data. It is clear that in average the restoration processing leads to obtain a considerable gain ($\triangle_{\text{Ai,i=1..8}}$) as shown in Table 4.1 in the final binarization results which validate our assumption that, the restoration of degraded document image can facilitate the task of many subsequent task such as information extraction, line segmentation, binarization etc. Figure 4.13 shows graphically the performance of different binarization algorithms on both original and restored data (We note here that the comparison is not necessarily the interest of our work). The algorithms show competitive results on the ensemble of the document images. Figure 4.14 shows an example of the results of these binarization algorithms. It is clear that the conventional binarization algorithm dealing directly with gray level document images find a huge difficulty to separate the main text from other interfering patterns having similar intensity levels (see Fig. 4.14, First row) which happens frequently with historical document images. The use of MS imaging system is of great interest in document image processing because of the additional information which can be offered allowing the separation between the main text and other interfering patterns (see Fig. 4.14 second row).

In all cases, the quality of the enhancement process depends mainly on the size of the generated mask and the appearance of inpainted areas which itself depends on the size of the corresponding degraded areas. With a small size mask, as in the case of annotations or thin tears, interpolated information in inpainted areas is similar to that in the original surrounding areas. In fact, the human perception will not be too bothered by these negligible transitions. The images in Fig. 4.17 ([1b], [2b], [3b],[4b]) and 4.18 ([10b], [12b], [14b], [15b]) represent a good example of this. Some magnified regions are shown in Fig. 4.15.

With a large size mask, on the other hand, as in case of holes, tears, or large stains, we have large inpainted areas. Since the TV inpainting model used for this work is based on interpolation or spread information that is not degraded in degraded areas, if these latter are too broad, then the

Table 4.1 F-measure and the gain ($\triangle_{Ai,i=1..9}$) values of the algorithms: A1 (Lu *et al.*, 2010), A2 (Farrahi Moghaddam and Cheriet, 2010b), A3 (Gatos *et al.*, 2004), A4 (Wolf *et al.*, 2002), A5 (Farrahi Moghaddam and Cheriet, 2012), A6 (Rivest-Hénault *et al.*, 2011), A7 (Toennies, 2005, P.202), A8 (Abutaleb, 1989b) and A9 (O'Gorman, 1994). The means (**Mean**) and standard deviations (**STD**) are calculated for more insight on the performance of each algorithm

	Image	a4	a6	a8	a10	a11	a12	a13	a14	a15	**Mean**	**STD**
Before restoration	A1	82.91	81.68	81.03	61.72	86.58	74.58	87.02	71.73	86.99	79.36	8.53
	A2	84.99	58.43	75.02	55.27	68.15	73.22	86.85	69.87	81.72	72.61	11.04
	A3	80.25	86.5	79	66.88	65.24	71.83	58.04	78.06	78	73.76	8.94
	A4	78.66	86.54	81.92	59.85	67.8	67.42	56.1	74.39	79.85	72.5	10.36
	A5	85.07	87.97	84.72	76.58	78.26	82.92	58.68	72.72	80.17	78.56	8.83
	A6	84.76	84.03	80.06	76.24	68.05	79.17	58.27	78.58	77.64	76.31	8.32
	A7	87.8	87	83.52	72.58	68.06	73.2	57.93	71.67	77.59	75.48	9.65
	A8	84.46	84.82	79	11.66	66.11	15.89	29.28	67.4	67.05	56.18	29.18
	A9	76.33	35.84	78.67	4.53	66.24	40.18	28.62	70.69	64.45	51.73	25.52
After restoration	A1	84.66	85.47	85.13	71.88	88.89	84.56	88.07	77.93	92.73	84.37	6.16
	A2	90.21	76.86	76.32	74.05	85.15	87.39	89.96	79.5	89.79	83.25	6.56
	A3	83.66	87.1	81.77	80.36	79.25	79.07	77.37	80.8	82.76	81.35	2.9
	A4	84.35	86.74	87.94	63.01	84.19	79.64	74.39	81.12	82.79	80.42	7.68
	A5	87.63	88.61	92.35	85.94	87.16	87.3	75.46	84.56	85.95	86.11	4.56
	A6	88.76	86.56	92.28	84.82	85.17	84.47	75.72	83.56	85.31	85.18	4.44
	A7	88.26	87.24	85.01	83.58	78.21	80.14	70.15	73.67	80.68	80.77	6.06
	A8	89.4	86.41	81.77	34.26	82	48.94	38.62	71.42	70.75	67.06	21.09
	A9	84.44	81.63	84.49	36.35	81.42	48.86	30.42	74.38	69.08	65.68	21.45
Gain	$\triangle_{A1}$	1.75	3.79	4.1	10.16	2.31	9.98	1.05	6.2	5.74	5.01	
	$\triangle_{A2}$	5.22	18.43	1.3	18.78	17	14.17	3.11	9.63	8.07	10.63	
	$\triangle_{A3}$	3.42	0.6	2.77	13.48	14.01	7.24	19.32	2.74	4.75	7.59	
	$\triangle_{A4}$	5.7	0.2	6.02	3.16	16.4	12.22	18.29	6.73	2.94	7.96	
	$\triangle_{A5}$	2.56	0.64	7.63	9.36	8.91	4.38	16.78	11.84	5.78	7.54	
	$\triangle_{A6}$	3.99	2.54	12.22	8.58	17.12	5.3	17.44	4.98	7.67	8.87	
	$\triangle_{A7}$	0.42	0.24	1.5	11	10.14	6.94	12.22	2	3.09	5.29	
	$\triangle_{A8}$	4.94	1.59	2.77	22.6	15.89	33.05	9.34	4.02	3.7	10.88	
	$\triangle_{A9}$	8.11	45.79	5.82	31.8	15.19	8.68	1.84	3.69	4.64	13.95	

inpainted (interpolated) areas will be smoothed and, consequently, the enhanced image loses its texture in them.

A second factor influencing the quality of the enhanced image is the precision of the binarization transform used to extract the mask. The binarization transform is assumed to be robust and accurate in extracting the whole domain occupied by the degraded areas. If any degraded pixel is not located by the binarization transform, it remains visible in the enhanced image. Fig. 4.16 shows a magnified region extracted from one of our document images, as well as the

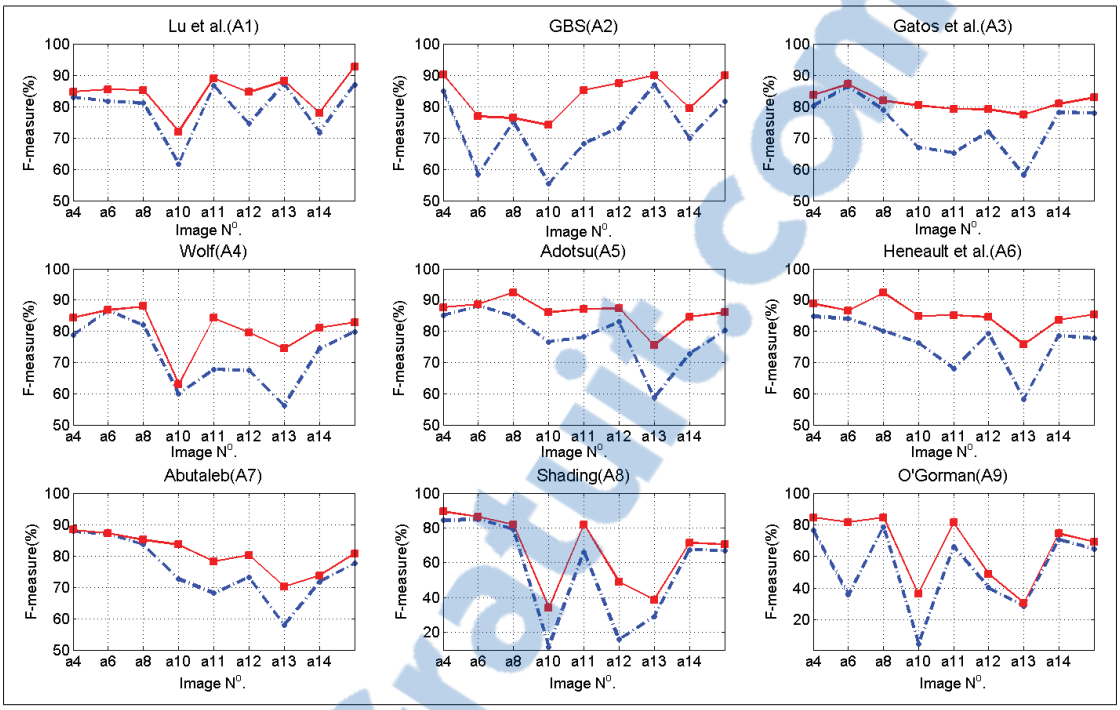


Figure 4.13 Graphs showing the impact of the restoration step on the result of different binarization algorithms. The continuous red curves correspond to the results after restoration while the dotted blue curves correspond to the results of the binarization before restoration

corresponding inpainting results. It is clear that the inpainting result shown in Fig. 4.16 (d) is more suitable than that shown in Figure 4.16 (c) due to the accuracy of the mask segmentation.

Another point worth discussing pertains to the position of the degraded areas in relation to the text. If there is no collision between the degraded areas and the main text, the interpolation process only takes into account the information belonging to the background and propagate it within these degraded areas. This leads to the formation of homogeneous areas with the background.

The proposed restoration algorithm, implemented in Matlab 7.0, takes 71s on average for an image of 600×600 pixels, (ie. 32s are allocated to the loading of the data and the selection of the optimal IR band. 31s are allocated to the flattening step and approximately, 8s are allocated to the inpainting step which needs a maximum of 100 iterations) on an AMD Athlon (TM)64×2 Dual Core Processor 6000+,3.01 GHz,3.25GB of RAM with Windows XP.

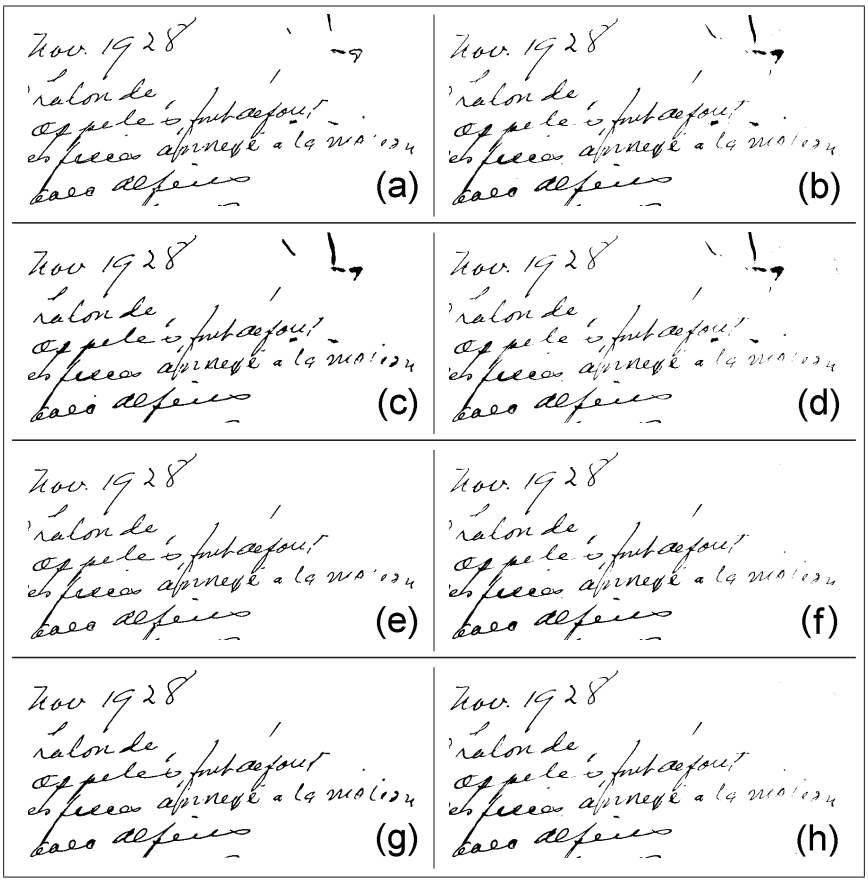


Figure 4.14 Effect of the restoration step on the binarization process (the image shown in Fig. 4.17 (4a) is chosen as an example). (a-d), binarization outputs of the algorithms A1, A3, A6 and A9 before restoration. (e-h), binarization outputs of the same algorithms after restoration

4.7 Conclusion

We have developed a new algorithm to restore and enhance the visual quality of degraded historical document images collected from the BAnQ. When few spectral bands are available (i.e., with a simple gray level or color space), we have first noticed that it is very difficult to find an efficient segmentation strategy for correctly estimating the different (classes or) objects (i.e., TEXT/INK, BACKGROUND and various DEGRADATIONS) of the historical documents, which will be then useful in our subsequent unsupervised restoration model. In light of that, we have proposed a new method dealing with multi channel images produced by our multispectral imaging system. The latter, in addition to the information provided by color channels, is able to provide additional information from infrared channels. In a multispectral representation space, the different objects or classes of the document react differently from one another according

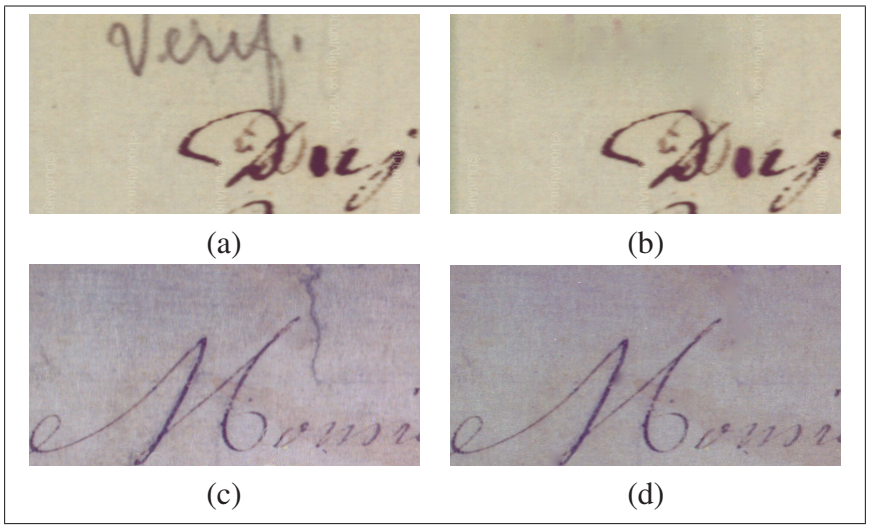


Figure 4.15 Example of thin size degradations and the corresponding inpainting result; (a,c) degraded images; (b,d) inpainted images

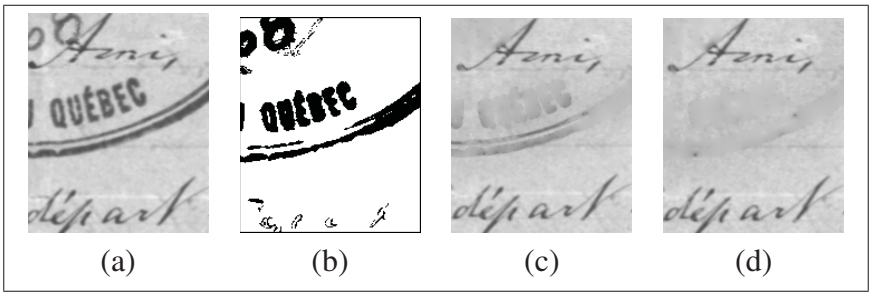


Figure 4.16 Influence of the binarization transform on the inpainting process: (a) original image with main text and degradations (stamp and annotations); (b) example in which the mask is not well extracted; (c) inpainted image using the mask shown in (b); (d) inpainted image when the mask is well extracted (not shown here)

to their reflectance properties, due to the chemical composition of their respective materials. Thus, ink (text) and black spots, for example, may appear differently in the infrared channels, although they will have a similar intensity in the visible color channels. This physical property, characterizing most historical documents from the $16^{\text{th}} - 19^{\text{th}}$ centuries in the BAnQ is due to the fact that they are mostly handwritten with iron-gall based ink, which is visible in the color channels, and gradually disappears in the infrared ones. Unlike text, degradations are visible in both the color and infrared channels.

Based on this key characteristic, we have proposed a new restoration model in which the information in the infrared channels is used to enhance the text in the visible color channel. Our

algorithm typically looks for the optimal IR channel in which text is completely absent, and then extracts the degradation areas by a binarization transform, to generate a binary mask used to locate the areas which should then be inpainted. Slight degradations, which do not need to be inpainted (because of their low variation), are first eliminated using a preliminary unsupervised semi-local correction. By this way, if a document image suffers only from slight degradations, only a slight correction step will take place.
The major problem arises when the degradation collides with text. This problem can be resolved by isolating the text pixels from overlapped area using a classical end-members extraction technique. The degraded pixels are then compared to the end-member belonging to the INK class in order to separate the pixels of the text. These pixels are then excluded from the previous mask, and the refined mask generated is then used in the final inpainting step.
The proposed algorithm is not complex, easy to implement and adaptive. Furthermore, in various cases, it is able to remove degradation while approximately preserving the original view of the document, except in inpainted areas, where the interpolated information represents a small smoothness especially when the degraded areas are thick. Otherwise, the quality of the enhanced document images is largely sufficient, according to the feedback of our collaborator.

In the future, we will be collecting a large set of multispectral images from a wide range of historical documents to test the effectiveness of the proposed algorithm and make them available for other researchers.

Acknowledgments

The authors would like to thank the NSERC of Canada and FQRNT of Quebec for their financial support. We would like to extend our thanks to the BAnQ for providing valuable ancient documents. The authors thank also the Professor Max Mignotte from university of Montreal for his thorough and constructive comments and suggestions. The authors would like to thank Dr. R. Farrahi Moghaddam and Dr. D. Rivest-Hénault from Synchromedia laboratory for providing the AdOtsu and the level set program codes.

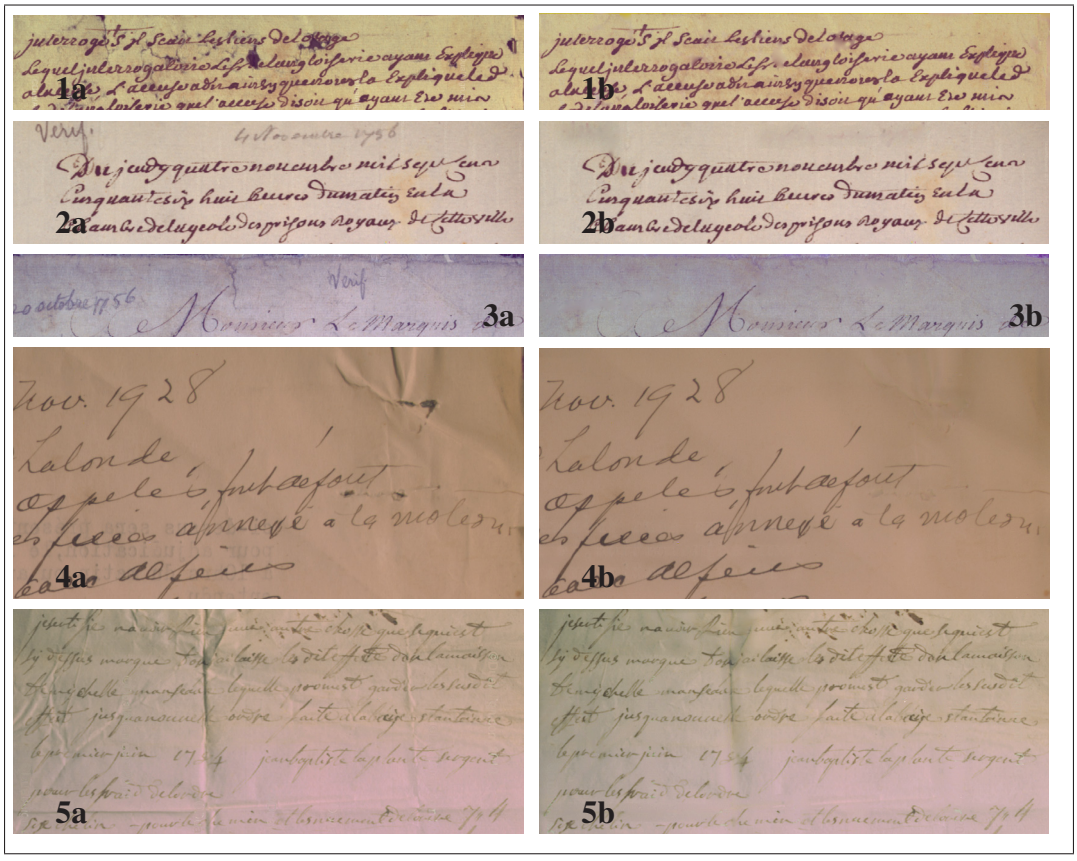


Figure 4.17 Result of the proposed algorithm. $[Xa]$, original image, $[Xb]$, enhanced image. $[X]$ is the number of the image

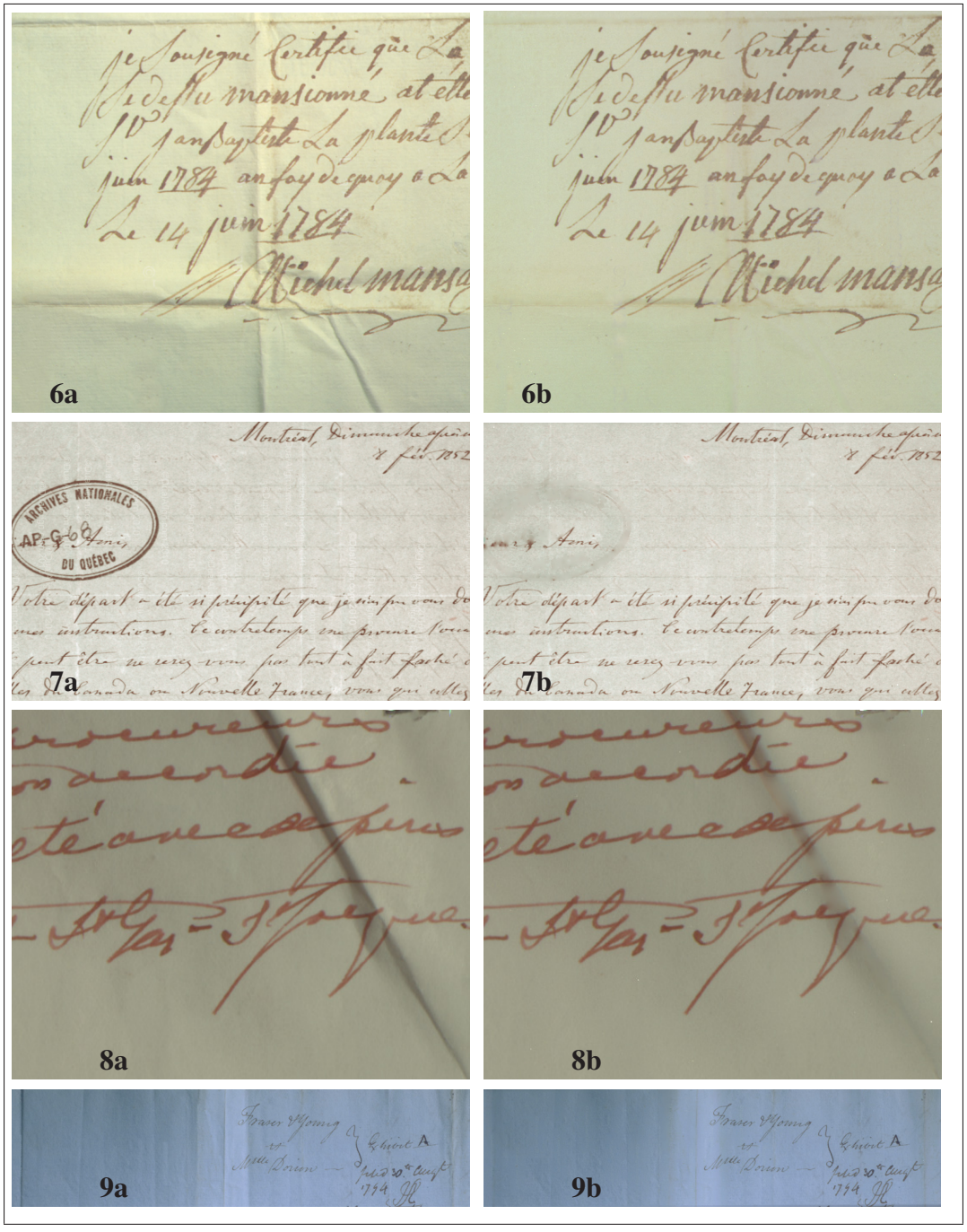


Figure 4.18 Result of the proposed algorithm. $[Xa]$, original image, $[Xb]$, enhanced image. $[X]$ is the number of the image

Figure 4.19 Result of the proposed algorithm. $[Xa]$, original image, $[Xb]$, enhanced image. $[X]$ is the number of the image

CHAPTER 5

ARTICLE III: REFERENCE DATA ESTIMATION

Rachid Hedjam[1] and Mohamed Cheriet[1],

[1] Département de génie de la production automatisée, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Submitted to IEEE Pattern Analysis and Machine intelligence, Trans. on (TPAMI); 2013.

Abstract

The manual generation of reference data in restricted representation spaces, such as gray-scale images, is highly subject to mislabeling and judgment errors. There are two factors that make the process very difficult: a lack of information to distinguish among the data, and the presence of degradation in the data, and so there is always the potential for producing inaccurate outputs. This paper addresses the issue of reference data estimation from the references of multiple experts. A method is proposed to accurately estimate reference data from multispectral (MS) images, which provide additional information in a multidimensional representation space. Data fidelity and a priori information about the agreement of the experts are incorporated in a simple Bayesian model to estimate new a posteriori labels. The method is evaluated based on the assumption that mislabeling in training set leads to a decrease in interclass variability, and consequently a decrease in the performance of the subsequent classification tasks. A real case study is used to validate this hypothesis, which shows promising results.

Keywords

Reference data estimation. Ground-truth. Document image analysis. Document image binarization. Historical document images. Multispectral document imaging.

5.1 Introduction

The performance of an algorithm is usually validated by comparing its output with reference data generated manually or semiautomatically by a human (i.e. ground-truth), which are believed to accurately reflect the true benchmark information to be processed. These data are also used, among other things, to compare algorithms to track their progress over time towards a level of human performance. They should indicate the existence of a degree of coverage at the

location of each sample. It would be of great interest to researchers to obtain accurate reference data with the most representative information about the targeted samples, which would facilitate comparison of the results of their algorithms. Reference data are used in natural image segmentation (Martin *et al.*, 2001), in computational color science (Ciurea and Funt, 2003), in remote sensing (Ghoggali and Melgani, 2009), and in document analysis (Sezgin and Sankur, 2004; P. Stathis and Papamarkos, 2008; Gatos *et al.*, 2009a; Pratikakis *et al.*, 2010, 2011; Smith, 2010). Whatever method is used to generate these data, manual or semiautomatic, as in document image binarization, medical image segmentation, and remote sensing image classification (Richards and Jia, 1999), the generation process is still very subjective. Even if the reference data are accurately generated, they can be viewed in dramatically different ways by different individuals, at different times, and under different conditions. Reference data generation is based on many factors, such as the purpose of the intended application, the cognitive aspect of the problem, the expertise of the expert, etc., and so virtually all reference data are inaccurate to some extent. If they are extremely inaccurate, or even bad, the validation process results can be misleading, in the sense that the algorithm outputs could be labeled poor when they are in fact good. The inaccuracy of reference data can also be attributed to a ***mislabeling problem***, which can have a direct negative impact on the validation process. To solve this problem, it is crucial that we develop automatic techniques for validating the reference data that we generate.

The problem of mislabeling is studied in (Wilson, 1972; John, 1995; Breslow and Aha, 1997; Brodley and Friedl, 1999; Li *et al.*, 2007; Ghoggali and Melgani, 2009). Wilson (Wilson, 1972) demonstrates that using only a few preclassified samples for a 3-NN classifier applied on a whole set of learning samples considerably improves the performance of a 1-NN classifier. Li (Li *et al.*, 2007) considered this problem as a class noise problem, in which the noise is modeled and incorporated into the Kernel Fisher discrimination model. In a C4.5 decision tree framework (Quinlan, 1986) that incorporates a pruning scheme, John (John, 1995) proceeds iteratively to eliminate the training samples that are not from the majority class, and then rebuilds the decision tree classifier from the reduced training set. Brodley presents a procedure in (Brodley and Friedl, 1999) for identifying mislabeled samples using an ensemble of classifiers called filters (k-NN, C4.5 decision tree, linear classifiers). The mislabeled samples are the ones that are misclassified by one of these filters. More recently, Ghoggali and Melgani (Ghoggali and Melgani, 2009) have studied the problem of mislabeling in the context of ground-truth validation for image classification in remote sensing. Their method is aimed at interacting with the ground-truth expert by providing binary information of the "validated"/"invalidated" type for each learning sample. The idea behind their method is to consider the detection of the mislabeled learning samples as an optimization problem under the genetic algorithm framework,

where the goal is to find the best learning subset, in terms of statistical separability between classes.

The method proposed in this article is based on the work of Ghoggali and Melgani (Ghoggali and Melgani, 2009). However, rather than calling on a single expert to validate the labels, we use a number of references generated by multiple experts to estimate new labels, on the assumption that, when the expert judgments are combined (agreement), they provide sufficient cues to lead to the construction of a comprehensive theory (Kaikova and Terziyan, 1997; Taylor *et al.*, 1995). The main idea is to use the votes (labels) submitted by different experts as a priori information to infer the a posteriori probability of labels under a Bayesian framework. The independence of experts is assumed. In our work, new reference data from multiple expert references are estimated as an optimization problem of multivariate classification in a multidimensional feature space. Two terms work together to define the energy function to be optimized, the first is related to data fidelity and the second to the full uncertainty with respect to expert voting that can be included in the form of a priori information. We provide the details of this process in section 5.2. Our motivation for developing this method is this: A single expert generation of reference data is likely to be biased by his knowledge of the reference data generation process and of the subsequent use of the intended application, not to mention his personal preferences, as explained in (Smith, 2010). The idea of estimating an accurate reference from multiple expert references originated in the medical imaging field (Yang *et al.*, 1995; Warfield *et al.*, 2004; Li *et al.*, 2009; Yuan *et al.*, 2009), the aim being to estimate accurate data based on the intra- and inter-variability of the experts. Our work differs from the state-of-the-art research in two important respects. First, it addresses MS images (multidimensional data), whereas other work processes gray-scale images, which means that the type of data processed is different. Second, it incorporates both original (i.e. gray-scale image) and binary data from multiple initial references in a single optimization model, whereas other works mainly consider binary data (binary references) in their models, and not original data. To the best of our knowledge, no similar work has yet been proposed to estimate a reference data for MS images that combines binary and original data.

The proposed reference estimation approach is evaluated based on the degree to which the final estimated labels could influence the separability of the new estimated classes, and their potential impact on the classification process. These factors are discussed in the next sections. The remainder of the paper is organized as follows: our proposed methodology is explained in section 5.2; the evaluation process is described in section 5.2.2; a case of study is presented in section 5.3; and our conclusions are provided in section 5.4.

5.2 Reference estimation methodology and its evaluation

The aim of this work is to propose a new statistical reference data estimation model for MS image analysis. It uses multiple references of experts as the initial input data to arrive at a more accurate standard reference. Two notions are introduced: accurate labeling, and mislabeling. A sample is considered accurately labeled (a full complement of votes) if all the experts agree on its label. A sample is considered mislabeled (less than a full complement of votes) if at least one of the experts disagrees with the others on its label. In spite of the fact that expert references always contain mislabeled samples, they are always more accurate than machine references. A combination of multiple expert references ensures a subset of accurately labeled samples that can be exploited for learning classifiers, in order to infer new accurately labeled samples to replace mislabeled samples. The process is as follows: a classifier model is trained with some accurately labeled samples known a priori, and used to estimate new labels for those that have been mislabeled. We note here that only mislabeled samples are involved in the new label assignment process. Thus, the estimation of new labels (targeted references) from existing accurately labeled samples can be achieved by combining two terms: data fidelity, which connects data to their most probable classes (i.e. likelihood), and a priori information about the number of votes a label receives (i.e. how many experts agree on the label). The explanation of how this combination works is, in principle, provided by the Bayesian framework.

5.2.1 General framework

The d-dimensional MS image is defined on a mesh, the vertices of which are a set S of sites (pixels) s such that $S = \{s = (i, j); 1 \leq i \leq N; 1 \leq j \leq M\}$. We now consider a couple of random fields $F = (X, Y)$, with $Y = Y_s, s \in S$ being the field of observation located on the lattice S, and $X = X_s, s \in S$ being the label field (class label). Each Y takes its value in $\lambda = \{0, .., 255\}^d$ and each X in the set of labels $\{c_0, .., c_k\}$. The classification process involves estimating label X from observation Y. It can also be viewed as a statistical labeling problem according to a global Bayesian framework in which the following a posteriori distribution has to be maximized (Geman and Geman, 1984):

$$\begin{aligned} \hat{x}_{MAP} &= \arg\max_x P_{X|Y}(x|y, \Theta) \\ &= \arg\max_x P_{Y|X}(y|x) P_X(x) \end{aligned} \tag{5.1}$$

where the $P_{Y|X}(y|x)$ is the likelihood function (i.e. the probability of the samples given the labels) and $P_X(x)$ is the a priori information about the labels. Assume that the samples given labels are conditionally independent, i.e.

$$P_{Y|X}(y|x) = \prod_{s \in S} P_{Y_s|X_s}(y_s|x_s) \tag{5.2}$$

for the likelihood model, we take the Gaussian law to describe the intensity distribution within each class, as follows:

$$P_{Y_s|X_s}(y_s|x_s) = \frac{1}{(2\pi)^{d/2}|\Sigma_{x_s}|^{1/2}} \exp\left(-\frac{1}{2}(y_s - \mu_{x_s})^\top \Sigma_{x_s}^{-1}(y_s - \mu_{x_s}) \right) \tag{5.3}$$

where μ and Σ are respectively the vector mean and the covariance of a class. Consequently, to revise the labels of the mislabeled samples, the likelihood term in Eq. 5.3 is modified as follows:

$$P_{Y_s|X_s}(y_s|x_s) = \max\left\{ P_{Y_s|X_s}(y_s|x_s), \mathsf{X}(m_v = V_s(x_s) \right\} \tag{5.4}$$

where $\mathsf{X}(a)$ is the indicator function, which is equal to 1 if a is true and 0 otherwise, m_v is the total number of experts, and $V_s(x_s)$ is the occurrence of the label x (the number of votes it received) assigned to the pixel at site s.

The a priori term is related to the probability of a label (the number of votes it received), and can be given as follows:

$$P_X(x) = \frac{1}{Z} \exp\{-\beta \sum_s (m_v - V_s(x))\} \tag{5.5}$$

where Z is a normalization factor, and β controls the importance of the a priori term. Clearly, if $m_v = V_s(x)$ (a full complement of votes), Eqs. 5.4 and 5.5 are set to their maximum values, keeping the energy at the maximum. This means that the current label doesn't change. In contrast, if $V_s(x) < m$ (less than full complement of votes), the energy function is computed according to the corresponding labels, and the most probable label wins.

Now, finding $\hat{x}_{MAP}$ is a classical problem of combinatorial optimization. Several algorithms, such as graph cuts (Boykov and Funka-Lea, 2006), loopy belief propagation (Yedidia *et al.*, 2003), ICM (iterated conditional modes) (Besag, 1975), SA (simulated annealing) (Metropolis *et al.*, 1953), etc., have been proposed to resolve this optimization problem. In our work, we used the ICM algorithm:

$$
\begin{aligned}
P_{X|Y}(x|y) &\propto \left(\prod_{s\in S} P_{Y_s|X_s}(y_s|x_s)\right) \times \left(\frac{1}{Z}\exp\left\{-\beta\sum_s\Big(m_v - V_s(x)\Big)\right\}\right) \\
&\propto \left(\prod_{s\in S} \max\Big\{P_{Y_s|X_s}(y_s|x_s), \mathsf{X}(m_v = V_s(x))\Big\}\right) \times \left(\frac{1}{Z}\exp\left\{-\beta\sum_s\Big(m_v - V_s(x)\Big)\right\}\right) \\
&\propto \left(\exp\left[-\Big(-\log\prod_{s\in S}\max\Big\{P_{Y_s|X_s}(y_s|x_s), \mathsf{X}(m_v = V_s(x))\Big\}\Big)\right]\right) \\
&\times \left(\frac{1}{Z}\exp\left\{-\beta\sum_s\Big(m_v - V_s(x)\Big)\right\}\right) \\
&\propto \frac{1}{Z}\exp\left[-\left(-\log\prod_{s\in S}\max\Big\{P_{Y_s|X_s}(y_s|x_s), \mathsf{X}(m_v = V_s(x))\Big\} + \beta\sum_s\Big(m_v - V_s(x)\Big)\right)\right] \\
&\propto \frac{1}{Z}\exp\left[-\left(\sum_{s\in S}\left(-\log\max\Big\{P_{Y_s|X_s}(y_s|x_s), \mathsf{X}(m_v = V_s(x))\Big\}\right)\right.\right. \\
&\left.\left. + \beta\sum_s\Big(m_v - V_s(x)\Big)\right)\right] \qquad (5.6)
\end{aligned}
$$

Finally, the classification is given by the maximization of the a posteriori probability, as follows:

$$\begin{aligned}\hat{x}_{MAP} &= \arg\max_{x}\Big\{P_{X|Y}(x|y)\Big\} \\ &= \arg\max_{x}\left\{\frac{1}{Z}\exp\left[-\left(\sum_{s\in S}\left(-\log\max\Big\{P_{Y_s|X_s}(y_s|x_s),\mathsf{X}(m_v=V_s(x))\Big\}\right)\right.\right.\right. \\ &\left.\left.\left.+\beta\sum_{s}\Big(m_v-V_s(x)\Big)\right)\right]\right\}\end{aligned} \tag{5.7}$$

Eq. 5.7 can be reformulated as the following energy function:

$$\begin{aligned}\hat{x}_{MAP} = \arg\min_{x}\left\{\sum_{s\in S}\left(-\log\max\Big\{P_{Y_s|X_s}(y_s|x_s),\mathsf{X}(m_v=V_s(x))\Big\}\right.\right. \\ \left.\left.+\beta\Big(m_v-V_s(x)\Big)\right)\right\}\end{aligned} \tag{5.8}$$

5.2.2 Evaluation

The estimated reference data is evaluated based on the assumption that mislabeled samples in a training set lead to a decrease in inter-class variability and a corresponding decrease in the performance of the subsequent classification tasks (Ghoggali and Melgani, 2009). Validation of this assumption is achieved in two steps: first, a measure of divergence is calculated; and second, training samples are selected from each class, and then used in the classification phase. This procedure is followed for each reference (the estimated reference, the expert references, and the reference generated by the majority voting method (MV) on the expert references). MV assigns a label to a sample if more than m experts (in general $m = \frac{m_v+1}{2}$, where m_v is the total number of experts) vote for this label. Of course, if the mislabeled samples are corrected by the proposed method, the estimated reference will lead to the greatest divergence, and therefore the best classification performance. The Bhattacharyya divergence (Kailath, 1967) is a suitable indication of divergence, and is a special case of the Chernoff distance, i.e. an upper bound of the probability of error of the Bayes classifier. For two multivariate Gaussian classes i and j (of given reference data), the Bhattacharyya divergence has a closed-form expression:

$$\mathcal{B} = \frac{1}{4}(\mu_i - \mu_j)^\top (\Sigma_i + \Sigma_j)^{-1}(\mu_i - \mu_j) + \frac{1}{2}\ln\left|\frac{\Sigma_i + \Sigma_j}{2}\right| - \frac{1}{4}\ln|\Sigma_i \Sigma_j| \quad (5.9)$$

where μ_i and Σ_i are the vector (spectral signature) mean and the covariance matrix of the class i respectively; and $|.|$ is the determinant. The higher the value of c_i, the greater the divergence between i and j. To show how the proposed mislabeling correction method positively impacts the performance of subsequent classification tasks, we compare the performance of two classifiers, k-nearest neighbor (k-NN) and multilayer feedforward neural network (MFNN) (Svozil *et al.*, 1997), before and after mislabeling. A case study on historical document image binarization is given in the next section.

5.3 Application: historical document image analysis

One of the more important applications in the image analysis field is historical document image binarization. In general, a document foreground is written or printed in different levels of gray from those of the background. Binarization consists of separating the foreground pixels from the background pixels. The simplest way to do this is to choose a threshold value and classify all the pixels with values above this threshold as background, and all the other pixels as foreground. Binarization is a crucial issue, because of its impact on subsequent applications, such as document recognition (i.e. OCR) and understanding. If the binarization output is not accurate, the OCR results won't be accurate either. Many binarization algorithms have been proposed in the last decade (Wang *et al.*, 2003; Farrahi Moghaddam and Cheriet, 2010a,b; Hedjam *et al.*, 2011b; Hedjam and Cheriet, 2011b,a; Lu *et al.*, 2010; Gatos *et al.*, 2008, 2009a). To track their progress over time towards a level consistent with human analysis, these algorithms should be compared to reference data or ground-truths. A common way to do this is to have an expert manually labels the document image pixels as foreground or background. Unfortunately, when dealing with intensity-based images (gray-scale or color images), the task of manual labeling becomes difficult in the presence of degradation effects, such as lack of contrast, interfering patterns, and ink fading, which are common on historical document images. As a result, manual ground-truthing is always highly subject to the possibility of labeling errors. For example, a pixel might suggest the presence of an object which is, in fact, absent, or it could be labeled as representing an object, when, in reality, it doesn't. These two errors occur frequently, in both manual and semiautomatic ground-truthing. In addition, when text

and background are labeled in document image binarization, the pixels located on the border between the text and the background have a high probability of being labeled incorrectly (Smith, 2010). This is because it can be difficult to decide at which gray level a pixel should be considered text or background. In contrast, the pixels located near the middle of the text have high probability of being labeled correctly. If they are, they are referred to as accurately labeled pixels. Where the expert encounters particular difficulty is in the vicinity of pixels located on the borders of the text strokes. The reason for this is the limited representation space, which offers only a subset of the information that is available by combining the responses of all the visible light into three spectral images (RGB) or less [38]. Although the RGB color space is the most common choice for computer graphics, it is not very efficient for dealing with real-world images, because the RGB channels contain redundant luminance information. Consequently, based on color, or only on gray-scale intensity, confusion arises as to what is text and what is background. This makes separating the two very difficult, especially with severely degraded historical document images. To cope with this limitation, multispectral (MS) and hyperspectral images are used instead, as they can provide a detailed quantitative measurement of the spectral responses to help analyze the scene. In that space, the objects that appear to have similar photometric properties under visible light (RGB) can behave differently when viewed under IR or UV light. The availability of multiple information further helps in the generation of ground-truths by mapping the spectral signatures of document pixels to references known a priori.

Ten case studies are reported in this work. Each case consists of one MS document image with 8 bands: 1 at a UV wavelength (400 nm), 3 at visible wavelengths (BLUE=400 nm, GREEN=500 nm, RED=600 nm), and 4 at IR wavelengths (700 nm, 800 nm, 900 nm, 1000 nm and 1100 nm). The 8 bands are aligned and corrected for chromatic aberrations, dark current noise, and uneven illumination defects. The data contain various types of degradation, and the documents were written in different time periods. The MS images are based on historical documents collected from the *Bibliothèque et Archives nationales de Québec* (BAnQ) [1] archives, and were written between the 17^{th} and 20^{th} centuries. Most of the 20^{th} century documents were character-printed, and so are less degraded than the earlier ones, which were handwritten with iron gall-based (ferro-gallic) ink. The handwritten documents are highly degraded. The majority of them were written with ink made with iron, salts, and tannic acid from vegetable sources. It was the standard writing and drawing ink from about the 12^{th} century to the 19^{th} century, and remained in use well into the 20^{th} century. It has a particular photometric characteristic when examined under IR and UV light. For each MS document image,

[1] http://www.banq.qc.ca/accueil/index.html

eight ground-truths were generated semiautomatically in two steps, and considered as expert ground-truths (denoted $\mathsf{G}_i, i = 1..8$). In the first step, eight different binarization algorithms were applied to produce the initial binarization maps, the corresponding parameters of which were set to their optimal values. In the second step, a manual refinement was applied, aimed at removing artifacts and recovering potentially missing text strokes, in order to create clean ground-truths. The objective is to generate ground-truths that reflect reality as best as possible, in the sense that ground-truths differ from one another at the pixels located on the borders of the text (which is mostly true). An MV method is applied to estimate a ground-truth (denoted G_m) from the expert ground-truths. Finally, the proposed method is applied to estimate the targeted ground-truth (denoted G_e).

Table 5.1 shows the divergence measures of each ground-truth in all the cases. We can see that the ground-truth with the highest divergence measure is G_e. There may be some exceptions, such as in cases 7 and 9, where the ground-truths were generated by experts 2 and 5 where the divergence measures equal to that of the estimated ground-truth. On average, the proposed method provides images that lead to higher divergence measures. Also, on average MV is able to generate ground-truths better than individual experts.

Table 5.1 Divergence based comparison between different ground-truths

Cases	**Manual ground-truths**								G_m	G_e
	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8		
Case0	2.2	1.7	1.5	1.3	1.5	1.5	1.8	1.7	1.9	**2.5**
Case1	1.9	2.6	2.1	2.2	2.5	2.4	2.0	2.4	2.5	**2.7**
Case2	1.8	1.8	1.9	1.8	1.8	1.7	1.9	1.8	1.9	**2.1**
Case3	2.1	2.1	1.7	1.8	2.0	2.0	1.9	1.6	2.2	**2.4**
Case4	1.8	1.6	1.7	1.7	1.7	1.7	1.4	1.7	1.7	**2.0**
Case5	2.3	2.0	2.2	1.8	2.3	1.8	2.2	1.6	2.2	**2.6**
Case6	2.0	2.2	2.1	2.2	2.1	1.9	2.2	2.0	2.2	**2.3**
Case7	3.2	**6.8**	3.5	3.1	3.2	3.1	3.2	2.9	3.2	**6.8**
Case8	2.8	2.0	1.9	2.4	2.4	2.5	2.6	2.7	2.6	**3.0**
Case9	2.6	3.3	3.3	3.2	**3.7**	3.2	2.5	2.6	3.6	**3.7**
mean	2.3	2.6	2.2	2.2	2.3	2.2	2.2	2.2	2.4	**3.0**

For particular applications, such as document image binarization, the contour of a stroke can be considered a very important feature, in addition to inter-class divergence (Lu *et al.*, 2010), as it plays a key role in ground-truth generation, because this is where the labeling confusion arises. For this reason, we also compare the various ground-truths based on contour infor-

mation, mapping the contour of each ground-truth to that of the original spectral images (i.e. reference contour). According to our MS data acquisition system, the text is sharpest under visible light, and so the reference contour is computed from the visible bands (a combination of color spectral images). A simple Canny edge detector can be useful in this computation. The *F-measure* (van Rijsbergen, 1979), the NRM (Negative Rate Metric) (Young and Ferryman, 2005), and the PSNR (Peak signal-to-noise ratio) are used as measures of spatial matching between different contours. We define the true positives, or the number of tests confirmed by the reference, as TP; the true negatives, or the number of tests rejected by the reference, as TN; the false positives, or the number of tests not matched in the reference, as FP; and the false negatives, or the number of tests erroneously accepted as belonging to the reference, as FN. The *F-measure* is a measure of test accuracy, and includes both the precision and the recall of the test to compute the accuracy score:

$$F = 2 \cdot \frac{precision \times recall}{precision + recall} \tag{5.10}$$

where

$$precision = \frac{TP}{TP + FP} \quad \text{and} \quad recall = \frac{TP}{TP + FN} \tag{5.11}$$

NRM is based on pixel-wise mismatches between the reference map and the test map, and is computed as:

$$NRM = \frac{1}{2}\left(\frac{FN}{TP + FN} + \frac{FP}{FP + TN}\right) \tag{5.12}$$

The PSNR measure between a reference map r and a test map t, both of size $N_x \times N_y$, is defined by:

$$PSNR(r,t) = 10 \ln_{10}\left(\frac{1}{MSE(r,t)}\right) \tag{5.13}$$

where

$$MSE(u,t) = \frac{1}{N_x \times N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \Big(r_{ij} - t_{ij} \Big) \tag{5.14}$$

and MSE is the Mean Square Error between r and t. The higher the value of PSNR, the less difference there is between r and t; and the smaller the value of PSNR, the greater the difference between r and t.

Tables 5.2, 5.3 and 5.4 show the *F-measure*, NRM, and PSNR values of each ground-truth in all the cases respectively. We can see from these tables that, in general, the estimated ground-truth leads to better matching of the contours with the reference than the other ground-truths. The exception is Expert 1 in Case 5, but, on average, the contours of the estimated images are more accurate than those of the other ground-truths.

Table 5.2 F-measure based comparison between the reference contour and those of different ground-truths

Cases	**Manual ground-truths**								G_m	G_e
	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8		
Case0	63.9	56.4	50.1	50.5	50.5	51.1	61.2	60.7	74.5	82.1
Case1	71.7	76.4	74.6	76.2	76.2	76.4	74.5	74.5	86.3	87
Case2	75.8	73	74.4	72.1	71.1	69.3	74.3	75.2	78.4	79.1
Case3	78.3	75.9	69.5	64.1	72.2	72.2	73.7	67.6	79.9	81.4
Case4	83.2	78.7	80.9	82.8	82.5	81.3	79.8	82.5	84.1	85.5
Case5	77.2	70.4	73	62.6	74.2	51.4	73.5	51.6	74.5	75.4
Case6	79.6	75	72.4	75.7	68	66.6	75.1	72.7	77.3	80.8
Case7	80.4	81.6	59.8	72.2	74.4	75.6	74.4	64.7	76	86
Case8	81.1	66.6	64.9	78.6	79.1	79.1	69.9	80.7	81.8	84.6
Case9	76.7	75.4	73.5	80.4	78.8	79.5	69.2	59.3	81.6	82.1
mean	76.8	72.9	69.3	71.5	72.7	70.3	72.6	69	79.5	82.4

In order to visually inspect the correspondence of the contours, the contour maps generated from G_m and G_e are mapped to that generated from the visible bands (see Fig. 5.1). The procedure is designed to create an RGB image in which the R layer contains the black map, the G layer contains the reference contour map, and the B layer contains the G_m contour map. The same procedure is performed for G_e. The presence and absence of contours are indicated by 1 and 0 respectively. Consequently, there are 4 possible alternatives for the **(r,g,b)** triplet:

Table 5.3 NRM based comparison between the reference contour and those of different ground-truths

Cases	**Manual ground-truths**								G_m	G_e
	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8		
Case0	0.24	0.27	0.3	0.3	0.3	0.29	0.25	0.25	0.14	0.1
Case1	0.18	0.16	0.17	0.16	0.16	0.16	0.17	0.17	0.07	0.06
Case2	0.12	0.14	0.13	0.14	0.15	0.16	0.13	0.13	0.09	0.09
Case3	0.12	0.14	0.17	0.2	0.16	0.16	0.15	0.18	0.1	0.1
Case4	0.09	0.11	0.1	0.09	0.09	0.1	0.1	0.09	0.08	0.07
Case5	0.12	0.16	0.14	0.2	0.14	0.25	0.14	0.25	0.13	0.13
Case6	0.1	0.13	0.14	0.12	0.16	0.17	0.13	0.14	0.11	0.1
Case7	0.12	0.1	0.22	0.16	0.15	0.15	0.15	0.21	0.14	0.07
Case8	0.12	0.19	0.2	0.13	0.12	0.12	0.18	0.12	0.11	0.08
Case9	0.13	0.12	0.13	0.1	0.11	0.1	0.18	0.22	0.09	0.09
mean	0.13	0.15	0.17	0.16	0.15	0.17	0.16	0.17	0.11	0.09

Table 5.4 PSNR based comparison between the reference contour and those of different ground-truths

Cases	**Manual ground-truths**								G_m	G_e
	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8		
Case0	13.2	12.3	11.7	11.7	11.8	11.8	12.9	12.8	15.5	17
Case1	15.6	16.4	16.1	16.4	16.4	16.4	16.1	16.1	19.3	19.5
Case2	16.1	15.6	15.8	15.5	15.3	15.1	15.8	16	16.7	16.9
Case3	16.4	16	15	14.3	15.4	15.4	15.6	14.7	16.9	17.3
Case4	18.8	17.7	18.2	18.7	18.6	18.3	18	18.6	19	19.4
Case5	16.2	15.1	15.5	14.1	15.7	12.9	15.6	12.9	15.7	15.9
Case6	18.4	17.5	17.1	17.7	16.4	16.2	17.6	17.1	18	18.7
Case7	16.1	16.1	12.9	14.5	14.9	15.1	14.9	13.7	15.2	17.3
Case8	15.5	13	12.8	14.9	15.1	15.1	13.6	15.4	15.7	16.4
Case9	15.6	15.2	14.8	16.2	15.9	16	14.5	13.1	16.4	16.6
mean	16.2	15.5	15	15.4	15.5	15.2	15.4	15	16.8	17.5

1. (0,1,1): **TP** (reference contour is detected)
2. (0,1,0): **TN** (reference contour is not detected)
3. (0,0,1): **FP** (noise is introduced)
4. (0,0,0)

Fig. 5.1 shows that G_e more accurately detects contours altered by the correction of mislabeled pixels in the vicinity of the pixels located on the borders between text and background. G_e generates a higher number of TP (Cyan) contours, and a lower number of TN (Green) and

FP (Blue) contours than G_m. This leads to better different measures (F-*measure*, NRM and PSNR).

We found experimentally that there is a direct relationship between inter-class divergence and the detected contours: divergence increases if the contours are accurately detected, and vice-versa. To support this, we compute the correlation coefficient between the *F-measure* for example and the divergence values computed on G_m and G_e (see Fig. 5.2). The correlation coefficient ρ_m is computed between the 10th column (G_m)of Table 5.1 and the 10^{th} column (G_m) of Table 5.2; and the correlation coefficient ρ_e is computed between the 11^{th} column (G_e) of Table 5.1 and the 11^{th} column (G_e) of 5.2. To be able to compare the calculations, the *F-measure* and the divergence measures must first be normalized. Fig. 5.2 shows the normalized values of these measures for the ten cases studied, for both G_m (Fig. 5.2 (a)) and G_e (Fig. 5.2 (b)). As we can see, because the G_e contours are more accurately detected than those of G_m, the value of ρ_e is about $3/2$ than that of ρ_m, thanks to the correction of mislabeled samples. This makes it possible to accurately locate the boundaries between text and background, and also increases the divergence between these two classes.

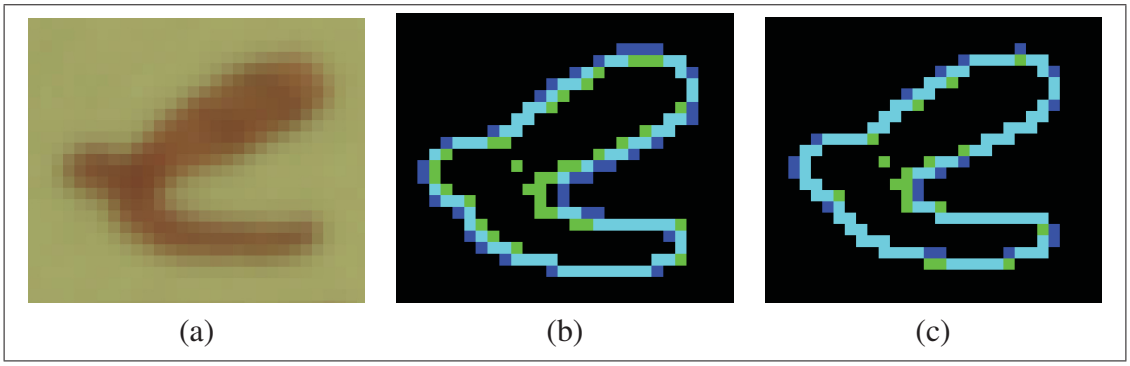


Figure 5.1 Visual contour-based comparison. (a) original image, (b) mapping between the contours of G_m and the the contours of the original image, (c) mapping between the contours of G_e and the the contours of the original image

Now, to show experimentally how the proposed method for correcting mislabeling positively impacts the performance of subsequent classification tasks, we compared the performance of two classifiers, namely k-NN and MFNN (Svozil *et al.*, 1997), on mislabeled data (i.e. G_i, $i = 1:8$) and on corrected data (G_m and G_e). For each piece of data, a set of training samples is picked automatically and randomly from the foreground and background classes. For the test phase, a set of test samples is picked from the accurately labeled domains (see Fig. 5.3

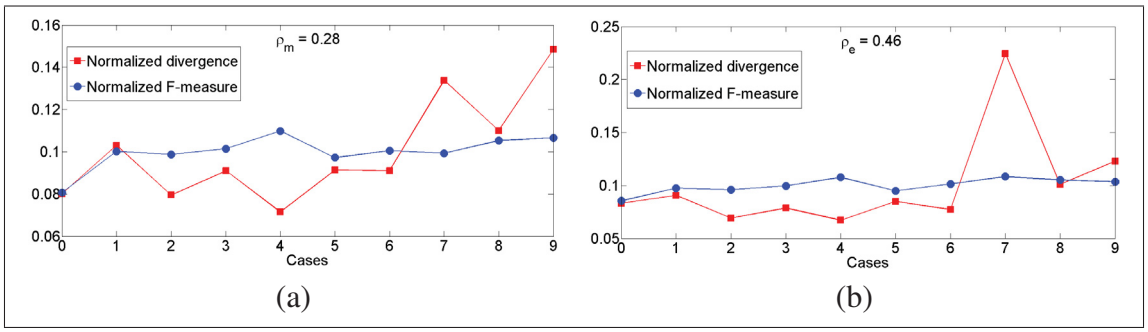


Figure 5.2 Correlation between F-measure on contours and inter-classes divergence measures. (a) majority voting method; (b) proposed method

(b)). This ensures that the training and testing data are disjoint. This experiment is performed twice. In the first experiment, there are 50 training samples and 50 test samples. In the second experiment, there are 100 of each. The classification parameters are adjusted as follows: the values of k in the k-NN classifier are 1, 3, 5, and 7. In the MFNN (a matlab toolbox is used by default with 10 hidden layers) half the training set is devoted to validation and half to testing. The results, in terms of the classification error rate, are reported in Table 5.5. The values are the average error rates over the 10 cases studied. For example, 8.79 is the average error rate of 1-NN, and 0.12 is the average error rate of MFNN, both over the 10 ground-truths generated by Expert n^o. 1. The values 8.13 and 3.45 are the error rates of 7-NN over the 10 ground-truths estimated by the Majority Voting method and the proposed method respectively.

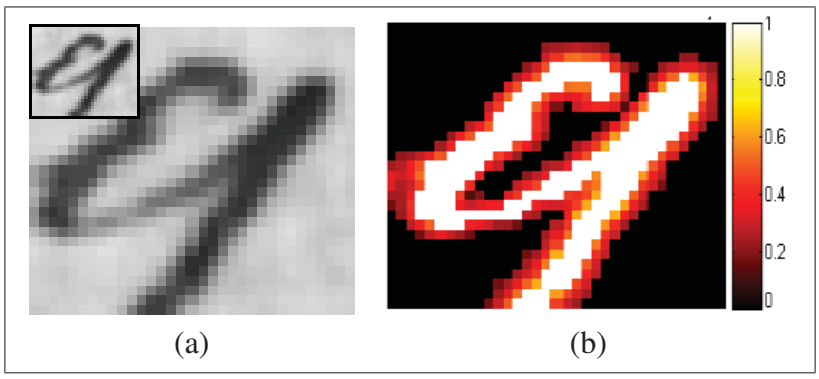


Figure 5.3 Domain definition. (a) visible band; (b) average ground-truth where the white color (1) means accurately-labeled text, black (0) means accurately-labeled background, and between 0 and 1 means partially-labeled text or background

From Table 5.5, we can see that our proposed method provides accurate ground-truths, leading to fewer classification errors than with MV. The comparison illustrates that individual experts' ground-truths lead, on average, to the worst classification errors, and, although MV labeling performs better than expert labeling, the results are worse than with the proposed method. The disadvantage of MV is that it is only designed to provide a trade-off that minimizes the average disagreement between its estimated ground-truth and the expert ground-truths, and ignores the data similarities (data fidelity). It is also possible that more than half of the experts will vote for labels that do not represent the right classes. The proposed method, in addition to considering the occurrence of labels (number of votes), addresses data fidelity in a weighted model (i.e. Eq. 5.8) to estimate the a posteriori labels. This leads to improved accuracy of the estimated ground-truths and validates our hypothesis.

A simple comparison of MFNN and k-NN reveals that the former is more efficient (lower error rate) and seems less sensitive to the number of training and test samples. As for the value of k, it does play a role (but not a major one) in reducing the classification error. In our case, low values lead, in general, to better classification performance.

The most important parameter in the proposed model is β, which acts as a weighting factor between the data fidelity and a priori information terms. It acts as a tradeoff between these two terms, in that the estimation model generates smooth outputs while keeping the most likely data membership. That is, the higher the value of this parameter, the greater the influence of the expert decision on the data, and therefore the fewer artifacts and holes in the outputs. The lower the value of this parameter, the less dominant the expert decision, and therefore the more accurate the data. Fig. 5.4 shows a typical example of how β influences the estimated ground-truths. The aim is to clean the artifacts of the image, as shown in Fig. 5.4(a-e), while preserving

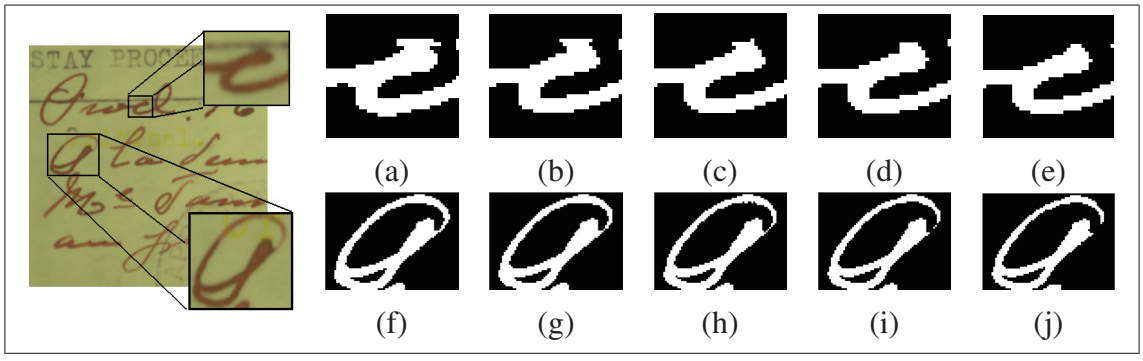


Figure 5.4 Influence of β on the reference data estimation process (from left to right: $\beta = 0.1, 0.5, 1, 5, 10$)

the connections between the strokes, as shown in Fig. 5.4(f-j). Visually, the compromise shown in Fig. 5.4(c,h) seems reasonable. Experimentally, the value of β is set to 1 or 5 for the majority of the case studies. It is also possible to learn this value from the data, in which case, estimating a reference becomes a learning problem, and is outside the scope of this work.

5.4 Conclusion

In this paper, a new method for reference data estimation is introduced, which uses multiple reference data generated manually or semiautomatically by several experts to estimate new more accurate data. An expert's reference data are supposed to be quasi-optimal and better than machine-generated reference data, which is true in most cases. However, expert references are subject to mislabeling and judgment errors, because they are based on the individual expert's experience and many other factors, especially in the case of degraded data that contain confusing samples. In the proposed approach, more experts are involved, in order to overcome the subjectivity related to the manual generation of reference data. Each expert is asked to generate one or more samples of reference data. Labeling uncertainty that result from combining expert judgments (based on voting) can be included in the form of a priori information, which is updated based on observational data in a multidimensional representation space. In this regard, a Bayesian framework is proposed to estimate the a posteriori probability of the pixel labels that will represent the final estimated reference data.

The proposed method was evaluated based on the assumption that mislabeling in a training dataset decreases inter-class variability, and so increases intra-class variability. In practice, this assumption is validated by a real application of document image binarization, which is the most

Table 5.5 Classification error rate (%)

			Error (%)									
classifiers	Tr|Tst	k	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_m	G_e
k-**NN**	50|50	1	8.79	8.078	7.64	8.6	7.42	9.72	7.72	7.25	7.42	**3.32**
		3	8.36	9.865	10.2	8.01	7.71	11.7	7.59	8.43	7.65	**3.39**
		5	10.7	11.48	9.35	8.74	8.4	11.3	9.18	9.15	8.7	**3.41**
		7	9.12	10.36	10.7	10.9	8.38	11.1	9.52	8.29	8.13	**3.45**
	—-	–										
	100|100	1	6.96	7.193	6.21	6.56	5.83	7.27	6.07	6.5	5.61	**3.35**
		3	6.14	6.532	6.78	5.94	6.48	7.82	6.16	6.49	5.97	**3.33**
		5	6.38	6.601	6.73	5.51	5.78	7.42	6.49	6.34	6.3	**3.39**
		7	5.93	6.712	7.94	5.69	7.06	7.5	7.21	7.35	6.27	**3.44**
MFNN	50|50		0.12	0.268	0.26	0.16	0.12	0.26	0.2	0.26	0.28	**0.06**
	100|100		0.21	0.128	0.22	0.11	0.11	0.25	0.14	0.15	0.12	**0.06**

important application in document image analysis owing to its impact on subsequent tasks, such as OCR (optical character recognition), document understanding, etc. The estimated reference data (ground-truths) are compared individually to each expert's reference data, and also to the reference data estimated by the MV (majority voting) rule, in terms of divergence, contours, and impact on classification tasks. The results show that the proposed method is able to estimate more accurate reference data, as a result of combining the judgments of multiple experts and addressing data fidelity in a multidimensional representation space (multispectral images).

The advantages of this method are that it is simple, easy to implement, and has a positive impact on the performance of subsequent classification tasks. In addition, the Bayesian framework has the advantage of being open to including other terms, such as MRF, which are responsible for homogenizing the outputs. The difficulty of this method lies in finding experts. Manual generation is time-consuming and requires a great deal of patience.

In future work, we will investigate the learning process for the weighting parameter β of the proposed Bayesian model, in order to make the proposed algorithm fully automatic.

Acknowledgments

The authors would like to thank the NSERC of Canada for their financial support. We would like to extend our thanks to the BAnQ for providing valuable ancient documents.

CHAPTER 6

GENERAL DISCUSSIONS

The overall objective of this thesis has been to define a new image processing framework for the enhancement and restoration of HDI to make them legible and more accessible. Our proposed general methodology consists of three themes, which we covered in this work: Chapter 3 presented a new adaptive soft thresholding method for HDI binarization that is robust in terms of recovering weak connections between text strokes, and which produces more accurate outputs for better accessibility to data following treatment by high level analysis methods. Chapter 4 introduced a new variational method for HDI restoration that is robust in terms of eliminating degradation from HDI, while keeping their original appearance intact, which improves both their visual quality and their legibility. Finally, a new RD estimation method for HDI binarization was presented in chapter 5. Each theme is the subject of a separate published journal article. Although the themes seem to be independent, they are, in fact, complementary, and together they form our general framework. Below, we briefly discuss the strengths and weaknesses of the proposed method as reflected in each theme.

6.1 Adaptive soft thresholding for intensity-based HDI binarization

The first theme covered the issue of the historical document image (HDI) enhancement by means of binarization, with the aim of offering better accessibility to data following treatment by high level analysis methods. We have presented a new soft thresholding approach for text/background separation in an intensity-based representation space (gray-scale or color images). Briefly, the binarization process consists first of detecting the most likely text pixels (i.e. regions of interest – ROI), which are used as a mask to roughly separate text from background. Then, for each masked document image pixel, the mean and variance of each class (text and background) are estimated and interpolated over the unmasked document's image pixels. Finally, a maximum likelihood (ML) classifier is applied to classify the document image pixels based on their local class membership. However, to capture the weak text strokes (i.e. very thin strokes of low intensity), a morphology dilation is performed on the ROI initially detected, followed by a simple Otsu thresholding to include text pixels of low intensity in the parameter estimation phase. Locally, this process accurately separates text from background, leading to the preservation of weak connections between text strokes, which is an advantage that the other state-of-the-art methods do not possess. Preserving weak connections is of great interest in subsequent tasks involving character recognition and information retrieval. In ad-

dition, this binarization method has a major advantage over other methods, which is that it is based on soft thresholding, and so there is no requirement to set a large number of parameters. In fact, it is a parameterless thresholding method, which is easy to implement. We tested the proposed binarization method on the DIBCO'09 dataset, which consists of a set of real HDI captured with a color camera, converting the majority of the document images to gray-scale images. Our comparison results show that our method nearly outperforms the state-of-the-art, and is the best binarization method developed to date. However, it performs less well in some instances, especially in the handling of small looped characters that have been degraded by blur, which makes the local estimation of parameters unreliable. In the future, some preprocessing steps, such as denoising and deblurring, will be investigated, with the aim of producing sharp text contours. In addition, we will explore the idea of introducing Markov random fields (MRF) and a priori information on the spatial connectivity between the labels as an alternative solution to maximum likelihood (ML), in order to achieve a better classification. However, in spite of the achievements of existing binarization methods (including our own), the problem of text/background separation remains unresolved. What we can say is that, while the results of the state-of-the-art methods are encouraging, the question of generalizability has not yet been answered. This is because some binarization methods can be successful in some cases, but fail in many others, owing to the non linear and unpredictable nature of the degradation. It is very difficult, if not impossible, to develop reliable classification models based on features extracted based simply on intensity information. Two promising lines of enquiry have emerged to circumvent this problem:

1. Research based on reverse engineering, the principle of which is the following: if a degradation phenomenon can be modeled, then it is also possible to retrieve the original document image (i.e. before it suffered degradation) by reversing this model on the observed document image. Ink bleed-through is a typical example. If its diffusion through the fibers of the sheet can be simulated, then it is feasible to reverse the diffusion towards the verso side of the document image.

2. Research into multispectral imaging that allows objects to be analyzed based on their photometric properties, rather than on their intensity. This technology uses visible light and invisible light, such as infrared (IR) and ultraviolet (UV). The document in question is analyzed in terms of its photometric response along the electromagnetic spectrum from UV, to visible, to IR. This means that a pixel is represented by a vector of many values (its spectral signature), rather than as a single scalar value, as in the case of intensity-based document images. Each value corresponds to the portion of the light reflected by the

surface of the object at a specific wavelength. This technology has one major advantage, which is that objects in a multispectral (MS) space representation that appear to have the same photometric properties under visible light (RGB) can behave very differently when viewed under UV or IR. Given that the ink and the degradation are containing different chemical components, their spectral signatures will also be different.

6.2 Variational method of multispectral HDI restoration

Another critical need for historians and librarians is the ability to visualize their documents on digital screens. For better visualization and legibility, the degradation on these documents needs to be removed, the main concern being to do so while preserving their original appearance (the second theme of the thesis). To achieve this, we have proposed a novel restoration model, and demonstrated that it is possible to remove the degradation from the multispectral images of documents without changing their original appearance. However, this cannot be achieved without also applying a solid mathematical model that allows the incorporation of the photometric properties of the objects in a variational image restoration formalism. The basic idea behind this approach is simply that the ink, which is iron gall-based, can be seen in the visible spectral images (color) and begins to disappear from the first IR spectral image, while the degradation can be seen in all the spectral images. Taking this key feature into account, and considering degradation as missing data in the visible spectral images, we have developed a restoration model based on an inpainting framework aimed at isolating the degradation in the IR spectral images and then inpainting them from the visible spectral images. The restored visible spectral images are used to reconstruct the HDI to be displayed. A problem arises when the degradation covers a portion of the text. In such cases, the inpainting process can delete useful information (text). A preliminary solution based on the extraction of end-members can circumvent this obstacle, because it allows ink pixels to be detected, and the algorithm can detach them from the inpainting mask before the inpainting takes place. We have also shown that inpainting, which is among the few variational models that are readily open to both theoretical analysis and efficient computational implementation, does not seem to work well with large missing data (larger than the inpainting scale). This is because inpainting only involves the first geometric information using the shortest possible connection to interpolate the level lines (Chan and Shen, 2002b); otherwise, the technique could generate smooth areas instead of reproducing the original missing textures. Consequently, high order geometric information (i.e. length) is mandatory if large missing data are to be reproduced accurately. Overall, the restoration model we have proposed can remove degradation while preserving the original appearance of the document. Moreover, the quality of the enhanced document images is, for

the most part, satisfactory, according to the feedback we have received from our collaborator (BAnQ: Bibliothèque et Archives nationales de Québec).

6.3 Reference data estimation in a multispectral representation space

Since the performance of the binarization algorithm has a direct impact on OCR engines, an objective evaluation based on RD (i.e. ground-truths) is mandatory, and RD are essential for HDI research. The most common method of generating RD is to involve an expert who is able to segment the data of interest according to the end use of the processed data. Unfortunately, human RD generation is subject to mislabeling and judgment errors on the part of the expert, as well as being a subjective process influenced by his personal preferences and his expertise, all of which affect the quality of the processed data. To address these limitations, we have defined a new method of RD estimation based on multiple RD generated by a number of experts (the third theme of the thesis). The labeling uncertainty that may result from combining expert judgments (based on voting) can be included in the form of a priori information, which is updated based on observational data in a multidimensional representation space. To summarize, what we have proposed is a Bayesian framework to estimate the a posteriori probability of the pixel labels that will represent the final estimated RD.

Our proposed method was evaluated based on the assumption that mislabeling in a training dataset decreases inter-class variability, and so increases intra-class variability. In practice, this assumption is validated by a real application of document image binarization, which is the most important application in document image analysis, owing to its impact on subsequent tasks, such as OCR, document understanding, etc. The estimated RD (ground-truths) are compared individually to each set of expert RD, and also to the RD estimated by the MV (majority voting) rule, in terms of divergence, contours, and impact on classification tasks. The results show that the proposed method is able to estimate more accurate RD, as a result of combining the judgments of multiple experts and addressing data fidelity in a multidimensional representation space (MS images).

The advantages of this method are that it is simple, easy to implement, and has a positive impact on the performance of subsequent classification tasks. In addition, the Bayesian framework has the advantage of being open to the inclusion of other terms, such as MRF, which are responsible for homogenizing the outputs. The difficulty with it lies in finding experts. Manual generation is time-consuming and requires a great deal of patience. In future work, we will investigate the learning process for the weighting parameter β of the proposed Bayesian model, in order to fully automate the proposed algorithm.

GENERAL CONCLUSION

In this thesis, we have addressed three important themes related to the image processing of historical documents, which constitute the most important element of the visual cultural heritage. The three themes studied in this thesis are the cornerstones of the low level processing of historical document images (HDI). We have introduced these themes in a particular sequence to emphasize the importance of choosing an adequate representation space in which the data are considered for processing. Roughly speaking, gray-scale and color images have the advantage of showing data appropriately for visualization and reading purposes. Unfortunately, they are not ideal for classification or analysis, because of the high correlation among the red, green, and blue (RGB) components. Also, the measurement of color in an RGB space does not represent color differences on a uniform scale, which makes it impossible to evaluate the similarity between colors from their distance apart in this space (Cheng *et al.*, 2001), and therefore it is difficult to separate text from degraded background. To solve this problem, physics-based imaging techniques are used, which are based on the reflection of light from the materials making up the objects. Among these techniques are multispectral imaging and hyperspectral imaging. Although they have advantages that techniques based on gray-scale or color do not posses, too many assumptions must be made regarding the type of material, the light source, and the illumination conditions that cannot always be substantiated in the real world.

In restricted representation spaces, such as gray-scale or color spaces, the text extracted from a document with a degraded background by conventional methods based on hard thresholding is not accurate. By contrast, local adaptation of soft thresholding to variations in document intensity has been shown to be effective, as it is robust in handling local intensity variations statistically, thanks to its accurate local estimation of the classification model parameters. This makes it possible for the binarization process to separate text from background, and preserve weak connections between text strokes with a high degree of accuracy, even in cases of severe degradation, and is a significant improvement over existing binarization techniques.

As for HDI restoration, physics-based imaging techniques (i.e. multispectral imaging) are required that provide reflectance measurements of focused ROI over a wide field of view (high resolution). MS imaging provides us with additional information about a document in the infrared (IR) and ultraviolet (UV) range of the electromagnetic spectrum, as objects that look similar in visible light can look very different under IR and UV light. Integrating this information into a solid inpainting-based variational restoration model increases the legibility of degraded HDI.

Dealing with HDI in an MS representation space is also important in the quantitative estimation of RD, which is necessary for evaluating and validating the processing algorithms. We have demonstrated that, in this space and with the introduction of the notion of multiple reference data (i.e. generated by a number of experts), it is possible to estimate new, more accurate reference data than those estimated by the majority voting rule or generated individually by experts.

Open questions

There are many problems in Document Image Analysis and Processing such as enhancement, restoration, recognition, etc., which are not resolved. This is due to several reasons. To know these reasons, we must find answers to these open questions:

1. Is it possible to revise the physics of image acquisition as well as those related to the mathematical modeling of their formation in order to have better subjective and quantitative control of these images?

2. Is it possible to develop generic methods to solve more than one type of degradations of documents? Otherwise, several questions arise: how many type of degradations are there and how many methods should be developed?

3. If we accept the current situation, and we continue to live with what we have as physical devices of image acquisition and mathematical solutions (statistical models, pattern recognition techniques, image processing approaches, etc.), what are the attributes or the features that we need to capture for providing accurate datasets allowing to design robust and reliable systems, for learning as well as for test; and also to respond to specific questions posed by end-users about their document?

Summary of contributions

In this thesis, we have made the following contributions:

1. Defined a new model of HDI enhancement using intensity-based binarization. The principle is to adapt a soft thresholding technique in a maximum likelihood framework to handle local variation in document intensity, in order to capture the weak connections between the text strokes, ultimately providing better access to historical data.

2. Devised a new approach to HDI restoration in a multispectral representation space. The principle is to incorporate IR information in a solid variational restoration model based on an inpainting framework to remove degradation from the color spectral images, with a view to improving the legibility of historical data.

3. Defined a novel reference data estimation model for HDI binarization. The principle is to incorporate multiple sets of reference data generated by a number of experts and multispectral information in a single Bayesian classification framework to estimate new, more accurate reference data. For the first time, a quantitative meaning is imparted to reference data, overcoming thus the limit of the subjectivity notion tied to the manual generation of the reference data.

Articles in peer review journals

1. Rachid Hedjam and Mohamed Cheriet. Reference data estimation in multidimensional representation space: application to historical document image analysis. Submitted to IEEE Pattern Analysis and Machine intelligence, Trans. on (TPAMI); 2013.

2. Rachid Hedjam and Mohamed Cheriet. Historical document image restoration using multispectral imaging, Pattern Recognition Journal. Elsevier Pattern Recognition (PR), 46(8), August 2013, Pages 2297–2312. DOI=10.1016/j.patcog.2012.12.015 http://dx.doi.org/10.1016/j.patcog.2012.12.015.

3. M. Cheriet, R. Farrahi Moghaddam and Rachid Hedjam. A learning framework for the optimization and automation of document binarization methods. Elsevier Computer Vision and Image Understanding (CVIU), 117(3), March 2013, Pages 269–280. DOI=10.1016/j.cviu.2012.11.003 http://dx.doi.org/10.1016/j.cviu.2012.11.003

4. Rachid Hedjam, Reza Farrahi Moghaddam, and Mohamed Cheriet. 2011. A spatially adaptive statistical method for the binarization of historical manuscripts and degraded document images. Elsevier Pattern Recognition (PR). 44(9), September 2011, 2184-2196. DOI=10.1016/j.patcog.2011.02.021 http://dx.doi.org/10.1016/j.patcog.2011.02.02.

Articles conference proceedings with a reading committee

1. Rachid Hedjam and Mohamed Cheriet. Ground-truth estimation in multispectral representation space: application de historical document image binarization. ICDAR'13. Washington DC, USA, 25-28, August 2013 (accepted).

2. Rachid Hedjam and Mohamed Cheriet. Hyperspectral Band Selection Based on Graph Clustering, Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on. Montreal, Canada, 5-7, Jul. 2012. (Best conference paper).

3. Rachid Hedjam and Mohamed Cheriet. Novel Data Representation for Text Extraction from Multispectral Historical Document Images. Document Analysis and Recognition (ICDAR), 2011 International Conference on. Beijing, China, pp. 18-21 Sept. 2011; doi: 10.1109/ICDAR.2011.43 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6065298&isnumber=6065247

4. Rachid Hedjam and Mohamed Cheriet. Combining statistical and geometrical classifiers for text extraction in multispectral document images. In Proceedings of the 2011 Workshop on Historical Document Imaging and Processing (HIP'11). ACM, New York, NY, USA, pp. 98-105. DOI=10.1145/2037342.2037359 http://doi.acm.org/10.1145/2037342.2037359.

5. Rachid Hedjam and Mohamed Cheriet. Segmentation-based document image denoising. Visual Information Processing (EUVIP), 2010 2nd European Workshop on pp. 61-65, 5-6 July 2010; doi: 10.1109/EUVIP.2010.5699134 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5699134&isnumber=5698737

6. Rachid Hedjam, Reza F. Moghaddam and Mohamed Cheriet. Text extraction from degraded document images. Visual Information Processing (EUVIP), 2010 2nd European Workshop on, pp.247-252, 5-6 July 2010; doi: 10.1109/EUVIP.2010.5699135 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5699135&isnumber=5698737

7. Rachid Hedjam and Max Mignotte. A hierarchical graph-based markovian clustering approach for the unsupervised segmentation of textured color images. Image Processing (ICIP), 2009 16th IEEE International Conference on, pp.1365-1368, Cairo, Egypt, 7-10 Nov. 2009; doi: 10.1109/ICIP.2009.5413555 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5413555&isnumber=5413332

8. Rachid Hedjam, Reza F. Moghaddam and Mohamed Cheriet. Markovian clustering for the non-local means image denoising," Image Processing (ICIP), 2009 16th IEEE International Conference on, pp.3877-3880, Cairo, Egypt, 7-10 Nov. 2009; doi: 10.1109/ICIP.2009.5414041 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5414041&isnumber=5413332

9. T. Abboud, R. Hedjam, R. Noumeir and A. Berinstain. Segmentation d'images de plantes capturées par un système d'imagerie fluorescente, 25 th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 29 avril-2 mai 2012. Montreal, Canada.

Internship

- Two-month (June-July 2010) internship at T2I laboratory at the University of Paris 13 with the aim to investigate the methods of measuring the perceptual quality of the digital images.

Awards

1. FQRNT, doctoral research scholarship.
2. École de Technologie Supérieure (ÉTS), Internal Scholarship.
3. Best conference oral paper. Hyperspectral Band Selection Based on Graph Clustering, Rachid Hedjam; Mohamed Cheriet (ISSPA'12, Montreal, Canada).
4. First rank in The international Competition on Qantitative Evaluation of Binarization Algorithms of Images of Historical Documents with Bleeding Noise (in ICFHR'10-International Conference on Frontiers in Handwriting Recognition, India 2010). Reza Farrahi Moghaddam, Rachid Hedjam and Mohamed Cheriet.
5. Honored organizing committee member in ISSPA'12.

Paper reviewing

- Elsevier, International Journal on Document Analysis and Recognition (2 papers).
- Elsevier, Digital Signal Processing Journal (1 paper).
- International Conference on Document Analysis and Recognition (3 papers).
- International Workshop on Document Analysis and System (1 paper).
- International Workshop on Historical Document Processing (2 papers).
- International Workshop on Signal Processing and their Applications, (5 papers).
- International Conference on Industrial Engineering and Manufacturing (1 paper).

ANNEX I

MS IMAGING SYSTEM, SET-UP AND ACQUISITION

We intend to produce, using our MS imaging system, a set of MS document images with ground-truth information about the foreground at each pixel that are freely accessible to the research community via a web site[1] . The data contain various types of degradation, and the documents were written at different periods of time. The MS images are based on historical documents collected from the BAnQ (Bibliothèque et Archives Nationales du Québec). The documents are written between the 17^{th} and 20^{th} centuries. Most 20^{th} century documents were machine printed, and so are less degraded than the earlier ones, which were handwritten with iron gall-based (ferrogallic) ink. The latter are highly degraded. The ink with which the majority of the historical documents stored in the BAnQ is made from iron salts and tannic acid from vegetable sources. It was the standard writing and drawing ink from about the 12th century to the 19th century, and remained in use well into the 20th century. It has a particular photometric characteristic when examined under infrared or UV light.

1 Material description

In this subsection we describe our acquisition system including setup, recording and calibration of the data. Briefly, the setup of our MS imaging system used in documents analysis is shown schematically in Fig. A I-1(a), where the document under investigation is placed on a support facing the camera from above. Two tunable lamps are used to illuminate the document, which are usually positioned at a 45^o angle, on the upper-left and upper-right sides of the camera, and close to the document, which allows to produce an ideal diffuse illumination where each area of the document is uniformly illuminated from all directions. The MS imaging system uses a set of 8 chromatic filters (motorized and controlled by the software of the camera) to produce 8 high resolution spectral images at specific wavelengths in ultraviolet (UV), visible (BLUE, GREEN, RED) and infrared (IR) as summarized in Table A-I-1. In addition to the filters, our MS imaging system is composed of a CCD camera, a lens in front of the camera. The camera's sensor is a **Chroma** X3 KAF 6303E (Kodak), shown in Figure A I-1(b), with a high quantum efficiency of 1.100 nm, is front-side illumination, and a resolution of 3072×2048 (6 megapixels) pixels of 9×9 microns. To ensure that this tool is non destructive (Klein *et al.*, 2008), a low intensity, wavelength tunable light source is generally used, with the temperature and relative humidity in the room maintained in an ambient state ($18^o - 20^o$), and the document not ex-

[1] http://www.synchromedia.ca/databases/HISTODOC2

posed to the light for an extended period. The MS imaging is entirely enclosed in a light-proof cabinet, in order to prevent any stray light from external sources affecting the measurements (Klein *et al.*, 2008).

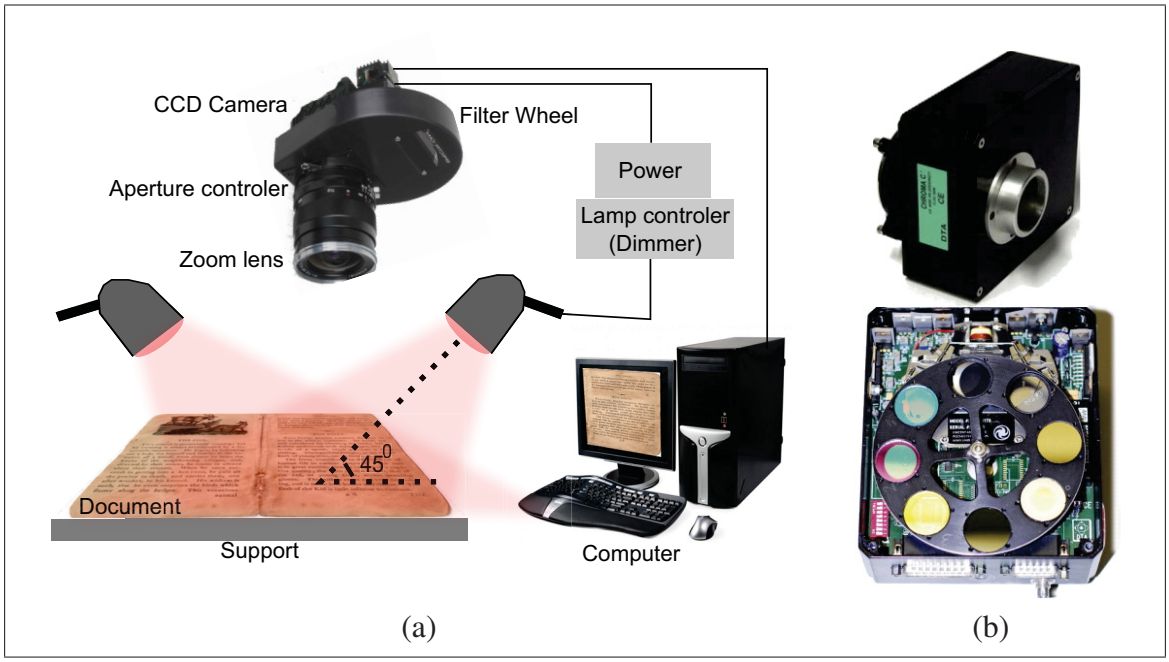


Figure-A I-1 The MSI setup (a); and the Chroma CX3 provided with filter-wheel (b). Image (b) is from (Bianco *et al.*, 2012)

During the exposure time, the incident photons reach the CCD elements (pixel sites) and then are converted into an electric charge. The latter are converted by an analogue-to-digital converter (ADC) into the digital pixel value to derive the spectral reflectance value of the area of the document under investigation. The number of electrons collected is linearly depends on light level and exposure time and non-linearly on wavelength. Above a certain threshold, the pixels will be saturated. One of the characteristic of our MS imager is that the spectral response (or QE: quantum efficiency) is very low for IR bands (almost 0 for 1100nm, see Fig. A I-2), to ensure a high Signal-to-Noise ration (SNR) for these spectral images, the pixel sites should collect as much light energy as possible without reaching the saturation level. To meet this condition, we must choose a proper exposure time which is in turn depends on the light level, the photometric proprieties of the object and on the current filter. Our system offers us the possibility to adjust the light via a dimmer and choose a proper exposure time for each filter. For our experiment, the exposure time (ET) for each filter is summarized in Table A I-1.

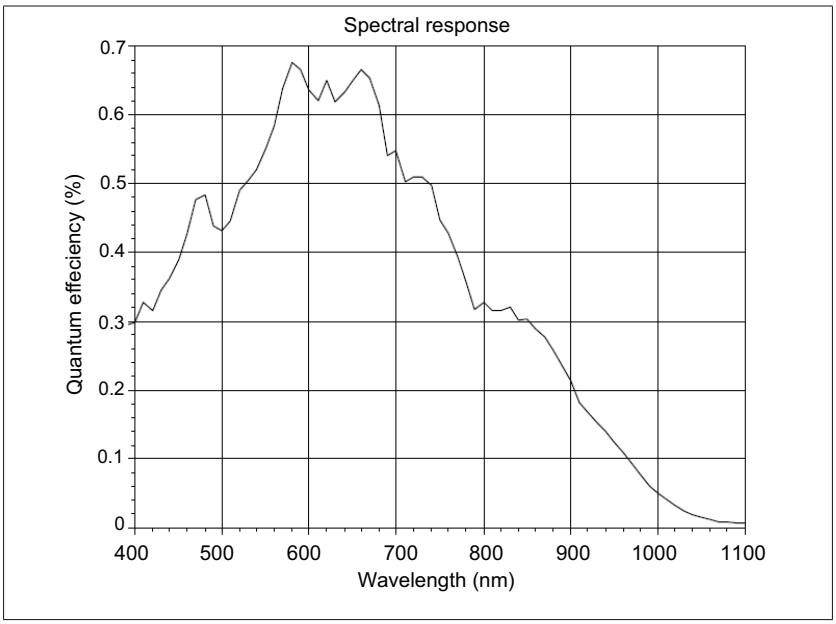


Figure-A I-2 Spectral response of Chroma X3 KAF 6303E.

2 Camera settings, data acquisition and calibration

First of all the camera should be connected properly and powered. The data acquisition is controlled by a computer. For the UV lighting, BlackLight BULB style bulbs are used. As for the VIS and IR lighting, the lamps are Quartz-Tungsten Halogen-based and emit a smooth light spectrum range from visible to infrared. The shooting (acquisition) is influenced by different factors and parameters such as Exposure time (ET), the CCD temperature, the camera gain, the focusing of the target and so on. It is not very easy to find out the right combination of the values to set for all these factors in order to obtain satisfying images. To reach the better conditions of the shooting, preliminary modifications on the different values is made.

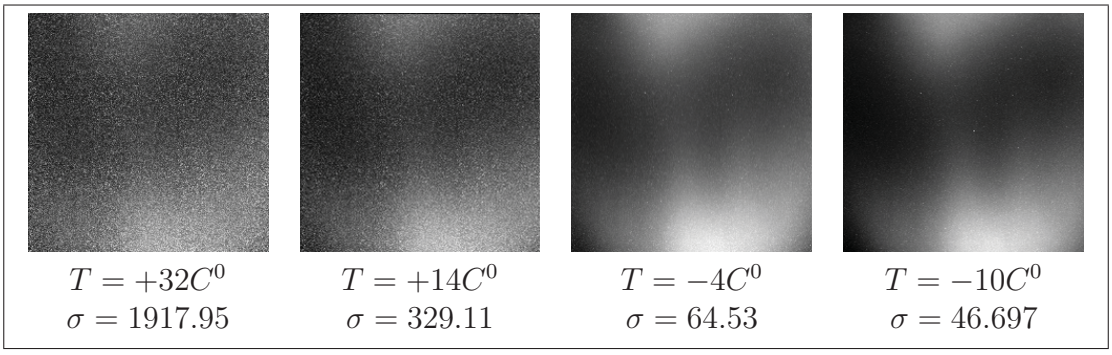
The **temperature** is the main source of Dark Current noise. Figure A I-3 shows the effect of the temperature on the generation of noise. The standard deviation (σ) confirm the previous considerations. Cooling the CCD sensor give the advantage to reduce the factor of the temperature, consequently increasing the SNR. The Exposure time (**ET**) must to be set carefully. Long ET is enough to have a significant signal but not so much to avoid the saturation conditions of the pixels. The **gain**, which represents the number of electrons necessary to create a voltage relative to 1 ADU (analogue to digital unit) is set to 12 e/ADU. The **focusing** allows to find the target (whole object or Region Of Interest, ROI) in the field of view (FoV) of the recorded image. Because our MS imaging system uses several filters with different lens thickness and

Tableau-A I-1 Chroma X3 filter description and acquisition set-up.

		Parameter set-up			
Filter	Wavelength(nm)	ET(ms)	Dimer (position)	Focus	Aperture
F1 (UV)	400	100		71	f/8
F2 (BLUE)	500	50	1	71	f/8
F3 (GREEN)	600	50	2.5	71	f/8
F4 (RED)	700	50	3.5	71	f/8
F5 (IR)	800	20	3	70	f/8
F6 (IR1)	900	20	3	71	f/8
F7 (IR2)	1000	100	3	70	f/8
F8 (IR3)	1100	200	1.5	69	f/8

is not equipped with Auto-Focus, the adjustment of lens, for each filter switch, must be done manually and verify if the image is in-focus. A number are engraved on the lens barrel like 60, 69, 70, 71 and so one.

Figure-A I-3 Effect of the temperature on the generation of noise



The final parameter is the **aperture** which is referred to the lens diaphragm opening that regulates amount of light passes through the camera lens. The aperture size is generally specified in terms an *f-stop* values engraved on the lens barre like f/1.0, f/1.4, f/2.0,... and so on (see Figure A I-4). For a given ET, the smaller the aperture the darker the image. For our experiment, the aperture is fixed to f/8.0 for every filter. Table A I-1 summarizes the parameter set-up of our imaging system.

Calibrating the recorder spectral images must be done firstly before any subsequent processing to obtain the true spectral reflectance values for each pixel location on the document (see Fig. A I-5). This can be achieved by recording two frame images under the same recording conditions

(ie. same parameters): dark frame image (D) and white reference frame (S). The dark frame is measured by capturing a dark image by blocking the camera by using a cap; while the white reference frame is characterized simply by acquiring a light image of a homogeneous white surface. The spectral reflectance value at each location pixel, x in a specific wavelength, λ is given by:

$$u(x,\omega) = \mu_h(\omega)\frac{t_h}{t_o}\frac{O(x,\omega,t_o) - D(x,t_o)}{S(x,\omega,t_s) - D(x,t_s)} \qquad \text{(A I-1)}$$

where O is the recorded imperfect spectral image, t_o and t_s are the exposure times used respectively for the spectral image of the object (document) and the spectral image of the white reference frame. Experimentally, $t_o = t_s$. $\mu_h(\omega)$ is the intensity value of the white reference frame. In our work we consider it as the mean value of S as indicated by the documentation of our MS imaging system. For more detail, the reader can refer to (Klein *et al.*, 2008).

3 Registration

Another problem is the chromatic aberration, which is explained by the geometric difference (rotation, translation, etc.) between the spectral images acquired by the various filters. In any MS imaging system, the filters are of different thicknesses, and the refractive index of the glass

Figure-A I-4 Maximum and minimum camera aperture.

depends on the wavelength of each filter. This makes a difference in the acquisition geometry of the filters, and affects further analysis of the document image, because even a simple pixel-to-pixel comparison of two spectral images will fail. An example of this problem is shown in Fig. A I-6(a). In order to compensate for this problem, and to align all the spectral images as much as possible, we use the i2KAlign software, developed by DualAlign LLC ™ [2].

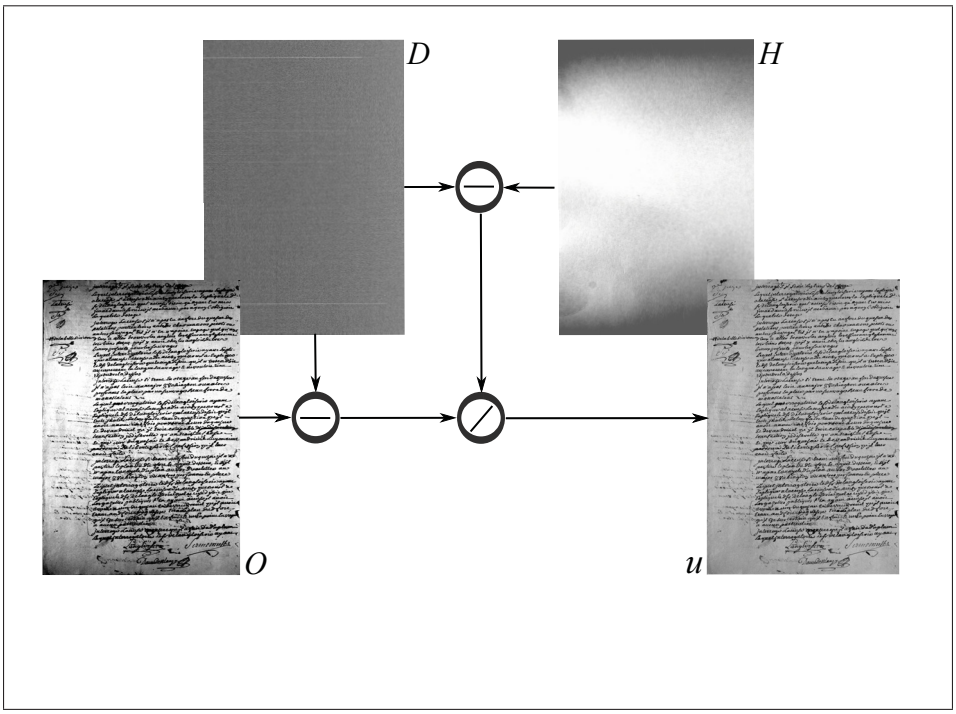


Figure-A I-5 Spectral image calibration procedure.

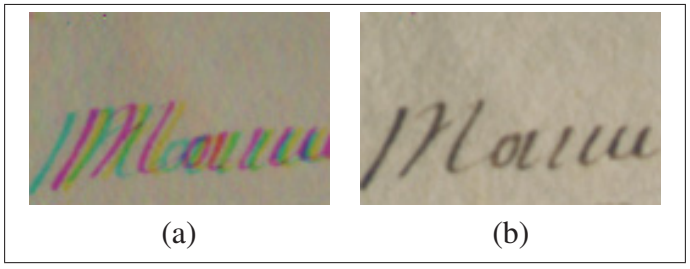


Figure-A I-6 Chromatic aberration correction; (a) aberration effect on a document image combined from BLUE, GREEN and RED channels, (b) after chromatic aberration correction by alignment.

[2] http://www.dualalign.com/

4 Mathematical model of the MS document image

Mathematically speaking, a MS image (see Figure A I-7 (a)) is described as follows: $u(s)$, where $s = [x, y]^\top \in \Omega \subset \mathbb{R}^B$. Ω is the domain of the MS image. Each pixel s is characterized by B independent spectral values (or spectral reflectance denoted by $u(s)$) represented by a vector of B components (for our MS imaging system $B = 8$). The MS data set is then embedded in an B-dimensional vector space and the spectral signature of a pixel corresponds to a particular location in this space. The spectral signature (see Figure A I-7 (b)) of a pixel x can be denoted

$$u(s) = (u_1(s), u_2(s), ..., u_B(s)) \tag{A I-2}$$

and then used as a feature vector in subsequent treatment procedures.

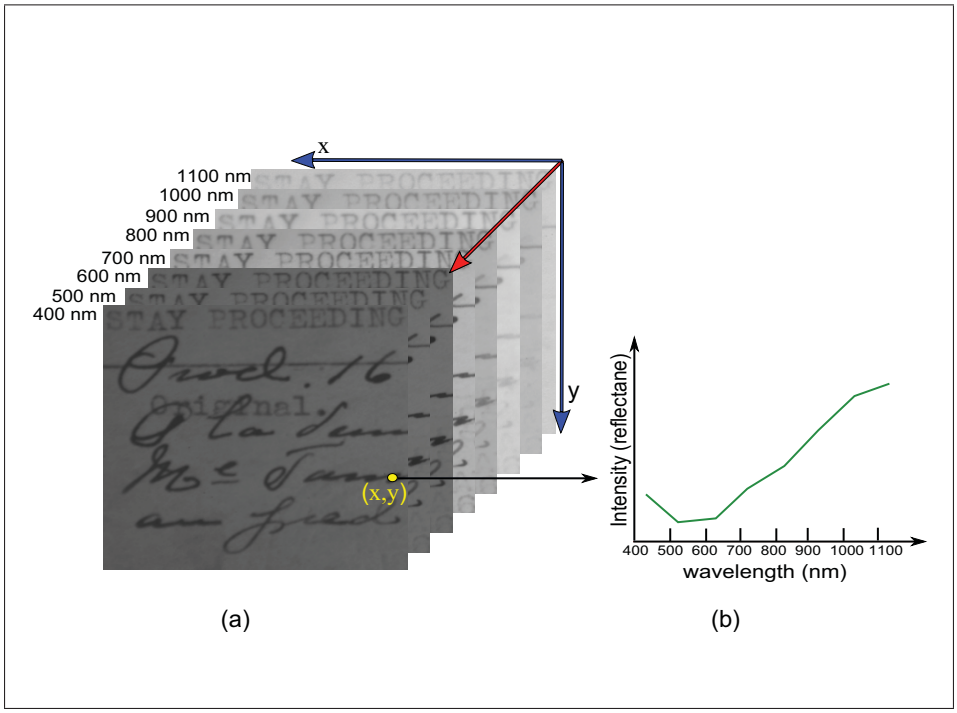


Figure-A I-7 MS document image is shown in (a), and the spectral signature of the pixel s at location (x, y) is shown in (b).

5 Spectral data analysis

One interesting characteristic of the documents collected from the BAnQ is that the main (original) text is written with ink based on iron gall material and doesn't contain much carbon. Carbon absorbs IR radiation and reflects UV radiation, while iron transmits IR radiation, which is then reflected by the underlying constituents. This explains why the main text disappears from the spectral images when we move towards longer wavelengths (see Figure A I-8). In other words, the IR region makes iron gall ink transparent. In contrast, the annotations are written with pencil containing carbon, which allows them to be visible in the IR regions.

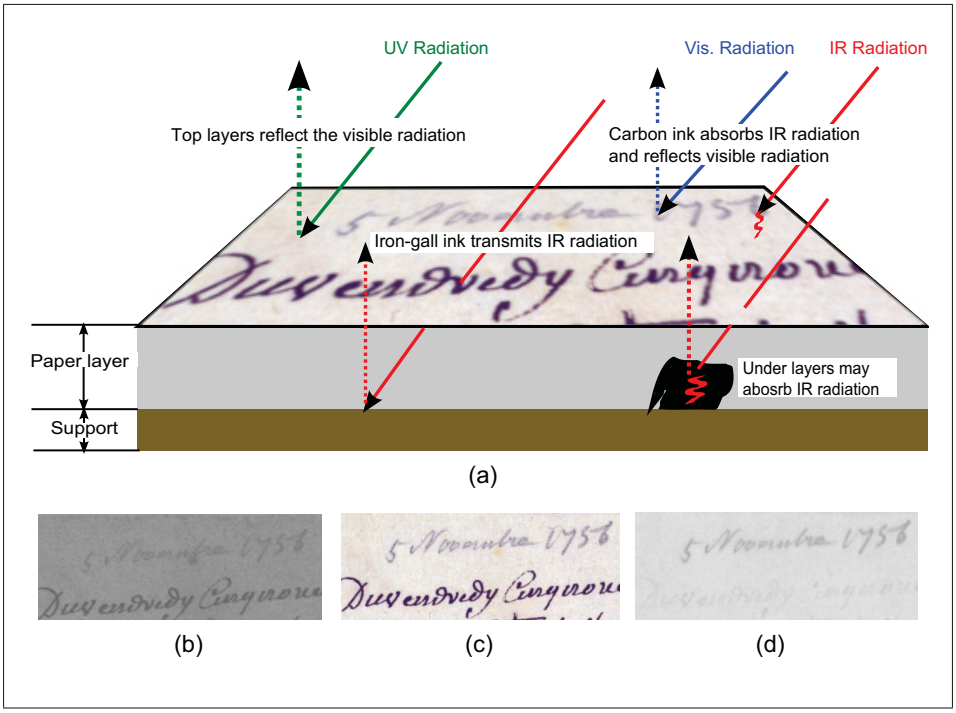


Figure-A I-8 Difference between iron-gall ink (original text) and carbon based ink (annotation). (a) illustration of the photometric propoerties of the document content. (b)UV spectral image, (c) visible spectral image, (d) IR spectral image. It is clear that the annotation (written in french: *5 novembre 1756*) stills visible at IR spectral image while the original text disappears

Now, we aim to explain briefly the spectral reflectance of the main element to be studied which is the original ink used to writing. The spectral analysis of the data provide the reader with important information about the different document image objects. Fig. A I-9 shows the intensity distribution of three samples of pixels picked manually from three different objects: original text, stamp, and background. The histogram of each sample at different wavelengths represents

the tonal distribution, which can then be used as a feature for object separation (Hedjam *et al.*, 2010; Hedjam and Cheriet, 2011b,a; Hedjam *et al.*, 2011b).

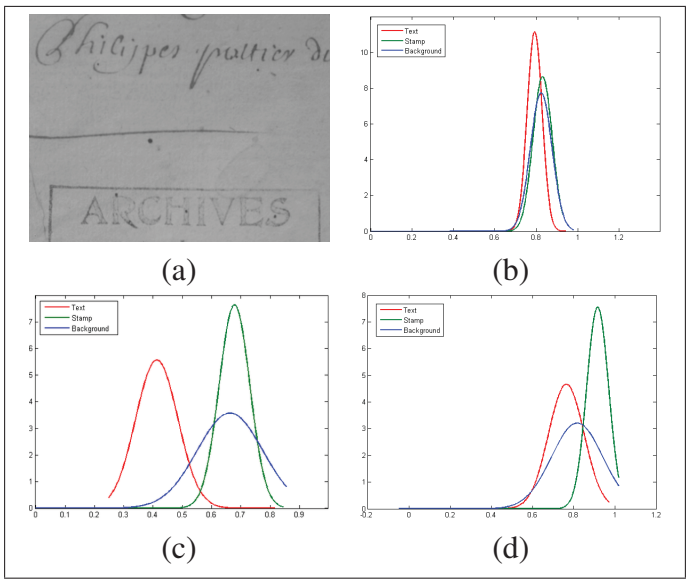


Figure-A I-9 Histograms of three samples of pixels picked from (a): histograms of the text sample are plotted with red color, the histograms of background sample are plotted with blue color and those of stamp sample are plotted with green color. (c) at GREEN wavelength, the histogram of the text sample is well separated from others; (d) histogram of the stamp sample is bit separated from others at the IR wavelength; (b) unfortunately, all three histograms are overlapped at UV wavelength.

Even though the text is clearly differentiable from the other objects as shown in Fig. A I-9 for example, this alone is not sufficient to make a definitive determination. One of the main reasons for the instability of the results is that the photometric properties of the original text are influenced by the volumetric concentration of the ink flows in each text zone. Whenever the volumetric concentration is not uniform throughout the text, the distribution of its pixel intensities is multi-modal, and vice versa. Fig. A I-10 shows the intensity distribution of iron gall text pixels in two different document images examined at 750 nm (IR wavelength). These pixels are selected using a semi-automated process, as follows: one of the spectral images in which the text is clearly apparent is binarized (in general, the green band is an appropriate choice); then, the possible artefacts produced in the final binary image are manually removed, leaving only the original text pixels; and, finally, a morphological thinning process is performed on these pixels to leave the relevant ones (see Fig. A I-10(b,e)). This methodology ensures the selection of pixels with higher and lower volumetric concentrations of ink. However, using a

Gaussian mixture function in its accurate representation, which allows the decomposition of the intensities found in the iron -gall ink, we can show that the intensity distribution is more multi-modal when the ink volumetric concentration varies over the text pixels, as shown in Fig. A I-10(f).

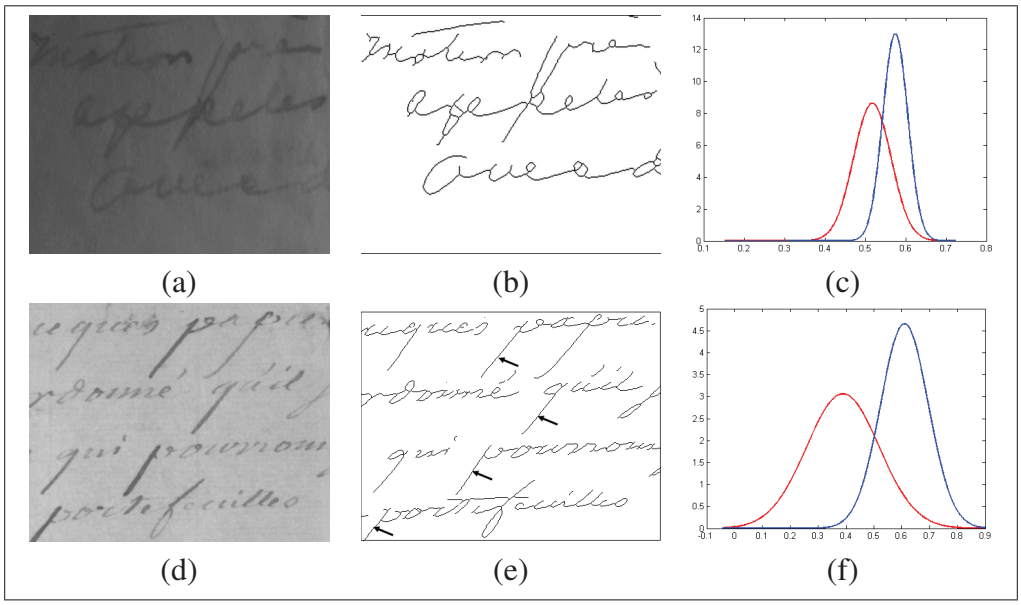


Figure-A I-10 Examples of Gaussian mixture models of iron-gall ink at IR wavelength. The image in second line presents more variability (non-uniformity) in ink volume concentration, thus, the corresponding text intensity distribution is more bimodal.

ANNEX II

AUTOMATIC FINDING OF THE THRESHOLD τ.

Automatic finding of the threshold τ.

As noted before, the accurate IR channel can have three kinds of intensity variance. The first variance, σ_b, is related to the intensity change in non-degraded background areas, the second variance, σ_s, is related to the intensity change in slight degraded areas and the third one, σ_h, is related to the intensity change in strong degraded areas. It is worth noting that in general, $\sigma_h > \sigma_s > \sigma_b \simeq 0$. The correction of slight degradations consists in deciding whether the local variance, calculated on a neighborhood around each pixel, is greater than a threshold τ or not. This threshold can be learned from the histogram of local variances distribution. Our hypothesis is that, the non-degraded background pixels are more dominant than those belonging to the degraded areas. This can be shown by the proportion of counts around the highest peak in the histogram of local variances distribution (see Fig. A II-1). To avoid changing the intensity of the non degraded areas, by local correction of slight degradations, the threshold τ, can, so, be the variance corresponding to the first peak following the highest peak. Fig. A II-1 shows two histograms of the distributions of local variances of three accurate IR channels, taken randomly from the images we have. It was found that the threshold, τ, is approximately close to 0.015 in most cases.

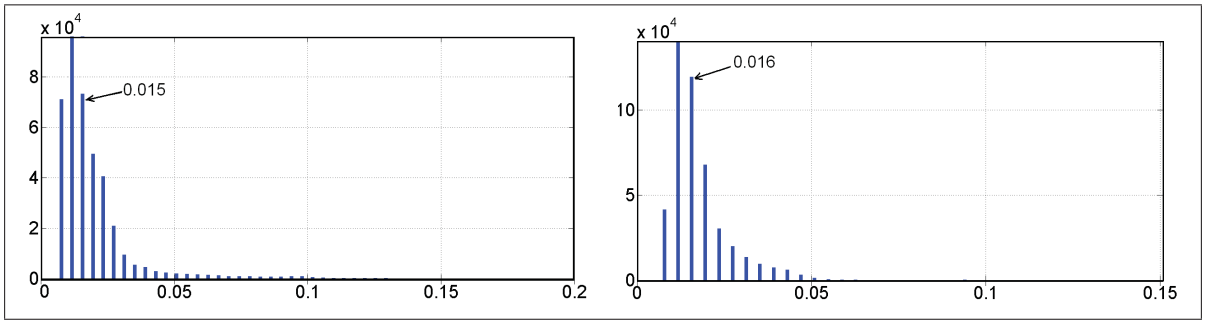


Figure-A II-1 Unsupervised estimation of the threshold τ.

ANNEX III

EXPERIMENTAL SET-UP FOR IRR, UVR AND UVF IMAGING TECHNIQUES.

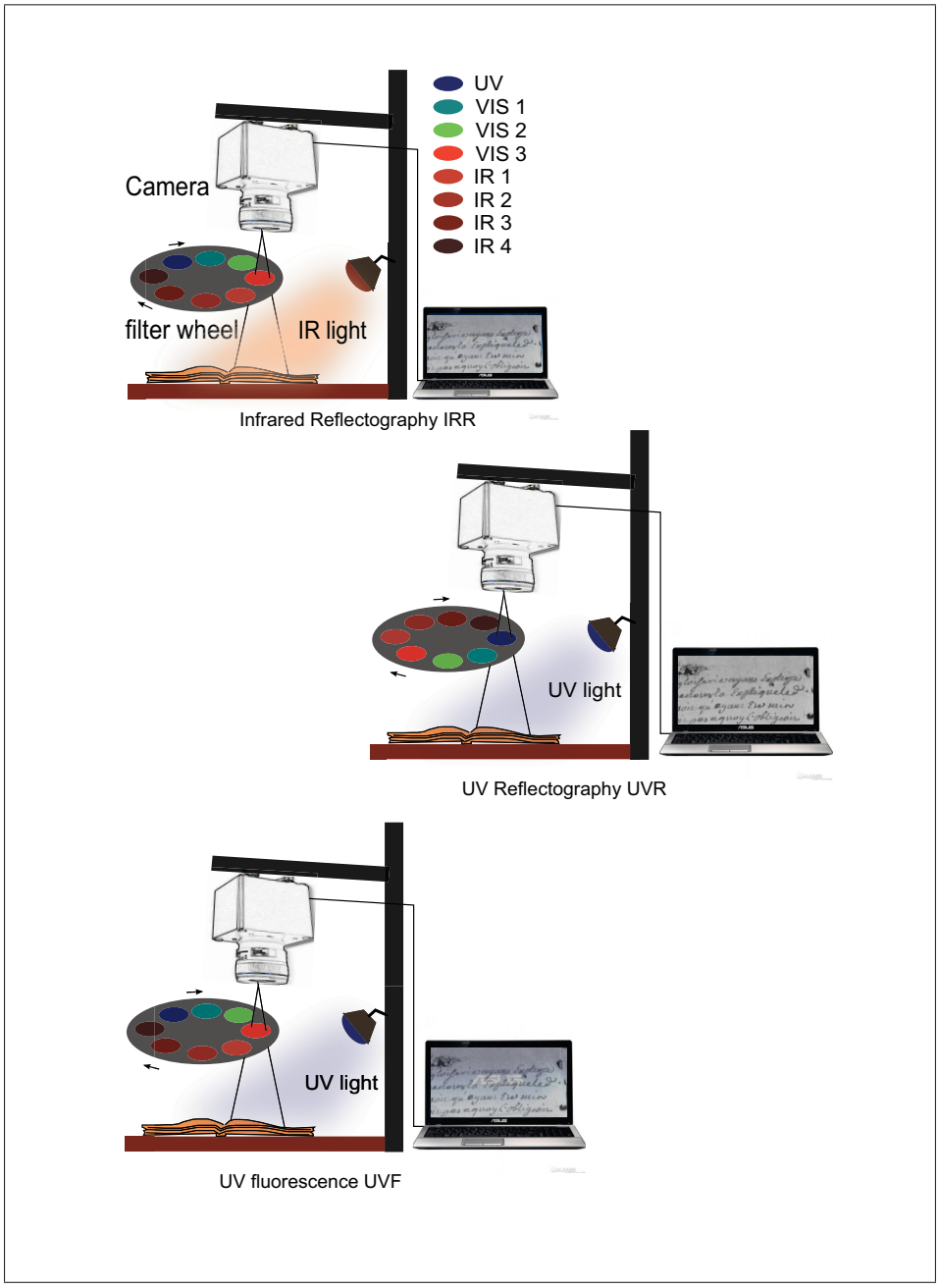


Figure-A III-1 Experimental set-up for IRR, UVR and UVF imaging techniques.

BIBLIOGRAPHY

Juma Al Majid center for Culture and Heritage. 55106, Dubai, UAE. <http://www.almajidcenter.org/>.

Abutaleb, Ahmed S. July 1989a. "Automatic thresholding of gray-level pictures using two-dimensional entropy". *Computer Vision, Graphics, and Image Processing*, vol. 47, n° 1, p. 22–32.

Abutaleb, Ahmed S. 1989b. "Automatic thresholding of gray-level pictures using two-dimensional entropy". *Computer Vision, Graphics, and Image Processing*, vol. 47, n° 1, p. 22-32.

andJ. Bartl, M. Hain and V. Jacko. 2003. "Multispectral analysis of cultural heritage artefacts". *Measurement Science Review*, vol. 3, n° 3, p. 9-12.

Antonacopoulos, A. and D. Karatzas. 2004. "Document image analysis for World War II personal records". In *Document Image Analysis for Libraries, 2004. Proceedings. First International Workshop on*. p. 336 - 341.

Antonacopoulos, Apostolos and Andy Downton. April 2007. "Special issue on the analysis of historical documents". *IJDAR*, vol. 9, n° 2, p. 75–77.

Backer, Marina V, Zoya Levashova, Vimalkumar Patel, Brian T Jehning, Kevin Claffey, Francis G Blankenberg, and Joseph M Backer. 11 March 2007. "Molecular imaging of VEGF receptors in angiogenic vasculature with single-chain VEGF-based probes". *Nature Medicine*, vol. 13, p. 504–509.

Baird, H. 2000. "The State of the Art of Document Image Degradation Modeling". In *Proc. of 4 th IAPR International Workshop on Document Analysis Systems*. (Rio de Janeiro, Brazil 2000), p. 1–16.

Bar-Yosef, Itay, Alik Mokeichev, Klara Kedem, Itshak Dinstein, and Uri Ehrlich. 2009. "Adaptive shape prior for recognition and variational segmentation of degraded historical characters". *Pattern Recogn.*, vol. 42, n° 12, p. 3348–3354.

Bei, Ling, Glenn I. Dennis, Heather M. Miller, Thomas W. Spaine, and Jon W. Carnahan. 2004. "Acousto-optic tunable filters: fundamentals and applications as applied to chemical analysis techniques". *Progress in Quantum Electronics, Elsevier*, vol. 28, n° 2, p. 67-87.

Bernsen, J. 1986. "Dynamic thresholding of grey-level image". In *Eighth International Conference on Pattern Recognition*.

Bertalmio, Marcelo, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. 2000. "Image inpainting". In *ACM SIGGRAPH'2000*. p. 417–424. ACM Press/Addison-Wesley Publishing Co.

Besag, Julian. Sep. 1975. “Statistical Analysis of Non-Lattice Data”. *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 24, n° 3, p. 179–195.

Bianco, Gianfranco, Fabio Bruno, and Maurizio Muzzupappa. 2012. “Multispectral data cube acquisition of aligned images for document analysis by means of a filter-wheel camera provided with focus control”. *Cultural heritage*.

Bouali, Marouan and Said Ladjal. Spectral inpainting for the restoration of missing data from multispectral satellite sensors: Case study on aqua modis band 6. *Archives ouvertes HAL*. Institut Télécom. http://hal-institut-telecom.archives-ouvertes.fr/hal-00639083.

Boussellaa, Wafa, Abderrazak Zahour, and Adel Alimi. 2007. “A methodology for the separation of foreground/background in Arabic historical manuscripts using hybrid methods”. In *Proceedings of the 2007 ACM symposium on Applied computing*. (Seoul, Korea 2007), p. 605–609. ACM.

Boykov, Yuri and Gareth Funka-Lea. 2006. “Graph Cuts and Efficient N-D Image Segmentation”. *International Journal of Computer Vision*, vol. 70, p. 109-131.

Breslow, L. A. and D. Aha. Jan. 1997. “Simplifying decision trees: A survey”. *Knowledge Engineering Review*, vol. 12, n° 1, p. 1–40.

Brettel, Hans, Jon Yngve Hardeberg, and Francis Schmitt. 1999. “Multispectral Image Capture Across the Web”. In *in Proceedings of IS&T and SID?s 7th Color Imaging Conference: Color Science, Systems and Applications*.

Brodley, Carla E. and Mark A. Friedl. 1999. “Identifying Mislabeled Training Data”. *journal of Artificial Intelligence Research*, vol. 11, p. 131–167.

Calpe-Maravilla, Javier, Joan Vila-Frances, Emilio Ribes-Gomez, Vicente Duran-Bosch, Jordi Munoz-Mari, Julia Amoros-Lopez, Luis Gomez-Chova, and Enrique Tajahuerce-Romera. 2004. “400- to 1000-nm imaging spectrometer based on acousto-optic tunable filters”. p. 460-471.

Casini, A. 2002. “Fourier transform interferometric imaging spectroscopy: a new tool for the study of reflectance and fluorescence of polychrome surfaces”. *Conservation Science Proceedings 2002, 249-253.*, p. 249-253.

Cha, Sung-Hyuk and Sargur N Srihari. 2002. “On measuring the distance between histograms”. *Pattern Recognition*, vol. 35, n° 6, p. 1355–1370.

Chalmers, John M., Howell G. M. Edwards, and Michael D. Hargreaves, 2012. *Infrared and Raman Spectroscopy in Forensic Science*. John Wiley and Sons.

Chan, Tony F. and Jianhong Shen. 2002a. “Mathematical Models for Local Nontexture Inpaintings”. *SIAM J. Appl. Math*, vol. 62, p. 1019–1043.

Chan, Tony F. and Jianhong Shen. 2002b. “On the Role of the BV Image Model in Image Restoration”.

Chang, C-I, 2003. *Hyperspectral Imaging*. Plenum Publishers.

Chang, Chein-I. aug 2000. " An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis ". *Information Theory, IEEE Transactions on*, vol. 46, n° 5, p. 1927 -1932.

Chen, Q., Q-S. Sun, P.A. Heng, and D-S. Xia. 2008. " A double thresholding image binarization method based on edge detector ". *Pattern recognition*, vol. 41.

Chen, Y., M. K. Mihcak, and D. Kirovski. *Certifying authenticity via fiber-infused paper*. Technical report. Microsoft Research.

Cheng, H. D., X. H. Jiang, Y. Sun, and Jing Li Wang. 2001. " Color image segmentation: Advances and prospects ". *Pattern Recognition*, vol. 34, p. 2259–2281.

Cheriet, M. and R. Farrahi Moghaddam. Auguest 25-29 2008a. " Low Quality Image Processing for DIAR. Issues and Directions ". In *EUSIPCO'08*. (Laussane, Switzerland 2008).

Cheriet, M. and R. Farrahi Moghaddam. 2008b. " Degradation Modeling and Enhancement of Low Quality documents ". In *WOSPA'2008*. (Sharjah, UAE, Invited paper 2008).

Cheriet, M. and R. Farrahi Moghaddam. 2008c. Diar: Advances in degradation modelling and processing. *Lecture Notes in Computer Science: ICIAR'2008*, volume 5112/2008, p. 1–10. Póvoa de Varzim, Portugal, Invited paper. doi: 10.1007/978-3-540-69812-8_1.

Cheriet, M., J.N. Said, and C.Y. Suen. 1998. " A recursive thresholding technique for image segmentation ". *IEEE Transactions on Image Processing*, vol. 7, n° 6, p. 918–921.

Cheriet, Mohamed, Reza Farrahi Moghaddam, and Rachid Hedjam. 2012. " A learning framework for the optimization and automation of document binarization methods ". *Computer Vision and Image Understanding*.

Ciurea, F. and B. Funt. Nov. 2003. " A Large Image Database for Color Constancy Research ". In *roceedings of the Imaging Science and Technology Eleventh Color Imaging Conference,*. (Scottsdale 2003), p. 160-164.

Comelli, Daniela, Gianluca Valentini, Austin Nevin, Andrea Farina, Lucia Toniolo, and Rinaldo Cubeddu. August 2008. " A portable UV-fluorescence multispectral imaging system for the analysis of painted surfaces ". *Rev. Sci. Instrum.*, vol. 79, n° 8, p. 086112–3.

Cortés, Alejandro Ribés. 2003. " Analyse multispectrale et reconstruction de la réflectance spectrale de tableaux de maitre ". PhD thesis, Ecole Nationale Supérieure des Télécommunications, France.

Couasnon, B., J. Camillerapp, and I. Leplumey. 2004. " Making handwritten archives documents accessible to public with a generic system of document image analysis ". In *Document Image Analysis for Libraries, 2004. Proceedings. First International Workshop on*. p. 270 - 277.

da Silva, Joao Marcelo Monte, Rafael Dueire Lins, Fernando Mario Junqueira Martins, and Rosita Wachenchauzer. January 2008. " A New and Efficient Algorithm to Binarize Document Images Removing Back-to-Front Interference ". *Journal of Universal Computer Science*, vol. 14, n° 2, p. 299–313.

de Campos, T. E. 2006. " 3D Visual Tracking of Articulated Objects ". PhD thesis, Department of Engineering Science - University of Oxford.

Deriche, R. and O. Faugeras. 1996. *Les EDP en traitement des images et vision par ordinateur*. Technical Report 2697. INRIA.

Drira, F. December 2007. " Contribution à la Restauration des Images de Documents Anciens ". PhD thesis, École Doctorale Informatique et Information pour la Société (EDIIS), LIRIS, UMR 5205 CNRS.

Drira, Fadoua, Frank Le Bourgeois, and Hubert Emptoz. 2006. " Restoring Ink Bleed-Through Degraded Document Images Using a Recursive Unsupervised Classification Technique ". *Document Analysis Systems VII*, p. 38–49.

Du, Hao, Xin Tong, Xun Cao, and Stephen Lin. 2009. " A prism-based system for multispectral video acquisition ". In *ICCV*. p. 175-182.

Dubois, E. and P. Dano. April 2005. " Joint compression and restoration of documents with bleed-through ". In *Proc. IS&T Archiving 2005*. (Washington DC, USA 2005), p. 170–174.

Dubois, E. and A. Pathak. April 2001. " Reduction of bleed-through in scanned manuscript documents ". In *Proc. IS&T Image Processing, Image Quality, Image Capture Systems Conference (PICS2001)*. (Montreal, Canada 2001), p. 177–180.

Duda, Richard O., Peter E. Hart, and David G. Stork, 2001. *Pattern Classification*. ed. 2^{nd}. Wiley.

Easton,R.L., Jr., K.T. Knox, and W.A. Christens-Barry. 2003. " Multispectral imaging of the Archimedes palimpsest ". In *Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. 32nd*. p. 111 - 116.

et al., M. Attas. 2003. " NearInfrared spectroscopic imagine in art conservation: investigation of drawing constituents ". *Journal of cultural heritage*, vol. 4, n° 2, p. 127-136.

Fabbri, Ricardo, Luciano Da F. Costa, Julio C. Torelli, and Odemir M. Bruno. 2008. " 2D Euclidean distance transform algorithms: A comparative survey ". *ACM Comput. Surv.*, vol. 40, n° 1, p. 1–44.

Farrahi Moghaddam, Reza and Mohamed Cheriet. 2009a. " Low quality document image modeling and enhancement ". *IJDAR*, vol. 11, n° 4, p. 183–201.

Farrahi Moghaddam, Reza and Mohamed Cheriet. 2009b. " RSLDI: Restoration of single-sided low-quality document images ". *Pattern Recognition*, vol. 42, p. 3355–3364.

Farrahi Moghaddam, Reza and Mohamed Cheriet. July 26–29 2009c. “ Application of Multi-level Classifiers and Clustering for Automatic Word-spotting in Historical Document Images ”. In *ICDAR’09*. (Barcelona, Spain 2009), p. 511–515.

Farrahi Moghaddam, Reza and Mohamed Cheriet. 2010a. “ A Variational Approach to Degraded Document Enhancement ”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, n° 8, p. 1347–1361.

Farrahi Moghaddam, Reza and Mohamed Cheriet. June 2010b. “ A multi-scale framework for adaptive binarization of degraded document images ”. *Pattern Recognition*, vol. 43, n° 6, p. 2186–2198.

Farrahi Moghaddam, Reza and Mohamed Cheriet. 2012. “ AdOtsu: An adaptive and parameterless generalization of Otsu’s method for document image binarization ”. *Pattern Recognition*, vol. 45, n° 6, p. 2419–2431.

Farrahi Moghaddam, Reza, David Rivest-Hénault, and Mohamed Cheriet. July 26–9 2009. “ Restoration and segmentation of highly degraded characters using a shape-independent level set approach and multi-level classifiers ”. In *ICDAR’09*. (Barcelona, Spain 2009), p. 828–832.

Faubel, Werner, Susanne Staub, Rolf Simon, Stefan Heissler, Andrea Pataki, and Gerhard Banik. July 2007. “ Non-destructive analysis for the investigation of decomposition phenomena of historical manuscripts and prints ”. *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 62, n° 6-7, p. 669–676.

Gatos, B., I. Pratikakis, and S.J. Perantonis. March 2006. “ Adaptive degraded document image binarization ”. *Pattern Recognition*, vol. 39, n° 3, p. 317–327.

Gatos, B., I. Pratikakis, and S.J. Perantonis. 2008. “ improved document image binarization by using a combination of multiple binarization techniques and adapted edge information ”. In *ICPR’08*. p. 1-4.

Gatos, B., K. Ntirogiannis, and I. Pratikakis. 2009a. “ ICDAR 2009 Document Image Binarization Contest (DIBCO 2009) ”. In *ICDAR’09*. p. 1375–1382.

Gatos, B., K. Ntirogiannis, and I. Pratikakis. 2009b. “ DIBCO 2009: document image binarization contest ”. *IJDAR*, p. –.

Gatos, Basilios, Ioannis Pratikakis, and Stavros J. Perantonis. 2004. An adaptive binarization technique for low quality historical documents. *Lecture Notes in Computer Science :Document Analysis Systems VI*, volume 3163, p. 102–113. Springer.

Geman, Stuart and Donald Geman. 1984. “ Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images ”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, p. 721-741.

Ghoggali, N. and F. Melgani. july 2009. “Automatic Ground-Truth Validation With Genetic Algorithms for Multispectral Image Classification”. *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, n° 7, p. 2172 -2181.

Goltz, Douglas, Michael Attas, Gregory Young, Edward Cloutis, and Maria Bedynski. 2010. “Assessing stains on historical documents using hyperspectral imaging”. *Journal of Cultural Heritage*, vol. 11, n° 1, p. 19–26.

Google, 2007. *Book Search Dataset*.

Hardin, R. W. Nov. 1999. *Optical tricks designed to foil counterfeiters*. OE Reports 191. International Society for Optical Engineering.

Havermans, J., H. Abdul Aziz, and H. Scholten. 2003a. “Non destructive detection of iron-gall inks by means of multispectral imaging. Part 2: Application on original objects affected with iron-gall-ink corrosion”. *Restaurator International Journal For The Preservation Of Library And Archival Material*, vol. 24, n° 2, p. 88–94.

Havermans, J., H. Abdul Aziz, and S. Scholten. 2003b. “Non destructive detection of iron-gall inks by means of multispectral imaging. Part 2: application on original objects affected with iron-gall-ink corrosion”. *Restaurator: international journal for the preservation of library and archival material*, vol. 24, n° 2, p. 88-94.

Hedjam, Rachid and Mohamed Cheriet. 2011a. “Combining statistical and geometrical classifiers for text extraction in multispectral document images”. In *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*. p. 98–105.

Hedjam, Rachid and Mohamed Cheriet. 2011b. “Novel Data Representation for Text Extraction from Multispectral Historical Document Images”. In *ICDAR*. p. 172-176.

Hedjam, Rachid, Reza Farrahi Moghadam, and Mohamed Cheriet. July 5–7 2010. “Text extraction from degraded document images”. In *EUVIP’10*. (Paris, France 2010), p. 248-253.

Hedjam, Rachid, Reza Farrahi Moghaddam, and Mohamed Cheriet. 2011a. “A spatially adaptive statistical method for the binarization of historical manuscripts and degraded document images”. *Pattern Recognition*, vol. 44, n° 9, p. 2184-2196.

Hedjam, Rachid, Reza Farrahi Moghaddam, and Mohamed Cheriet. 2011b. “A spatially adaptive statistical method for the binarization of historical manuscripts and degraded document images”. *Pattern Recognition*, vol. 44, n° 9, p. 2184-2196.

Ifarraguerri, Agustin and Chein-I Chang. 2, March 1999. “Multispectral and Hyperspectral Image Analysis with Convex Cones”. *IEEE trans. on geoscience and remote sensing*, vol. 37, n° 2, p. 756–770.

Jia, Xiuping and J.A. Richards. 1994. “Efficient maximum likelihood classification for imaging spectrometer data sets”. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, n° 2, p. 274–281.

Jiang, Xiaomo and Sankaran Mahadevan. 2007. “Bayesian risk-based decision method for model validation under uncertainty”. *Reliability Engineering & System Safety*, vol. 92, n° 6, p. 707–718.

John, George H. 1995. “Robust Decision Trees: Removing Outliers from Databases”. In *In Knowledge Discovery and Data Mining*. p. 174–179. AAAI Press.

Joo Kim, Seon, Fanbo Deng, and Michael S. Brown. July 2011. “Visual enhancement of old documents with hyperspectral imaging”. *Pattern Recogn.*, vol. 44, p. 1461–1469.

Joosten, Ineke. June 2008. “Applications of microanalysis in the cultural heritage field”. *Microchimica Acta*, vol. 161, n° 3, p. 295–299.

Kaikova, Helen and Vagan Terziyan. 1997. “Temporal Knowledge Acquisition From Multiple Experts”. In *In: Shoval P. & Silberschatz A. (Eds.), Proceedings of NGITS'97, Neve Ilan, June 30 - July 3*. p. 44–55.

Kailath, T. february 1967. “The Divergence and Bhattacharyya Distance Measures in Signal Selection”. *Communication Technology, IEEE Transactions on*, vol. 15, n° 1, p. 52 -60.

Kalacska, Margaret and G. Arturo Sanchez-Azofeifa, Feb. 26 2008. *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests*. ed. 1. CRC Press.

Kaplan, Wilfred, 2002. *Advanced Calculus*. Addison Wesley.

Kapur, J. N., Prasanna K. Sahoo, and A. K. C. Wong. 1985. “A new method for gray-level picture thresholding using the entropy of the histogram”. *Computer Vision, Graphics, and Image Processing*, p. 273-285.

Kerekes, John P. and David A. Landgrebe. 1991. “An analytical model of Earth-observational remote sensing systems”. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, n° 1, p. 125-133.

Kim, In-Kwon, Dong-Wook Jung, and Rae-Hong Park. January 2002. “Document image binarization based on topographic analysis using a water flow model”. *Pattern Recognition*, vol. 35, n° 1, p. 265–277.

Kim, Seon Joo, Shaojie Zhuo, Fanbo Deng, Chi-Wing Fu, and Michael Brown. November 2010. “Interactive Visualization of Hyperspectral Images of Historical Documents”. *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, p. 1441–1448.

Kim, Seon Joo, Fanbo Deng, and Michael S. Brown. 2011. “Visual enhancement of old documents with hyperspectral imaging”. *Pattern Recognition*, vol. 44, n° 7, p. 1461-1469.

Kittler, J. and J. Illingworth. 1985. “On threshold selection using clustering criteria”. *IEEE Trans. Syst. Man Cybern.*, vol. SMC-15, p. 652-655.

Klein, M. E., J. H. Scholten, G. Sciutto, Th. A. G. Steemers, and G. De Bruin. 2006. “The Quantitative Hyperspectral Imager - A Novel Non-destructive Optical Instrument for monitoring Historic Documents”. *International Preservation News*, vol. 40, p. 4-9.

Klein, M. E., B. J. Aalderink, R. Padoan, G. de Bruin, and T. A. Steemers. March 2008. “Quantitative Hyperspectral Reflectance Imaging”. *Sensors*, vol. 9, n° 8.

Knox, Keith T. July 1997. “Show-through correction for two-sided documents”. United States Patent 5646744.

Kokla, V., A. Psarrou, and V. Konstantinou. 2006. “Ink recognition based on statistical classification methods”. In *Document Image Analysis for Libraries, 2006. DIAL ’06. Second International Conference on*. p. 11 pp.–264.

Kuk, Jung Gap, Nam Ik Cho, and Kyoung Mu Lee. 2008. “MAP-MRF approach for binarization of degraded document image”. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. p. 2612–2615.

Landgrebe, David A. 1980. “The development of a spectral-spatial classifier for earth observational data”. *Pattern Recognition*, vol. 12, n° 3, p. 165-175.

Leedham, G., S. Varma, A. Patankar, and V. Govindaraju. 6–8 Aug. 2002. “Separating text and background in degraded document images - a comparison of global thresholding techniques for multi-stage thresholding”. In *Proc. Eighth International Workshop on Frontiers in Handwriting Recognition*. p. 244–249.

Lelore, Thibault and Frèdèric Bouchara. July 26–29 2009. “Document image binarisation using Markov Field Model”. In *ICDAR’2009*. (Barcelona, Spain 2009), p. 551–555.

Lettner, Martin and Robert Sablatnig. 2010. “Higher order MRF for foreground-background separation in multi-spectral images of historical manuscripts”. In *DAS’10*. (Boston, Massachusetts 2010), p. 317–324. ACM.

Lettner, Martin, Markus Diem, Robert Sablatnig, and Heinz Miklas. Aug 26–29 2008. “REGISTRATION AND ENHANCING OF MULTISPECTRAL MANUSCRIPT IMAGES”. In *Proceedings of EUSIPCO 2008*. (Lausanne, Switzerland 2008).

Leydier, Y., F. Le Bourgeois, and H. Emptoz. 2004. “Serialized Unsupervised Classifier for Adaptative Color Image Segmentation: Application to Digitized Ancient Manuscripts”. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR’04) Volume 1 - Volume 01*. (Washington, DC, USA 2004), p. 494–497. IEEE Computer Society.

Li, Xiang, Ben Aldridge, Lucia Ballerini, Bob Fisher, and Jonathan Rees. 2009. “Depth improves skin lesion segmentation”. In *Medical Image Computing and Computer Assisted Intervention*. p. 1101–1107.

Li, Yunlei, Lodewyk F. A. Wessels, Dick de Ridder, and Marcel J. T. Reinders. December 2007. “Classification in the presence of class noise using a probabilistic Kernel Fisher method”. *Pattern Recogn.*, vol. 40, n° 12, p. 3349–3357.

LIoud, D.E. Dec. 1985. *Automatic target classification using moment invariant of image shapes*. RAE IDN AW126, Farnborough, UK :.

Lu, Shijian, Bolan Su, and Chew Lim Tan. December 2010. “Document image binarization using background estimation and stroke edges”. *Int. J. Doc. Anal. Recognit.*, vol. 13, p. 303–314.

Maino, G. sept. 2007. “Digitization and Multispectral Analysis of Historical Books and Archival Documents: Two Exemplary Cases”. In *Image Analysis and Processing Workshops, 2007. ICIAPW 2007. 14th International Conference on*. p. 119-124.

Manso, M. and M.L. Carvalho. 2009. “Application of spectroscopic techniques for the study of paper documents: A survey”. *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 64, n° 6, p. 482 - 490.

Mardia, K.V. and T.J. Hainsworth. nov 1988. “A spatial thresholding method for image segmentation”. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 10, n° 6, p. 919 -927.

Marinai, Simone, Marco Gori, and Giovanni Soda. 2005. “Artificial Neural Networks for Document Analysis and Recognition”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, n° 1, p. 23–35.

Martin, D., C. Fowlkes, D. Tal, and J. Malik. July 2001. “A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics”. In *Proc. 8th Int’l Conf. Computer Vision*. p. 416–423.

Martinelli, Francesca, Emanuele Salerno, Ivan Gerace, and Anna Tonazzini. January 2012. “Nonlinear model and constrained ML for removing back-to-front interferences from recto-verso documents”. *Pattern Recogn.*, vol. 45, n° 1, p. 596–605.

Melessanaki, K., V. Papadakis, C. Balas, and D. Anglos. 2001. “Laser induced breakdown spectroscopy and hyper-spectral imaging analysis of pigments on an illuminated manuscript.”. *Spectrochim. Acta B*, vol. 56, p. 2337-2346.

Mendez-Rial, R., M. Calvino-Cancela, and J. Martin-Herrero. march 2012. “Anisotropic Inpainting of the Hypercube”. *Geoscience and Remote Sensing Letters, IEEE*, vol. 9, n° 2, p. 214 -218.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. 1953. “Equations of State Calculations by Fast Computing Machines”. *Journal of Chemical Physics*, vol. 21, n° 6, p. 1087–1092.

Mezghani, Neila, Amar Mitiche, and Mohamed Cheriet. 2008. “ Bayes Classification of Online Arabic Characters by Gibbs Modeling of Class Conditional Densities ”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, n° 7, p. 1121–1131.

Moghaddam, Reza Farrahi and Mohamed Cheriet. 2009. “ RSLDI: Restoration of single-sided low-quality document images ”. *Pattern Recognition*, vol. 42, n° 12, p. 3355 - 3364.

Moghaddam, Reza Farrahi and Mohamed Cheriet. 2012. “ AdOtsu: An adaptive and parameterless generalization of Otsu’s method for document image binarization ”. *Pattern Recognition*, vol. 45, n° 6, p. 2419-2431.

Moghaddam, Reza Farrahi, David Rivest-Hénault, Itay Bar Yosef, and Mohamed Cheriet. 2009. “ A Unified Framework Based on the Level Set Approach for Segmentation of Unconstrained Double-Sided Document Images Suffering from Bleed-Through ”. In *ICDAR*. p. 441-445.

Mukhopadhyay, S. and B. Chanda. 2003. “ Multiscale morphological segmentation of gray-scale images ”. *Image Processing, IEEE Transactions on*, vol. 12, n° 5, p. 533–549.

Nagy, G. jan 2000. “ Twenty years of document image analysis in PAMI ”. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, n° 1, p. 38 -62.

Nagy, G., S. Seth, and M. Viswanathan. july 1992. “ A prototype document image analysis system for technical journals ”. *Computer*, vol. 25, n° 7, p. 10 -22.

Nakayama, K., H. Hasegawa, and C .A. Hernandez. September 1993. “ Handwritten Alphabet and Digit Character Recognition Using Skeleton Pattern Mapping With Structural Constraints ”. In *Proc. ICANN’93*. (Amsterdam 1993), p. 941.

Nascimento, J.M.P. and J.M.B. Dias. april 2005. “ Vertex component analysis: a fast algorithm to unmix hyperspectral data ”. *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, n° 4, p. 898 - 910.

Niblack, W., 1986. *An Introduction to Image Processing*. Englewood Cliffs, NJ : Prentice-Hall.

Nishida, H. and T. Suzuki. 2002. “ Correcting show-through effects on document images by multiscale analysis ”. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. p. 65–68 vol.3.

Niyobuhungiro, Japhet. 2010. “ Image Restoration and Text Removal ”. Master’s thesis, National University of Rwanda.

O’Gorman, Lawrence. November 1994. “ Binarization and multithresholding of document images using connectivity ”. *CVGIP: Graph. Models Image Process.*, vol. 56, n° 6, p. 494–506.

O’Gorman, Lawrence, 1997. *Document Image Analysis*. IEEE Computer Society Executive Briefings, 1 edition, 128 p.

Ophir, Boaz and David Malah. 2007. “Show-Through Cancellation in Scanned Images using Blind Source Separation Techniques”. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. p. III-233–III-236.

Otsu, N. 1979. “A Threshold Selection Method from Gray-Level Histograms”. *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, p. 62–66.

P. Stathis, E. Kavallieratou and N. Papamarkos. 2008. “An evaluation technique for binarization algorithms”. *Journal of universal Computer Science*, vol. 14, p. 3011–3030.

Pelagotti, A., A. Del Mastio, A. De Rosa, and A. Piva. 2008a. “Multispectral imaging of paintings”. *Signal Processing Magazine, IEEE*, vol. 25, n° 4, p. 27 -36.

Pelagotti, A., A. Del Mastio, A. De Rosa, and A. Piva. july 2008b. “Multispectral imaging of paintings”. *Signal Processing Magazine, IEEE*, vol. 25, n° 4, p. 27 -36.

Pratikakis, I., B. Gatos, and K. Ntirogiannis. nov. 2010. “H-DIBCO 2010 - Handwritten Document Image Binarization Competition”. In *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*. p. 727 -732.

Pratikakis, Ioannis, Basilios Gatos, and Konstantinos Ntirogiannis. 2011. “ICDAR 2011 Document Image Binarization Contest (DIBCO 2011)”. In *ICDAR*. p. 1506-1510.

Quinlan, J.R. 1986. “Introduction of decision trees”. *Machine Learning*, vol. 1, p. 81–106.

Raudys, S.J. and A.K. Jain. mar 1991. “Small sample size effects in statistical pattern recognition: recommendations for practitioners”. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 13, n° 3, p. 252 -264.

Ribes, A., R. Pillay, F. Schmitt, and C. Lahanier. 2008. “Studying That Smile”. *Signal Processing Magazine, IEEE*, vol. 25, n° 4, p. 14 -26.

Richards, J. A. and X. Jia, 1999. *Remote Sensing Digital Image Analysis: An Introduction*. Berlin, germany : Springer-Verlag.

Rivest-Hénault, David, Reza Farrahi Moghaddam, and Mohamed Cheriet. 2011. “A local linear level set method for the binarization of degraded historical document images”. *IJDAR*, vol. Online First, DOI: 10.1007/s10032-011-0157-5.

Rivest-Hénault, David, Reza Farrahi Moghaddam, and Mohamed Cheriet. 2012. “A local linear level set method for the binarization of degraded historical document images”. *IJDAR*, vol. 15, n° 2, p. 101-124.

Rivest-Hénault, David, Reza Farrahi Moghaddam, and Mohamed Cheriet. In Press. “A local linear level set method for the binarization of degraded historical document images”. *IJDAR*, vol. DOI: 10.1007/s10032-011-0157-5.

Rowley-Brooke, Róisin and Anil Kokaram. 2012. “Bleed-through removal in degraded documents”. p. 82970T-82970T-10.

Salerno, Emanuele, Anna Tonazzini, and Luigi Bedini. April 2007. “ Digital image analysis to enhance underwritten text in the Archimedes palimpsest ”. *IJDAR*, vol. 9, n° 2, p. 79–87.

Sauvola, J. and M. Pietikainen. February 2000. “ Adaptive document image binarization ”. *Pattern Recognition*, vol. 33, n° 2, p. 225–236.

Senvaitiene, J., A. Beganskiene, and A. Kareiva. January 2005. “ Spectroscopic evaluation and characterization of different historical writing inks ”. *Vibrational Spectroscopy*, vol. 37, n° 1, p. 61–67.

Sezgin, Mehmet and Bulent Sankur. January 2004. “ Survey over image thresholding techniques and quantitative performance evaluation ”. *J. Electron. Imaging*, vol. 13, n° 1, p. 146–168.

Shafait, Faisal, Daniel Keysers, and Thomas M. Breuel. January 2008. “ Efficient implementation of local adaptive thresholding techniques using integral images ”. In *Document Recognition and Retrieval XV*. (San Jose, CA 2008).

Shen, Huanfeng and Liangpei Zhang. may 2009. “ A MAP-Based Algorithm for Destriping and Inpainting of Remotely Sensed Images ”. *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, n° 5, p. 1492 -1502.

Shiel, Patrick. July 2010. “ Hyperspectral Image Analysis for Questioned Historical Documents ”. Master’s thesis, National University of Ireland, Maynooth, Maynooth, Co.Kildare, Ireland.

Smith, Elisa H. Barney. 2010. “ An analysis of binarization ground truthing ”. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. (New York, NY, USA 2010), p. 27–34. ACM.

Solit, David B. *et. al.* January 2006. “ BRAF mutation predicts sensitivity to MEK inhibition ”. *Nature*, vol. 439, n° 7074, p. 358–362.

Steinherz, Tal, Nathan Intrator, and Ehud Rivlin. 2000. “ A Special Skeletonization Algorithm for Cursive Words ”. In *IWFHR’00*. p. 529–534.

Stratis, D I M Itra N, Kristine L Eland, J Chance Carter, S A M Uel, J T O M Linson, and S M Ichael Angel. 2001. “ Comparison of Acousto-optic and Liquid Crystal Tunable Filters for Laser-Induced Breakdown Spectroscopy ”. *Applied Spectroscopy*, vol. 55, n° 8, p. 999–1004.

Su, Bolan, Shijian Lu, and Chew Lim Tan. June 9–11 2010. “ Binarization of historical document images using the local maximum and minimum ”. In *DAS’10*. (Boston, Massachusetts 2010), p. 159–166. ACM.

Su, F. and A. Mohammad-Djafari. 2007. “ Bayesian Separation of Document Images with Hidden Markov Model ”. In *2nd International Conference on Computer Vision Theory and Applications*. (Barcelona, Spain 2007).

Svozil, Daniel, Vladimír Kvasnickab, and Jiri Pospichalb. Nov. 1997. "Introduction to multilayer feed-forward neural networks". *Chemometrics and Intelligent Laboratory Systems*, vol. 39, p. 43–62.

Tahtouh, Mark, Pauline Despland, Ronald Shimmon, John R. Kalman, and Brian J. Reedy. 2007. "The Application of Infrared Chemical Imaging to the Detection and Enhancement of Latent Fingerprints: Method Optimization and Further Findings". *Journal of Forensic Sciences*, vol. 52, n° 5, p. 1089–1096.

Tan, Chew Lim, Ruini Cao, Peiyi Shen, Qian Wang, Julia Chee, and Josephine Chang. 2000. "Removal of interfering strokes in double-sided document images". In *Applications of Computer Vision, 2000, Fifth IEEE Workshop on.* (Palm Springs, CA 2000), p. 16–21.

Tan, Chew Lim, R. Cao, and Peiyi Shen. 2002. "Restoration of archival documents using a wavelet technique". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, n° 10, p. 1399–1404.

Taxt, T., P.J. Flynn, and A.K. Jain. 1989. "Segmentation of Document Images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, p. 1322-1329.

Taylor, W. A., D. H. Weimann, and P. J. Martin. 1995. "Knowledge Acquisition and Synthesis in a Multiple Source Multiple Domain Process Context". *Expert Systems with Applications*, vol. 8, n° 2, p. 295–302.

Toennies, K.D., 2005, P.202. *Grundlagen der Bildverarbeitung*. Pearson Studium.

Tonazzini, A., I. Gerace, and F. Martinelli. april 2010. "Multichannel Blind Separation and Deconvolution of Images for Document Analysis". *Image Processing, IEEE Transactions on*, vol. 19, n° 4, p. 912 -925.

Tonazzini, Anna, Emanuele Salerno, Matteo Mochi, and Luigi Bedini. 2004. "Blind Source Separation Techniques for Detecting Hidden Texts and Textures in Document Images". *Image Analysis and Recognition*, p. 241–248.

Tonazzini, Anna, Emanuele Salerno, and Luigi Bedini. June 2007. "Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique". *IJDAR*, vol. 10, n° 1, p. 17–25.

van Rijsbergen, C. J., 1979. *Information Retrieval*. ed. 2nd. London : Butterworths.

Vonikakis, Vassilios, Ioannis Andreadis, and Nikos Papamarkos. 2011. "Robust document binarization with OFF center-surround cells". *Pattern Anal. Appl.*, vol. 14, n° 3, p. 219-234.

Walvoord, D.J. and R.L. Easton. 2008. "Digital Transcription of the Archimedes Palimpsest [Applications Corner]". *Signal Processing Magazine, IEEE*, vol. 25, n° 4, p. 100 -104.

Wang, Q., T. Xia, C.L. Tan, and L. Li. 2003. "Directional wavelet approach to remove document image interference". In *ICDAR03*. p. 736-740.

Wang, Qian and Chew Lim Tan. 2001. “Matching of double-sided document images to remove interference”. In *in IEEE Conference on Computer Vision and Pattern Recognition*. p. 1084–1089. a.

Warfield, Simon K., Kelly H. Zou, and William M. Wells. 2004. “Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation”. *IEEE TRANS. MED. IMAG*, vol. 23, p. 903–921.

Way, Ted W., Berkman Sahiner, Lubomir M, Hadjiiski, and Heang-Ping Chan. Feb. 2010. “Effect of finite sample size on feature selection and classification: A simulation study”. *Medical physics*, vol. 37, n° 2, p. 907–920.

Wilson, Dennis L. july 1972. “Asymptotic Properties of Nearest Neighbor Rules Using Edited Data”. *Systems, Man and Cybernetics, IEEE Transactions on*, vol. SMC-2, n° 3, p. 408 -421.

Wolf, C., J.-M. Jolion, and F. Chassaing. 2002. “Text Localization, Enhancement and Binarization in Multimedia Documents”. In *Proceedings of the International Conference on Pattern Recognition*. p. 1037-1040.

Wolf, Christian. 2009. “Document Ink bleed-through removal with two hidden Markov random fields and a single observation field”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Yang, Luren, Fritz Albregtsen, Tor Lønnestad, and Per Grøttum. 1995. “A Supervised Approach to the Evaluation of Image Segmentation Methods”. In *BP 101 - 54602 Villers-ls-Nancy Cedex (France) Unit de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France) Unit de recherche INRIA Rhne-Alpes : 655, avenue de l’Europe - 38330 Montbonnot-St-Martin (France) Unit d*. p. 759–765. Springer.

Yedidia, Jonathan S., William T. Freeman, and Yair Weiss. 2003. Exploring artificial intelligence in the new millennium. chapter Understanding belief propagation and its generalizations, p. 239–269. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. ISBN 1-55860-811-7. <http://dl.acm.org/citation.cfm?id=779343.779352>.

Young, D. P. and J. M. Ferryman. 2005. “PETS Metrics: On-Line Performance Evaluation Service”. In *Proceedings of the 14th International Conference on Computer Communications and Networks*. (Washington, DC, USA 2005), p. 317–324. IEEE Computer Society.

Yuan, Xiaojing, Ning Situ, and George Zouridakis. June 2009. “A narrow band graph partitioning method for skin lesion segmentation”. *Pattern Recogn.*, vol. 42, n° 6, p. 1017–1028.

Zhang, Li, Andy M. Yip, Michael S. Brown, and Chew Lim Tan. 2009. “A Unified Framework for Document Restoration using Inpainting and Shape-from-Shading”.