Modèle de dégradation d'images de documents anciens pour la génération de données semi-synthétiques

van Cuong Kieu

UNIVERSITÉ DE LA ROCHELLE
ÉCOLE DOCTORALE S2IM
Sciences et Ingénierie pour l'Information, Mathématiques

Laboratoire L3i
Informatique, Image et Interaction

Modèle de dégradation d'images de documents anciens pour la génération de données semi-synthétiques

THÈSE
par
Van Cuong KIEU

Présenté et soutenue publiquement pour l'obtention du — LE 25 NOVEMBRE 2014, DOCTORAT DE L'UNIVERSITÉ DE LA ROCHELLE

Directeur de thèse :	RÉMY MULLOT	L3i, Professeur à l'Université de La Rochelle
Co-directeur de thèse :	JEAN-PHILIPPE DOMENGER	LaBRI, Professeur à l'Université de Bordeaux
Encadrant de thèse :	NICHOLAS JOURNET	LaBRI, Mcf à l'Université de Bordeaux
Encadrant de thèse :	MURIEL VISANI	L3i, Mcf à l'Université de La Rochelle
Rapporteurs :	NICOLE VINCENT	LIPADE, professeur à l'Université Paris Descartes
	JOSEP LLADOS	CVC, professeur à l'Université Autonome de Barcelone
Examinateur :	BERTRAND COUASNON	Irisa, Mcf-HDR à l'INSA de Rennes
Invité :	JEAN-PHILIPPE MOREUX	BnF - Bibliothèque Nationale de France

25 Novembre 2014

Remerciement

Tout d'abord, je remercie l'ensemble de mon jury de thèse à savoir Nicole Vincent, Josep Llados, Bertrand Couasnon, Rémy Mullot, Jean-Philippse Domenger, Nicholas Journet, et Muriel Visani.

De plus, cette thèse a été réalisée en parallèle d'un projet ANR nommé DIGIDOC. Ce projet a été une réelle source d'inspiration et de motivation. J'aimerais remercier l'ensemble de ses membres et en particulier Jean-Yves Ramel, Anne Vialard et Vincent Rabeux. Je remercie également les collaborateurs : BnF [1], l'équipe CESR-Tours [2] et Arkhênum [3] qui partagent des informations, des images représentatives.

Cette thèse n'aurait pas été la même sans le soutien et le travail de mes directeurs et encadrants de thèse. Je remercie, Rémy Mullot et Jean-Philippe Domenger, mon directeur et co-directeur de thèse, pour leurs conseils avisés tout au long de cette thèse et pour leur présence dans mon jury. Un grand merci au Nicholas Journet et Muriel Visani pour les discussions informelles reflétant parfaitement leur disponibilité et leur plaisir de partage.

J'ai eu la chance de travailler au sein deux équipes Document-LaBRI et Adoc-L3i dont la bonne ambiance a largement contribué à mon bien-être qui a constitué pendant trois ans une sorte de "deuxième famille" à mes yeux. Je tiens à les remercier tous. Tout d'abord, merci à Jérôme Charton dont son aide dans le domaine de vision par ordinateur m'a motivé la partie importante de ma thèse. Je remercie également aux doctorants du groupe DIVA-Fribourg, seniors chercheurs du groupe CVC-Barcelone pour leurs travaux impressionnants dans la collaboration ensemble, en plus de ceux déjà cités : HAO Wei, Angelika Garz, Alicia Fornés, et Joan Pastor.

Je présente mes remerciements plus personnels à formuler, à ma famille tout d'abord. A mes parents que j'admire, parce que leur courage, leur humour et leur bonté me donnent le plus bel exemple qui soit. A ma sœur, que j'aime de tout mon cœur. Un grand merci à Tra KIEU, ce qui partage sa vie, ses joies, ses bonheur avec moi. Elle m'écoute, me comprend, et me faire rire.

Je tiens à remercier tous mes amis. Merci en particulier à vous deux Quang TRAN et Duc Manh NGUYEN, ceux qui ont soutenu ou vont soutenir bientôt leur thèse de maths, et ainsi que ceux qui ont choisi l'enseignement depuis déjà quelques années à Bordeaux I pour vos aides dans des problèmes "durs" de maths et de statistique. Un immense merci à

1. http ://www.bnf.fr/fr/
2. http ://cesr.univ-tours.fr/
3. http ://www.arkhenum.fr/

V. N. Tung NGUYEN, Kim-Dung BUI, et Q. Anh BUI pour leur grande gentillesse et le bon moment qu'on a partagé dans les matchs de foot, de badminton.

Merci à tous, et merci à tous ceux que j'aurais éventuellement oubliés de me pardonner.

Résumé

Le nombre important de campagnes de numérisation mises en place ces deux dernières décennies a entraîné une effervescence scientifique ayant mené à la création de nombreuses méthodes pour traiter et/ou analyser ces images de documents (reconnaissance d'écriture, analyse de la structure de documents, détection/indexation et recherche d'éléments graphiques, etc.). Un bon nombre de ces approches est basé sur un apprentissage (supervisé, semi-supervisé ou non supervisé). Afin de pouvoir entraîner les algorithmes correspondants et en comparer les performances, la communauté scientifique a un fort besoin de bases publiques d'images de documents avec la vérité-terrain correspondante, et suffisamment exhaustive pour contenir des exemples représentatifs du contenu des documents à traiter ou analyser. La constitution de bases d'images de documents réels nécessite d'annoter les données (constituer la vérité terrain). Les performances des approches récentes d'annotation automatique étant très liées à la qualité et à l'exhaustivité des données d'apprentissage, ce processus d'annotation reste très largement manuel. Ce processus peut s'avérer complexe, subjectif et fastidieux. Afin de tenter de pallier à ces difficultés, plusieurs initiatives de crowdsourcing ont vu le jour ces dernières années, certaines sous la forme de jeux pour les rendre plus attractives. Si ce type d'initiatives permet effectivement de réduire le coût et la subjectivité des annotations, restent un certain nombre de difficultés techniques difficiles à résoudre de manière complètement automatique, par exemple l'alignement de la transcription et des lignes de texte automatiquement extraites des images. Une alternative à la création systématique de bases d'images de documents étiquetées manuellement a été imaginée dès le début des années 90. Cette alternative consiste à générer des images semi-synthétiques imitant les images réelles. La génération d'images de documents semi-synthétiques permet de constituer rapidement un volume de données important et varié, répondant ainsi aux besoins de la communauté pour l'apprentissage et l'évaluation de performances de leurs algorithmes.

Dans la cadre du projet DIGIDOC (Document Image diGitisation with Interactive DescriptiOn Capability) financé par l'ANR (Agence Nationale de la Recherche), nous avons mené des travaux de recherche relatifs à la génération d'images de documents anciens semi-synthétiques. Le premier apport majeur de nos travaux réside dans la création de plusieurs modèles de dégradation permettant de reproduire de manière synthétique des déformations couramment rencontrées dans les images de documents anciens (dégradation de l'encre, déformation du papier, apparition de la transparence, etc.). Le second apport majeur de ces travaux de recherche est la mise en place de plusieurs bases d'images semi-synthétiques utilisées dans des campagnes de test (compétition ICDAR2013, GREC2013) ou pour améliorer

par ré-apprentissage les résultats de méthodes de reconnaissance de caractères, de segmentation ou de binarisation. Ces travaux ont abouti sur plusieurs collaborations nationales et internationales, qui se sont soldées en particulier par plusieurs publications communes. Notre but est de valider de manière la plus objective possible, et en collaboration avec la communauté scientifique concernée, l'intérêt des images de documents anciens semi-synthétiques générées pour l'évaluation de performances et le ré-apprentissage.

Abstract

In the last two decades, the increase in document image digitization projects results in scientific effervescence for conceiving document image processing and analysis algorithms (handwritten recognition, structure document analysis, spotting and indexing / retrieval graphical elements, etc.). A number of successful algorithms are based on learning (supervised, semi-supervised or unsupervised). In order to train such algorithms and to compare their performances, the scientific community on document image analysis needs many publicly available annotated document image databases. Their contents must be exhaustive enough to be representative of the possible variations in the documents to process / analyze. To create real document image databases, one needs an automatic or a manual annotation process. The performance of an automatic annotation process is proportional to the quality and completeness of these databases, and therefore annotation remains largely manual. Regarding the manual process, it is complicated, subjective, and tedious. To overcome such difficulties, several crowd-sourcing initiatives have been proposed, and some of them being modelled as a game to be more attractive. Such processes reduce significantly the price and subjectivity of annotation, but difficulties still exist. For example, transcription and text-line alignment have to be carried out manually. Since the 1990s, alternative document image generation approaches have been proposed including in generating semi-synthetic document images mimicking real ones. Semi-synthetic document image generation allows creating rapidly and cheaply benchmarking databases for evaluating the performances and training document processing and analysis algorithms.

In the context of the project DIGIDOC (Document Image diGitisation with Interactive DescriptiOn Capability) funded by ANR (Agence Nationale de la Recherche), we focus on semi-synthetic document image generation adapted to ancient documents. First, we investigate new degradation models or adapt existing degradation models to ancient documents such as bleed-through model, distortion model, character degradation model, etc. Second, we apply such degradation models to generate semi-synthetic document image databases for performance evaluation (e.g the competition ICDAR2013, GREC2013) or for performance improvement (by re-training a handwritten recognition system, a segmentation system, and a binarisation system). This research work raises many collaboration opportunities with other researchers to share our experimental results with our scientific community. This collaborative work also helps us to validate our degradation models and to prove the efficiency of semi-synthetic document images for performance evaluation and re-training.

Table des matières

Chapitre 1

Contexte et motivation

1.1 Pourquoi la communauté scientifique a-t-elle besoin de données synthétiques ?

Débutés il y a prêt de 30 ans, un grand nombre de projets de numérisation ont vu le jour [1]. L'objectif d'une campagne de numérisation est double. Il vise à la fois à conserver le patrimoine documentaire et à en permettre une diffusion massive *via* des bibliothèques numériques. L'une des premières initiatives de numérisation a été la création de la bibliothèque numérique de "la Mémoire Américaine [2]" qui a permis de numériser un grand nombre de données (documents papiers, audio, vidéos non numériques, etc), pour ensuite les diffuser gratuitement sur Internet. Le budget du projet s'élevait à 73 millions de dollars sur dix ans. Plus de 5.000.000 de ressources ont été numérisées. Ces documents provenaient de 44 bibliothèques.

La première bibliothèque numérique française (Association des Bibliophiles Universelles) [3] a été créée en 1993. Jusqu'en 2002, cette bibliothèque numérique contenait 288 textes (transcriptions, ouvrages divers) issus de 101 auteurs. En 1997, la Bibliothèque nationale de France (BnF) a lancé le projet Gallica [4] dont les missions sont non seulement de créer une bibliothèque encyclopédique en ligne, mais encore de préserver des témoignages du patrimoine mondial documentaire. Ce projet a permis de mettre en ligne plus de 2 millions de documents.

Les enjeux relatifs à de telles campagnes de numérisation sont la maîtrise du temps alloué à cette numérisation, mais également la qualité des images numérisées. A titre d'exemple, lors de sa première campagne de numérisation, `Google` a estimé que la numérisation de 7 millions d'ouvrages (dans le contexte technologique de 2002) prendrait un millénaire [5]. En plus de l'ambition de réduire le temps de numérisation tout en maintenant des exigences de qualité de production, a émergé la volonté de donner au grand public accès au contenu de cette masse de documents (texte, illustration, etc.). De nombreux enjeux scientifiques ont donc émergé ces dernières décennies, donnant naissance à une communauté scientifique focalisant ses recherches sur le traitement et l'analyse d'images de documents ("communauté document") dont le domaine de recherche va du processus physique de numérisation à la mise en place de solutions logicielles permettant d'analyser le contenu des images de documents (reconnaissance de texte, analyse d'illustration, etc.).

1. liste de projets http://fr.wikipedia.org/wiki/Bibliothèque_numérique
2. http://memory.loc.gov/ammem/about/index.html
3. http://abu.cnam.fr/
4. http://gallica.bnf.fr/
5. http://www.google.fr/googlebooks/about/history.html

Le développement des travaux de recherche liés à la numération et à l'analyse d'images de documents s'est concrétisé par l'investissement massif de certaines entreprises sur ce sujet : Google, ABBYY, BNF, etc. En 2004, le géant Google a lancé un grand projet appelé "Google Books"[6]. Ce projet visait à numériser/diffuser des documents (7 millions d'ouvrages) tout en innovant sur la partie service en proposant des plateformes d'analyse d'image de documents. En plus des 219 contributions scientifiques publiées[7], Google a traduit son investissement dans le domaine en rachetant une plateforme de reconnaissance de caractères et en en faisant un logiciel gratuit (OCRopus[8]). Ce dernier est capable d'analyser la structure d'images de documents et de reconnaître des caractères (OCR). Une autre preuve de la volonté de Google d'innover sur ce domaine est l'argent accordé à des projets liés à l'analyse d'images de documents. Par exemple, le programme "Google's Digital Humanities Awards"[9] a financé 24 projets donnant lieu à la numérisation et à l'analyse de 12 millions ouvrages.

L'union européenne est également un grand acteur du domaine de la numérisation et de l'analyse de documents. Elle se positionne, face à Google, comme un concurrent crédible à même lui aussi de conserver et de diffuser le patrimoine documentaire mondial. L'Europe a mis en place des projets importants comme le projet Europeana[10] en 2008 et le projet IMPACT[11] en 2010. Ces deux projets ont regroupé 26 bibliothèques nationales des pays européens où des millions d'ouvrages anciens attendent d'être numérisés. Ces deux projets ont donné lieu à la création de nombreux outils de recherche[12] et d'un nombre conséquent de publications scientifiques.

Ces dernières années, la communauté scientifique mondiale s'est structurée. Elle regroupe un nombre important de laboratoires qui contribuent et collaborent sur les problématiques liées à la numérisation et à l'analyse de documents. La figure 1.1 donne une idée des pays possédant des laboratoires actuellement actifs en analyse d'images de documents. Plusieurs conférences et workshop dédiés à cette thématique de recherche ont été créés depuis la première conférence internationale sur document (ICDAR) en 1991. Ces évènements permettent de réunir des chercheurs et des industriels du monde entier favorisant ainsi l'échange de connaissances ainsi que le transfert industriel.

6. http://books.google.com/
7. http://research.google.com/pubs/MachinePerception.html
8. http://code.google.com/p/ocropus/
9. http://googleblog.blogspot.fr/2010/07/our-commitment-to-digital-humanities.html
10. http://www.europeana.eu/portal/
11. http://www.impact-project.eu/home/
12. http://www.impact-project.eu/taa/tech/tools/

Nous avons identifié 4 thèmes principaux (en relation avec nos travaux de recherche) sur lesquels se penchent la communauté :

- Pré-traitement et restauration : mise en place de méthodes permettant d'améliorer la qualité des images de documents digitalisés et de préparer les documents pour des phases ultérieures de traitement / analyse / interprétation. Cela a donné naissance à un grand nombre d'algorithmes de binarisation, de débruitage, de correction des distorsions, etc. ;
- Analyse de la structure physique et logique des documents : méthodes qui permettent de segmenter les zones composant un document (textes et illustrations), puis d'en classifier les contenus et d'en extraire les liens logiques entre les différentes zones (caractères, symboles, tables, logos, etc.) ;
- Reconnaissance de texte : méthodes de reconnaissance de caractères (OCR) et de mots dans les documents imprimés ou manuscrits ;
- Indexation de contenu : méthodes permettant d'indexer le contenu de documents pour réaliser des recherches en-ligne or hors-ligne le plus rapidement et précisément possible ;

FIGURE 1.1: Liste (non exhaustive) des principaux pays actuellement actifs en analyse d'images de documents (après analyse de l'origine des publications à ICDAR 2013).

1.1.1 Problématique de la constitution de bases d'images avec vérité-terrain

De manière transverse à ces 4 grandes problématiques identifiées, la nécessité de constituer des bases d'images avec vérité terrain à des fins d'évaluation de performances ou d'apprentissage est cruciale. La communauté scientifique a un fort besoin de bases d'images de documents publiques, annotées avec la vérité-terrain correspondante, et dont le contenu soit suffisamment exhaustif pour proposer des exemples représentatifs des documents à traiter/analyser. La constitution de bases d'images de documents réels nécessite d'annoter les données (constituer la vérité terrain). Ce processus d'annotation reste très largement manuel et peut donc s'avérer complexe, subjectif et fastidieux. Cela conduit également à la limitation des bases d'apprentissage en termes de taille et de variabilité. A l'heure du **"big data"**, il est important de pouvoir tester, alimenter et comparer les innovations scientifiques sur un volume conséquent de données correspondant à des réalités industrielles.

Une solution originale apportée par la communauté pour résoudre ce problème de mise en place de bases avec vérité-terrain a été le *crowdsourcing*. Certaines solutions ont été proposées sous forme de jeux pour les rendre plus attractives et ont concrètement permis de constituer des bases d'images de documents. Si ce type d'initiative permet effectivement de réduire le coût et la subjectivité des annotations, il reste cependant un certain nombre de difficultés à résoudre avant de les générer de manière complètement automatique. Par exemple, dans le cadre de l'acquisition de vérité-terrain d'images de documents manuscrits, l'alignement de la transcription et des lignes de texte extraites des images reste toujours un problème.

Une alternative à la création systématique de bases d'images de documents étiquetés manuellement a été imaginée dès le début des années 90. Cette alternative consiste à générer des images synthétiques imitant les images réelles. La génération d'images de documents synthétiques permet de constituer rapidement un volume de données important et varié, répondant ainsi en partie aux besoins de la communauté pour l'apprentissage et l'évaluation de performances de leurs algorithmes.

1.1.2 Projet DIGIDOC

Les travaux de recherche sur la génération d'images de documents synthétiques présentés dans ce manuscrit, prennent place au sein du projet ANR DIGIDOC (2010-2014). Ce projet a regroupé quatre laboratoires français (L3i La Rochelle, LaBRI Bordeaux, LI Tours, Litis Rouen), deux partenaires industriels (I2S Bordeaux, Arkhênum Bordeaux) et la

Bibliothèque nationale de France (BnF). L'objectif du projet DIGIDOC (Document Image diGitisation with Interactive DescriptiOn Capability) [13] était de concevoir un prototype d'un scanner cognitif capable d'adapter ses réglages en fonction du document (le plus souvent ancien) à numériser, de l'usage ultérieur qui en sera fait (archivage, reconnaissance de texte, indexation, etc.) et de la qualité de l'image numérisée attendue par l'opérateur. Par conséquent, le projet DIGIDOC et plus particulièrement la mise en place du scanner cognitif couvre un grand nombre de problématiques de l'analyse d'images de documents (binarisation, segmentation, analyse de la structure physique / logique, reconnaissance de texte, etc.). Afin d'entraîner et de tester ce scanner cognitif, le projet DIGIDOC a clairement identifié le besoin de constituer des bases d'images volumineuses et variées avec une vérité terrain associée. Conscient de la difficulté à acquérir un nombre important d'images avec vérité-terrain issues de documents réels (surtout anciens), le consortium DIGIDOC a souhaité mettre en place un processus de génération de données synthétiques avec vérité-terrain associée.

1.2 Enjeux scientifiques relatifs à la génération d'images de documents synthétiques

Une base d'images de documents associée à leur vérité terrain peut donc être constituée selon deux processus : annotation d'images de documents réels ou génération d'images de documents synthétiques. Les enjeux scientifiques du premier processus sont clairement identifiés et une partie de la communauté travaille de manière active sur ce problème. Les travaux les plus aboutis sont certainement ceux présentés dans [Antonacopoulos *et al.*, 2006].

Le deuxième processus est celui lié à la génération d'images synthétiques. Il débute réellement au début des années 90 avec les travaux de [Baird, 2000]. Les publications ultérieures à ces travaux fondateurs ont donné lieu à un nombre important d'études qui ont montré à chaque fois l'efficacité de la génération de bases synthétiques d'images de documents pour l'évaluation de performances [Jenkins et Kanai, 1994], [Ho et Baird, 1995], [Smith et Andersen, 2005], [Liang *et al.*, 2008], [Moghaddam et Cheriet, 2009], [Fornés *et al.*, 2011], [Rabeux *et al.*, 2011] et pour l'enrichissement de bases d'apprentissage [Mori *et al.*, 2000], [Varga et Bunke, 2003a].

Scientifiquement, il reste encore de nombreux verrous à résoudre concernant l'étape de génération d'images synthétiques et de ses usages. Tout d'abord, la notion de "synthétique" n'est pas encore clairement définie. La plupart des études précédemment citées basent

13. http://www.doconcloud.org:8080/DoQuBookWeb/index.jsf

leurs propositions sur une dégradation d'images de documents réels pour en générer une version synthétique. Clairement, ces images contiennent une partie réelle (des images originales) et une partie synthétique (les dégradations). Dans ce document, nous qualifierons ces images de "semi-synthétique" et introduirons la notion du "taux d'information synthétique" pour déterminer quelle quantité d'information synthétique est intégrée dans l'image. Cette notion de "taux d'information synthétique" nous permet non seulement d'étudier plusieurs manières de générer des images de documents, mais également d'aborder les verrous scientifiques liés à l'usage de données semi-synthétiques. Les questions à aborder sont (entre autres) : une image synthétique doit elle être visuellement réaliste pour être utile ? A quel point peut-on dégrader synthétiquement une image réelle ? Pour évaluer les performances d'un algorithme ou alimenter une étape d'apprentissage, peut-on utiliser uniquement des données semi-synthétiques ou doit-on les mélanger avec des données réelles ? En cas de mélange, quelle est la proposition optimale d'images semi-synthétiques à intégrer ?

Le cadre du projet DIGIDOC a amené ces travaux de thèse à se focaliser sur la génération d'images de **documents anciens**. A notre connaissance, il existe peu de bases d'images de documents anciens réels annotées et publiques (nous les listerons dans le chapitre suivant). En termes de travaux sur la génération synthétique de documents anciens, nous sommes parmi les premiers à nous pencher sur le problème. En termes de modélisation de dégradations, l'état de l'art propose majoritairement des modèles reproduisant des défauts liés à la numérisation de documents contemporains sur des images binarisées. S'ils sont numérisés en couleur, les documents anciens sont très souvent manipulés en niveaux de gris par les algorithmes d'analyse et d'interprétation d'images. Le premier enjeu de cette thèse sera donc de pouvoir proposer des modèles adaptés aux spécificités des images de documents anciens (image en niveaux de gris, déformations diverses dues à l'âge du support physique ou au processus complexe de numérisation de documents anciens).

Le second enjeu est, quant à lui, lié à la problématique de l'utilisation des données générées. Sont-elles utiles ? Dans quels cadres ? Combien faut-il en générer ? Quelle variété ? Faut-il combiner les dégradations ? Dans quelles proportions ? Peut-on les utiliser avec des données réelles ?

1.3 Organisation du mémoire

Le manuscrit s'articule autour de 5 chapitres présentant plus précisément le contexte de notre étude et l'ensemble de nos contributions.

Dans le chapitre 2, nous proposons et illustrons la notion de “taux d'information synthétique” (TIS) qui est le pourcentage de facteur synthétique intégré dans une image réelle afin de générer un défaut factice (luminosité, transparence, taches, etc.). A partir de cette définition, nous présentons des méthodes de l'état de l'art permettant de créer des bases d'images de documents avec vérité-terrain à partir d'images de documents réels, de documents semi-synthétiques ou de documents synthétiques. Finalement, nous détaillons l'architecture d'un nouveau générateur d'images semi-synthétiques adapté aux documents anciens. Ce générateur propose trois modes de génération d'images semi-synthétiques : la dégradation automatique d'images de documents réels, la génération automatique d'images en utilisant des éléments réels extraits d'images réelles, la génération semi-automatique en utilisant une interface conviviale pour l'utilisateur.

Le chapitre 3 présente les dégradations les plus couramment observées dans les documents et plus particulièrement dans les documents anciens. Nous détaillons également comment ces dégradations peuvent impacter le bon fonctionnement d'algorithmes de traitement ou d'analyse de documents Enfin, nous présentons les travaux de l'état de l'art relatifs aux modèles permettant de simuler des dégradations dans les documents.

Dans le chapitre 4 nous présentons deux apports majeurs de ces travaux de recherche. Le premier est un modèle de dégradation de caractères reproduisant de manière réaliste ceux existants dans les documents réels : petites taches noires à proximité de la bordure des caractères (dues aux processus d'imprimerie ancienne ou d'écriture ancienne) et disparition de l'encre due à l'âge du document, allant parfois jusqu'à briser la connectivité de caractères. En plus d'être réaliste, le modèle proposé a été spécifiquement étudié pour permettre une paramétrisation simple permettant à l'utilisateur de le configurer sans avoir à comprendre l'algorithme de génération des défauts. Au travers de divers exemples, nous montrerons l'intérêt de notre modèle par rapport à ceux de l'état de l'art (fonctionnel sur des images en niveaux de gris, réaliste visuellement et facilement paramétrable). La seconde partie de ce chapitre 4 présente un autre apport de nos travaux : une méthode permettant de reproduire de manière synthétique les déformations du papier. Cette méthode utilise des maillages de documents scannés en 3D. Elle permet de reproduire des distorsions globales (courbures, ondulations) et des distorsions locales (plis, trous, concavités) typiques des documents anciens. Comme pour le chapitre précédent, nous montrerons l'intérêt de notre modèle par rapport à ceux existants (principalement le réalisme du document généré).

Le chapitre 5 clôture ce manuscrit en présentant les multiples usages qu'il est possible de faire avec des images de documents semi-synthétiques générées à partir de nos

méthodes / modèles. Au travers de campagnes d'évaluation de performances sur des images de documents de natures différentes (partitions musicales et documents imprimés) nous montrerons qu'il est possible d'obtenir une analyse très fine des performances et des limites d'un algorithme de traitement ou d'analyse d'images. Ce chapitre se termine sur deux tests visant à valider l'intérêt d'utiliser des images semi-synthétiques dans un contexte d'apprentissage pour la reconnaissance de caractères manuscrits ou la prédiction de résultats d'algorithmes de binarisation. Ces tests permettront d'apporter des éléments de réponse aux problèmes liés à la constitution de bases avec des documents synthétiques (proportion d'images réelles/synthétiques, quantité de dégradation, etc.) Ces travaux sont réalisés en collaboration avec plusieurs chercheurs appartenant à différents laboratoires (CENPARMI-Québec au Canada, CVC-Barcelone, DIVA-Fribourg, L3i-La Rochelle, LaBRI-Bordeaux).

Chapitre 2

Constitution de bases d'images de documents

2.1 Introduction

Ces dernières décennies, la communauté document a développé un nombre important d'algorithmes de traitement et d'analyse d'images de documents. Ils permettent d'améliorer la qualité des images de documents (méthodes de restauration, filtrage de bruits divers, etc.), d'analyser la structure des documents (extraction de régions texte/non-texte, extraction de lignes de texte, de caractères, d'illustrations, etc.), de reconnaître et d'indexer des documents (OCR, reconnaissance d'écriture, moteur de recherche en/hors-ligne, etc.).

Les algorithmes d'analyse de documents nécessitent la mise en place de bases d'images de documents pour évaluer leurs performances et pour entraîner les classifieurs afférents. La constitution de bases d'images de documents réels associées à leur vérité-terrain est ainsi devenue une nécessité. Hélas, l'annotation se fait encore et le plus souvent manuellement. Cette étape peut vite s'avérer subjective et fastidieuse. Cela conduit à limiter la qualité et l'exhaustivité des bases d'images de documents réels. Il est à noter également que les groupes de recherche mettent en place leurs propres bases d'images qui sont parfois privées. Cela pose donc des problèmes de partage des données et de comparaison des différents algorithmes entre eux.

Une nouvelle proposition, faite au début des années 90, consiste à générer des images semi-synthétiques ou synthétiques comme alternative au problème de la constitution de bases d'images de documents réels. Cette proposition a pour ambition de permettre la constitution de bases analogues aux bases réelles, tout en contrôlant le contenu généré. La constitution de telles bases passe par un mécanisme qui permet d'obtenir des bases volumineuses avec un minimum d'interventions manuelles.

Les travaux de recherche associés à cette thèse s'insèrent dans la problématique de la génération d'images de documents semi-synthétiques et synthétiques. Ces travaux se sont déroulés dans le contexte du projet DIGIDOC [1]. Ils se sont concentrés sur encore sujet peu traité, celui de la génération d'images de documents anciens semi-synthétiques et synthétiques.

Ce chapitre présente les enjeux associés au processus de génération d'images semi-synthétiques et synthétiques, et plus généralement à la constitution de bases d'images avec vérité-terrain. Nous présentons tout d'abord les enjeux de la constitution de bases d'images de documents réels dans la section 2.2, puis nous détaillons ceux de la génération de bases d'images de documents semi-synthétiques et synthétiques dans la section 2.3.

1. http://www.doconcloud.org:8080/DoQuBookWeb/index.jsf

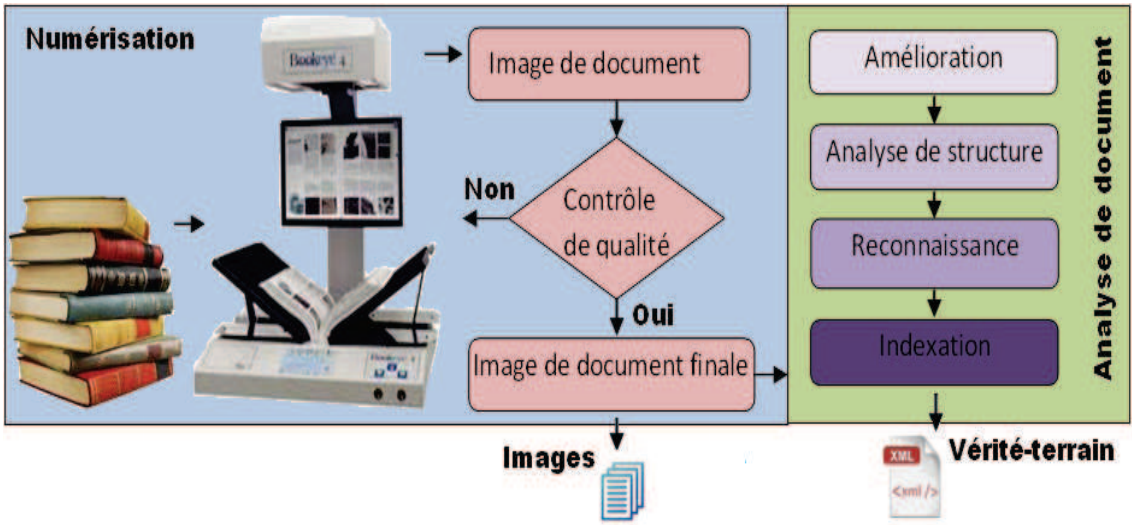


FIGURE 2.1: Les deux étapes du processus de constitution d'images de documents : la numérisation et l'analyse des images

2.2 Constitution de bases d'images de documents réels

2.2.1 Processus général

Le processus de constitution d'images de documents comprend classiquement deux étapes : la numérisation de documents et l'analyse des images en vue de leur indexation. Ces deux étapes sont présentées dans la figure 2.1 et résumées dans les sous-sections suivantes.

2.2.1.1 Numérisation de documents

La numérisation de documents comprend typiquement les trois étapes suivantes :

- Définition du cahier de charges : les exigences liées à la numérisation sont définies par le client. La première exigence est liée à la qualité des images, par exemple le taux de flou doit être classiquement inférieur à 5%. La seconde exigence concerne les formats d'images de sortie et leur vérité-terrain : par exemple l'image de sortie est fournie, dans la plupart des cas, dans un format non-compressé. Si une vérité terrain lui est associée, elle l'est dans un format structuré (XML).
- Numérisation des documents : cette étape permet de convertir un document physique en un document numérique en utilisant généralement un scanner, avec ou sans l'aide d'opérateurs humains. Les opérateurs choisissent le type de scanner adapté au document, par exemple un scanner automatique pour les documents

(a) (b) (c)

FIGURE 2.2: Exemples de scanners : (a) scanner automatique, (b) scanner grand format, (c) scanner manuel.

contemporains dans la figure 2.2-a, un scanner grand format comme montré dans la figure 2.2-b pour les cas particuliers que sont par exemple les documents architecturaux ou les cartes, un scanner manuel pour les documents anciens afin de les préserver (*cf.* la figure 2.2-c) ;

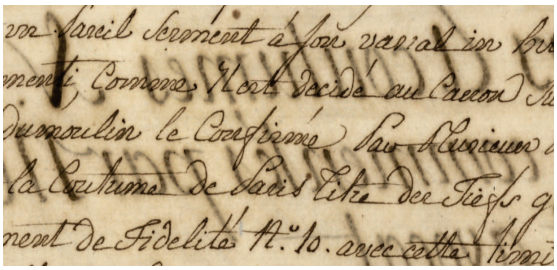
– Contrôle qualité : le contrôle de la qualité permet de vérifier que l'ensemble des images de sortie est conforme aux exigences du cahier de charges. Il peut engendrer des coûts supplémentaires dans le cas où des images de sortie sont refusées (les images refusées doivent être re-numérisées). En général, on applique deux types de contrôle : le contrôle automatique (*e.g.* détection du flou [Tong *et al.*, 2004]), puis le contrôle manuel. Ces contrôles sont essentiels afin de garantir la qualité des informations proposées. Les principales difficultés du contrôle portent sur la définition de critères objectifs permettant de valider ou non la conformité de l'image aux exigences du client, et l'échantillonnage des documents à contrôler afin de garantir la conformité au cahier de charges tout en évitant un contrôle exhaustif de documents ;

2.2.1.2 Analyse d'images de documents

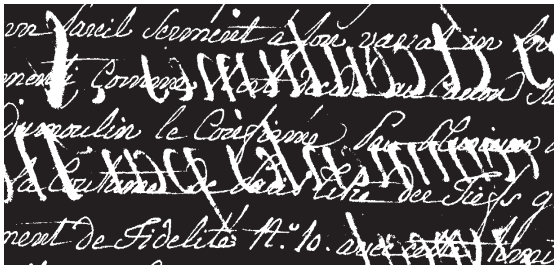
Après la numérisation, l'image est insérée dans une chaîne automatique ou semi-automatique d'analyse de documents qui permet d'extraire des informations qui pourront ensuite être utilisées par un mécanisme d'indexation. Nous décomposons cette chaîne selon les quatre étapes suivantes :

– Pré-traitements de l'image : cette étape permet d'améliorer la qualité d'une

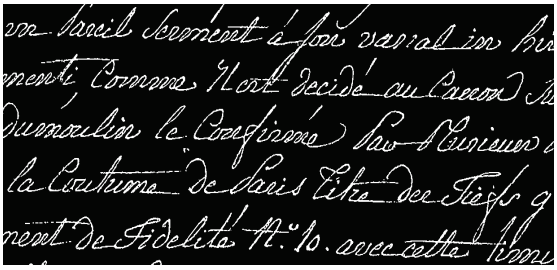
image pour s'adapter à l'application visée. La plupart des dégradations peuvent être numériquement corrigées, comme par exemple, les caractères coupés par les taches [Allier *et al.*, 2006], les plis ou courbures [Liang *et al.*, 2008], la transparence [Moghaddam et Cheriet, 2009], le bruit [Zhao et Pope, 2007], la luminosité [Thrin, 2003, Gouinaud *et al.*, 2011], le flou [Tong *et al.*, 2004, Srivastava *et al.*, 2009], etc. La figure 2.3 montre l'amélioration de la binarisation obtenue en combinant un algorithme d'élimination de transparence [Moghaddam et Cheriet, 2009] et un algorithme de binarisation adaptatif ;



(a) Image originale



(b) Résultat de la binarisation avec la méthode d'Otsu [Otsu, 1975]



(c) Résultat de la binarisation avec l'élimination de transparence [Moghaddam et Cheriet, 2009] et la méthode de Sauvola [Lazzara et Géraud, 2014]

FIGURE 2.3: La comparaison de l'amélioration obtenue après élimination de la transparence [Moghaddam et Cheriet, 2009] et la méthode de binarisation de Sauvola [Lazzara et Géraud, 2014] (c) binarisation par la méthode d'Otsu (b) image originale (a).

– Analyse de la structure de l'image de document : cette étape permet dans un premier temps de segmenter des éléments dans l'image du document, puis de les classer en plusieurs types : les caractères, les tableaux, les formules, les titres, les illustrations, etc. Grâce à cette étape, on obtient la plupart des informations physiques du document comme la position, la taille et la forme des éléments

segmentés ;

- Reconnaissance : cette étape permet d'extraire le contenu des documents. La reconnaissance peut consister en différentes tâches dépendant du type de documents et de l'application visée. Par exemple, la reconnaissance de caractères est utilisée sur les documents textuels imprimés (OCR) alors que la reconnaissance de symboles est utilisée pour les documents musicaux, architecturaux, etc. La performance d'une méthode de reconnaissance est directement liée à la qualité et à la variabilité des images d'entrée. Quand on parle de la qualité d'images, cela peut intégrer les dégradations dans les documents qui influent directement sur le taux de reconnaissance. Par exemple, une tache blanche qui coupe l'imagette de la lettre "l" en deux peut conduire à des erreurs d'OCR. La variabilité des images d'entrée est liée au type du document (document imprimé ou manuscrit, facture ou document administratif, journal, etc.), la langue du document, le style de l'écriture manuscrite, etc. Cette variabilité impacte directement les taux de reconnaissance. Cette phase de reconnaissance est donc assistée par un humain. On parle alors d'annotation et de transcription semi-automatique ou manuelle. Lors de cette phase d'annotation, l'opérateur humain pourra également corriger les résultats de l'analyse de structure ;
- Indexation : Les contenus des documents sont regroupés et indexés selon plusieurs critères pour la recherche. Par exemple, pour des documents de type texte, la transcription textuelle extraite automatiquement des images permet de faire des recherches par le contenu et/ou par mots-clés et des résumés de ces images de document ;

2.2.2 Bases d'images de documents réels existantes

Dans le cadre de l'évaluation de performances d'algorithmes d'analyse d'images de documents, les groupes de recherche construisent généralement leurs propres bases d'évaluation à partir de celles qui sont publiques. Il existe plusieurs bases d'images accessibles sur Internet [2, 3, 4]. Ces bases sont généralement dédiées à des cas d'usage précis de l'analyse de documents imprimés (Washington UW3 [Shahab *et al.*, 2010], LRDE [Lazzara *et al.*, 2011], RETAS-OCR [Yalniz et Manmatha, 2011], PaRADIIT [Roy *et al.*, 2011], les bases

2. http://www.iapr-tc11.org/mediawiki/index.php/Datasets_List
3. http://quod.lib.umich.edu/cgi/i/image/image-idx
4. http://www2.lib.udel.edu/eresources/digitalimages/

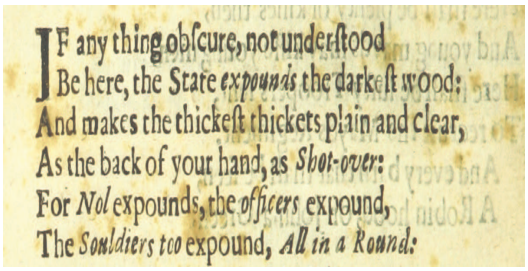
IMPACT [5], PRIMA [Clausner *et al.*, 2014], etc.), de documents manuscrits (bases de signature, de documents historiques : IAM [6], RIMES [Grosicki *et al.*, 2009], GERMANA [Perez *et al.*, 2009]), de documents graphiques (les bases de symboles chimiques [Nakagawa *et al.*, 2010], de logos [7], de symboles architecturaux [Delalandre *et al.*, 2010], de symboles musicaux CVC–MUSICMA [Fornés *et al.*, 2012]), de documents contenant des graphiques (la base Tobaco800 [8] [Tob, 2007] pour les logos), la base DIBCO [Pratikakis *et al.*, 2011] pour la binarisation, une base de lettrines (la base Navidomass [9]), etc. Nous résumons ci-dessous, par catégorie, quelques caractéristiques communes aux bases d'images de documents utilisées par la communauté.

La première catégorie regroupe les bases dédiées à l'analyse de documents imprimés. Ces documents utilisent des fontes régulières. Ils sont donc divisés en deux types : les documents imprimés anciens et les documents imprimés contemporains. Des bases d'images de documents imprimés anciens sont publiées comme la base PaRADIIT [Roy *et al.*, 2011] (*e.g.* la figure 2.4-a), la base DIBCO [Pratikakis *et al.*, 2011] contenant 36 images (*e.g.* la figure 2.4-b) ou des bases privées [10, 11, 12]. Ce type de base contient de nombreuses dégradations liées au processus d'impression ou au vieillissement du papier. Les bases d'images de documents imprimés contemporains comprennent des documents administratifs, formulaires, etc. Les deux bases IMPACT [13] et PRIMA [Clausner *et al.*, 2014] contiennent plus de 500.000 images qui sont annotées à l'aide de l'outil Aletheia [Clausner *et al.*, 2014] à des fins d'évaluation de performances d'OCR. On peut également citer la base Washington UW3 [Shahab *et al.*, 2010] contenant 1600 images de documents anciens, la base LRDE [Lazzara *et al.*, 2011] contenant 375 images, la base RETAS-OCR contenant 160 ouvrages scannés [Yalniz et Manmatha, 2011], etc.

La deuxième catégorie de base est dédiée à l'analyse de l'écriture manuscrites Dans le cas des documents anciens, les dégradations sont courantes. Par exemple, la figure 2.5-a montre une image d'écriture ancienne qui contient une dégradation diffuse de l'encre et une distorsion locale (pli du papier). On peut distinguer les bases relatives à la reconnaissance d'écriture en-ligne (c'est-à-dire contenant des informations sur la dynamique du tracé dans

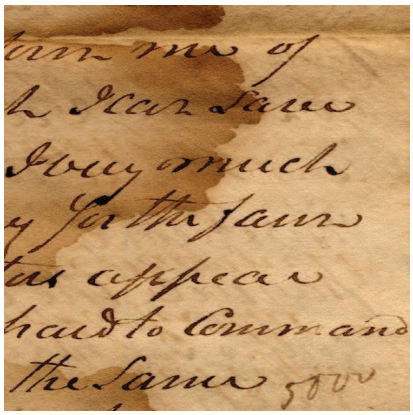
5. http://www.digitisation.eu/data/
6. http://www.iam.unibe.ch/fki/databases/iam-handwriting-database/iam-handwriting-database
7. http://www.eurecom.fr/huet/work.html
8. http://www.umiacs.umd.edu/~zhugy/tobacco800.html
9. http://navidomass.univ-lr.fr/ressources.html
10. http://digitalcollections.lib.washington.edu/cdm/search/collection/pioneerlife
11. http://www.archives.gov/historical-docs/
12. https://apps.carleton.edu/campus/library/dbs/historical/
13. http://www.digitisation.eu/data/

(a) Une image de la base PaRADIIT [Roy *et al.*, 2011]

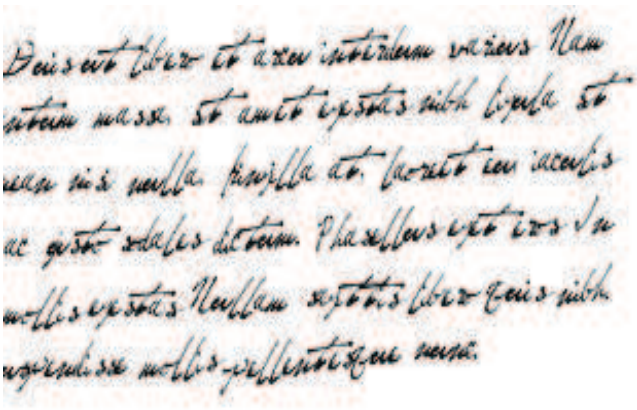


(b) Une image de la base DIBCO [Pratikakis *et al.*, 2011]

FIGURE 2.4: Exemples d'images de documents imprimés anciens : (a) une image de la base PaRADIIT [Roy *et al.*, 2011], (b) une image de la base DIBCO [Pratikakis *et al.*, 2011].



(a) Document manuscrit ancien



(b) Document manuscrit contemporain

FIGURE 2.5: Exemples de documents manuscrits.

des documents modernes, *e.g* les bases pour la compétition ICDAR en 2009 et 2011, pour la compétition ICFHR en 2010 et 2012 [14]), et d'écriture hors-ligne qui peuvent contenir des documents modernes ou anciens.

La troisième catégorie regroupe les bases dédiées à l'analyse de documents graphiques contenant, par exemple, des symboles, des graphes, ou des illustrations. Ce type de bases est varié, par exemple la base d'images de documents architecturaux [Delalandre *et al.*, 2010], la base CVC-MUSIMA [Fornés *et al.*, 2012] contenant 1000 images, la base de symboles chimiques [Nakagawa *et al.*, 2010] contenant 869 documents, ou une base de logos contant 999 images [15].

14. http://www.iapr-tc11.org/mediawiki/index.php/Datasets_List
15. http://www.eurecom.fr/~huet/work.html

2.2.3 Discussion autour de la constitution de bases d'images de documents

Le problème majeur de la constitution de telles bases est avant tout au problème de temps, le temps nécessaire à la numérisation mais aussi et surtout, au temps requis pour l'analyse et la génération de la vérité terrain associée.

Prenons l'exemple des trois bases d'images de documents imprimés UW-I, UW-II, UW-III [Phillips *et al.*, 1993] qui contiennent au total 9,5 millions caractères. Les trois bases permettent d'évaluer les performance d'OCRs. Dans l'article [Phillips, 1999], l'évaluation du temps relatif à la constitution de telles bases est donné. Les résultats montrent que le temps pour le contrôle de la qualité et l'annotation d'images représente 99% du temps total des traitements.

La complexité du contrôle de la qualité des bases dépend avant tout des informations qui doivent être acquises avec un taux de qualité suffisant pour leur exploitation. Pour chaque famille de documents, il faut sélectionner et annoter un nombre suffisant d'images permettant d'être représentatif de la réalité (ce nombre dépend bien évidemment de la variabilité des contenus). Concernant la constitution de bases d'images de documents anciens, s'ajoute une difficulté supplémentaire à savoir la prise en compte des dégradations liées au papier ou à l'encre. Il faut pouvoir non seulement sélectionner et annoter des images avec un contenu varié (structure, texte, langue, etc.), mais également un état de dégradation représentatif de la réalité (taches, papier déformé, trous, pliures, etc.).

Dans des bases d'images de documents anciens, le nombre d'exemples contenant des dégradations est important, et cela impacte le résultat des algorithmes de traitement ou d'analyse d'images. Ainsi, les études [Rice *et al.*, 1993, Blando *et al.*, 1995, Rice *et al.*, 1996, Kanungo *et al.*, 1998, Breuel, 2008, Liang *et al.*, 2008, Jacobi, 2011, Antonacopoulos, 2011] montrent que les performances d'algorithmes d'analyse de la structure diminuent quand le niveau de dégradation augmente. La combinaison de plusieurs types de dégradation dans un document diminue d'autant plus les performances. Plus les performances diminuent, plus le temps pour corriger manuellement les erreurs dans l'étape d'annotation et de transcription est grand (le cas extrême serait la ressaisie manuelle de l'ensemble du texte).

Une conséquence directe du temps nécessaire à la sélection d'images (représentatives de la réalité) et à leur annotation amène le plus souvent à limiter la taille de ces bases (d'autant plus que les documents sont anciens et donc potentiellement précieux ou rares [Kanungo et Haralick, 1996, Gang Zi, 2005, Beusekom *et al.*, 2008, Bal *et al.*, 2009]). Par conséquent, la variabilité et les dégradations dans les bases d'images réelles de documents

sont limitées. De plus, l'annotation semi-automatique ou manuelle, en plus de la question de la subjectivité, doit faire face au problème de l'erreur humaine. La qualité des bases est donc également liée à la quantité d'erreurs d'annotation qui peuvent être présentes.

La limitation de la taille des bases a des conséquences directes sur les méthodes de traitement ou d'analyse basées sur un apprentissage supervisé. Le contenu des bases d'apprentissage doit être suffisamment exhaustif pour présenter suffisament d'exemples représentatifs de la base réelle. Concrètement, les études menées dans [Baird, 2000, Mori *et al.*, 2000, Varga et Bunke, 2003a] montrent que plus la taille des bases d'apprentissage est grande et le contenu varié, plus les performances de ces méthodes sont améliorées. Des études réalisées ces 10 dernières années dans [Varga et Bunke, 2003a, Fischer *et al.*, 2013] montrent que la présence de plusieurs dégradations dans les images d'apprentissage permet d'améliorer la performance des algorithmes d'analyse de documents.

Cette section montre les difficultés et limites associées à la constitution de bases d'images réelles annotées. Une solution à ce problème est la génération d'images semi-synthétiques ou synthétiques. La notion "d'image synthétique" utilisée dès le début des années 90 désigne toute image contenant des éléments générés par ordinateur.

2.3 Bases d'images de documents semi-synthétiques et synthétiques

Dans cette section, nous définissions deux types d'images générées par ordinateur : les images semi-synthétiques et les images synthétiques. L'ambition du "synthétique" est de pallier le problème de la constitution de bases réelles annotées manuellement : diminuer le coût et le temps de leur constitution, obtenir un contenu varié, intégrer de nombreuses dégradations, etc. L'objectif est donc de générer de grandes bases d'images de documents et d'analyser ou d'améliorer les performances d'algorithmes de traitement ou d'analyse de documents.

2.3.1 Taux d'information synthétique

Dans cette section, nous essayons de caractériser les types d'images générées en introduisant la notion de "Taux d'information synthétique – TIS". Le taux d'information synthétique est le ratio d'information générée par un algorithme et injecté dans les images de documents réels. L'information générée par un algorithme est par exemple le bruit (généré par un modèle de bruit), les caractères dessinés par des éditeurs (Latex, Word-office),

TABLE 2.1: Le taux d'information synthétique (TIS) relatif à trois types d'images.

Type d'image	Réelle	Semi-synthétique	Synthétique
Taux d'Information Synthétique (TIS)	TIS $\approx 0\%$	$0 <$ TIS $< 100\%$	TIS $\approx 100\%$

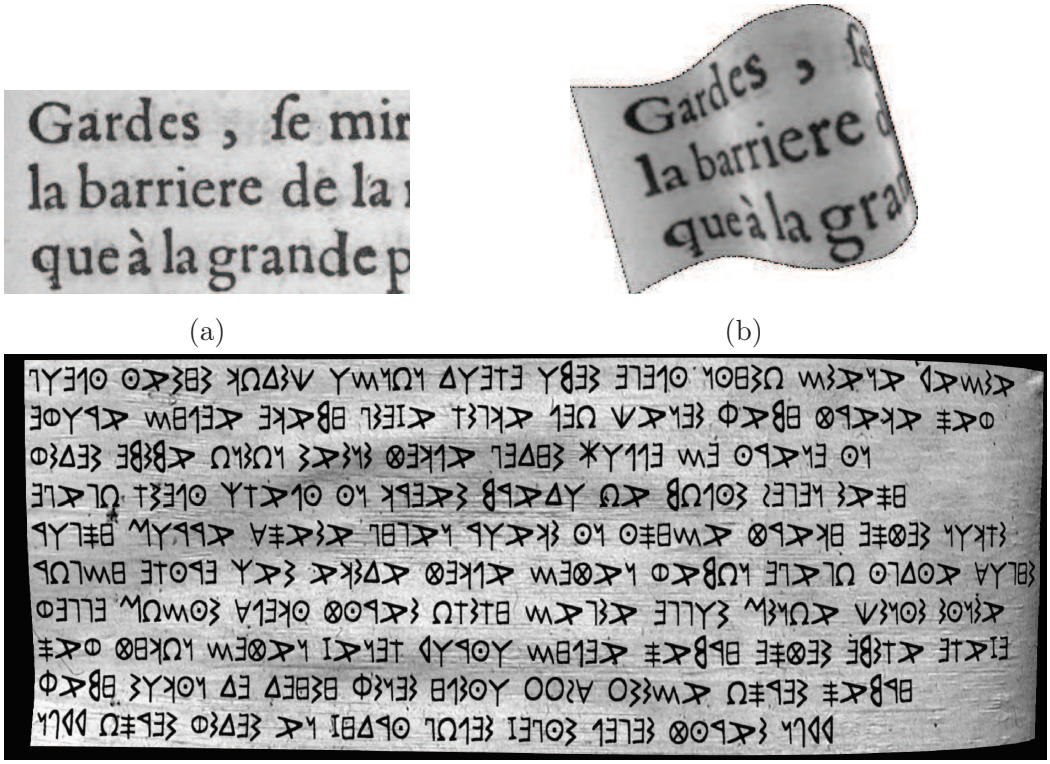
les images générées par modules de synthèse d'image (OpenGL, photoshop, 3Dsmax, etc.), etc. Par définition, le TIS dans les images de documents réels est égal à 0%. Nous considérons qu'elles sont semblables aux versions physiques, mis à part les artefacts liés à leur numérisation et les éventuels traitements postérieurs qui leur ont été appliqués.

Sur la base du taux d'information synthétique, nous classons les images en trois types (*cf.* la table 2.1) : les images réelles (TIS $\approx 0\%$) et les images semi-synthétiques (0% <TIS <100%), les images synthétiques (TIS $\approx 100\%$). Par conséquent, les images de documents semi-synthétiques sont des images dont une partie du contenu provient de documents réels, et l'autre est générée par des algorithmes. Les images de documents synthétiques sont des images dont l'ensemble du contenu est généré par des algorithmes. La figure 2.6 montre des exemples d'images appartenant à ces trois catégories. La figure 2.6-b montre une image semi-synthétique générée en dégradant l'image réelle 2.6-a. L'image 2.6-c est une image synthétique contenant un arrière-plan synthétique, des caractères générés par word-office, du bruit Gaussien, du flou et des distorsions rajoutées par ordinateur.

2.3.2 Génération de bases d'images de documents synthétiques

Cette section s'intéresse au processus de génération d'images synthétiques. Puisque les éléments d'une image synthétique sont totalement générés par ordinateur, ce processus est automatique et généralement rapide (même si l'apprentissage sous-jacent peut être long). Ce processus est montré dans la figure 2.7. D'abord, l'opérateur met en page le contenu du document en utilisant un éditeur comme Word-office ou Latex. Puis une image de document est exportée. Finalement, on applique un ou plusieurs modèles de dégradation pour générer l'image finale la plus réaliste possible. La figure 2.8-c montre une image synthétique générée par ce processus. Plusieurs générateurs sont proposés dans la littérature [Kanungo et Haralick, 1996, Ishidera et Nishiwaki, 2003, Jiuzhou, 2005, Gang Zi, 2005, Héroux *et al.*, 2007, Beusekom *et al.*, 2008] pour plusieurs types de documents comme les documents textuels, les schémas, les écritures, etc.

C'est dans ce cadre, que les auteurs de [Kanungo et Haralick, 1996, Gang Zi, 2005, Héroux *et al.*, 2007] proposent des générateurs basés sur l'éditeur LATEX pour générer des



(a) (b)

(c)

FIGURE 2.6: Exemples de trois images de documents avec un TIS différent : (a) image réelle, (b) image semi-synthétique, (c) image synthétique.

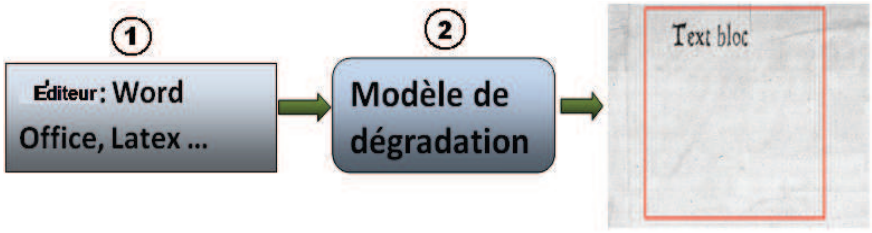


FIGURE 2.7: Le processus de génération d'images de documents synthétiques : (1) un document numérique créé par un éditeur, puis exporté sous le format d'une image (*e.g* PNG), puis (2) cette image est dégradée par un modèle pour obtenir une image de document synthétique.

(a)

(b)

(c)

FIGURE 2.8: Exemples d'images générées de manière synthétique : (a) document produit par Word-office, (b) image exportée avec ajout numérique d'un arrière-plan, (c) image dégradée.

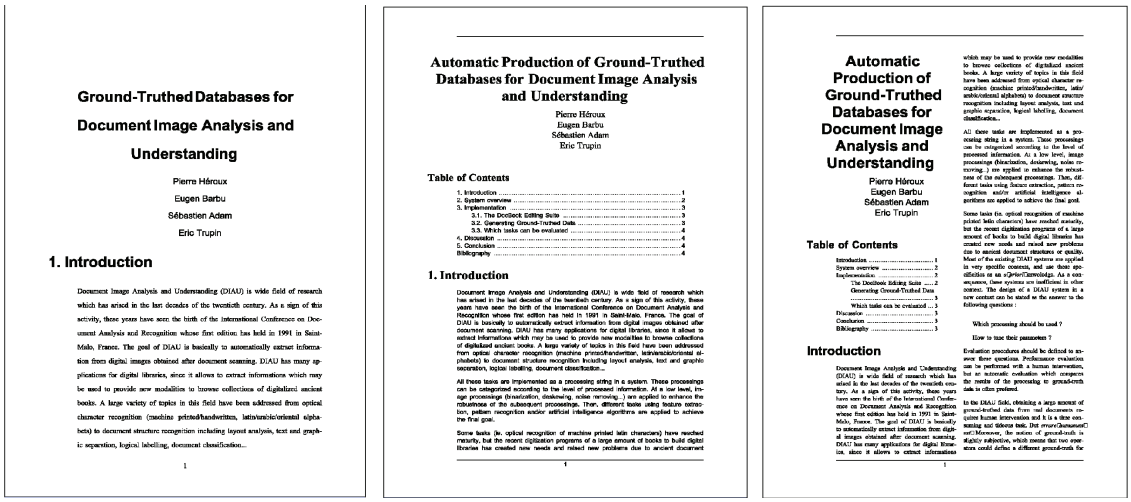


FIGURE 2.9: Trois images synthétiques créées par le générateur [Héroux *et al.*, 2007]. Elles ont le même contenu, mais avec des structures différentes.

fichiers de documents (DVI) permettant d'obtenir des images synthétiques au format TIFF (voir la figure 2.9 pour un exemple). Puis, des modèles de dégradation sont appliqués sur ces images, telles qu'une distorsion en 2D (rotation, translation), du bruit (Gaussien, bruit local [Kanungo *et al.*, 1993, Zhai *et al.*, 2003]), et du flou (flou Gaussien, flou médian). Ces images peuvent être utilisées pour évaluer des OCR.

L'étude [Beusekom *et al.*, 2008] présente un générateur d'images synthétiques de documents contemporains qui utilise un fichier PDF comme vérité-terrain. Ce fichier est imprimé pour produire un document physique. Ce document est retourné, puis scanné pour produire une image de document synthétique avec une rotation. Des bruits numériques [Baird, 1990] sont également ajoutés dans cette image pendant la numérisation.

Un générateur d'écriture manuscrite est proposé par les auteurs de [Ishidera et Nishiwaki, 2003]. Ce générateur permet de générer des mots cursifs en faisant varier l'espace inter-ligne, l'espace inter-mot, et la forme des mots. Un autre générateur est proposé dans l'étude [Graves, 2013] avec une démonstration sur le site [16]. Le modèle sous-jacent se base sur un réseau neuronal qui permet d'apprendre le style d'écriture, puis de générer des nouvelles séries de mots. La figure 2.10 montre des exemples de mots produits par ces deux générateurs.

Les études précédentes bénéficient des avantages du processus de génération d'images de documents synthétiques : rapidité et automaticité de la génération du contenu et d'une vérité-terrain. Cela permet de générer une grande base d'images synthétiques avec une va-

16. http://www.cs.toronto.edu/~graves

(a) Six images du mot "cork" synthétiques

(b) Une image de mots créés par le générateur [Graves, 2013]

FIGURE 2.10: Images d'écritures synthétiques : (a) six images du mot "cork" synthétiques créées par le générateur [Ishidera et Nishiwaki, 2003] en utilisant plusieurs styles d'écriture, (b) une image de mots créée par le générateur [Graves, 2013].

riabilité plus riche. La plupart des générateurs d'images synthétiques visent à générer des images de documents contemporains où l'arrière-plan est très clair et le premier-plan est homogène. A notre connaissance, aucune étude sur la génération d'images synthétiques ne s'adapte aux documents anciens, car ce type de document est souvent en niveaux de gris ou en couleurs avec l'arrière-plan très dégradé et le premier-plan variable. Cela pose des difficultés en terme de génération d'images totalement synthétiques.

C'est principalement pour ces raisons que certains chercheurs ont choisi de concevoir des générateurs d'images de documents semi-synthétiques. Dans la section suivante, nous allons étudier les systèmes de génération d'images semi-synthétiques ainsi que leur capacité d'adaptation aux documents anciens.

2.3.3 Génération de bases d'images de document semi-synthétiques

Par définition, une image semi-synthétique contient à la fois des éléments réels et des éléments synthétiques. La partie réelle peut être une image complète de document réel ou bien des éléments provenant potentiellement de plusieurs documents réels. Nous proposons donc de classer les méthodes de génération d'images semi-synthétiques en deux types : celles qui génèrent des images en dégradant des images réelles et celles qui combinent des éléments extraits d'images réelles.

2.3.3.1 Génération par dégradation d'images réelles

Une solution simple permettant de générer des images semi-synthétiques est de dégrader directement une image de document réel (*cf.* Figure 2.11) en utilisant des modèles de dégradation. Ces modèles permettent de reproduire des dégradations sur des documents

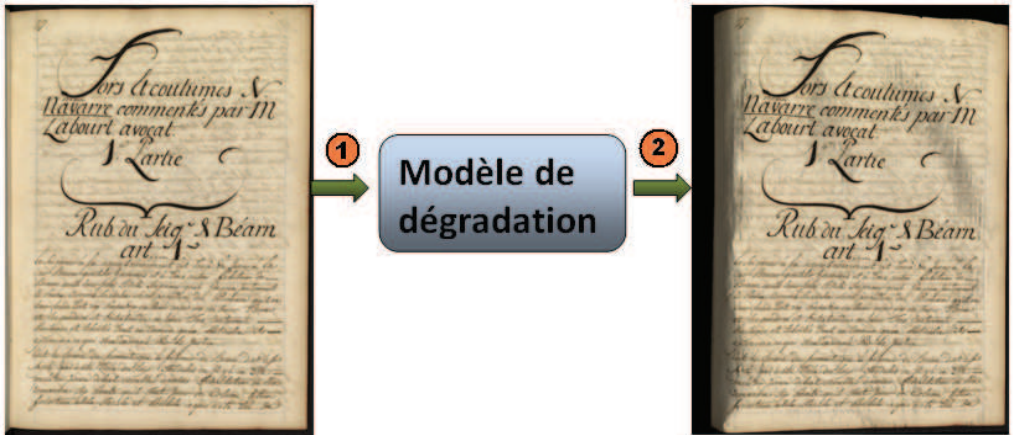


FIGURE 2.11: Dégradation d'une image de document réel par un modèle de dégradation : (1) une image de document réel en entrée, (2) une image semi-synthétique générée à partir de cette image.

réels. Cette solution a été proposée pour la première fois dans [Baird, 1990] pour créer des bases d'images de documents semi-synthétiques [Baird, 1990, Kanungo *et al.*, 1993, Phillips, 1999, Zhai *et al.*, 2003, Smith, 2008, Liang *et al.*, 2008, Moghaddam et Cheriet, 2009, Fornés *et al.*, 2011]. Elles ont la particularité d'être simples à implanter et à paramétrer, de générer rapidement des images et de fournir des images réalistes puisqu'elles sont directement issues de documents réels. La variété des images semi-synthétiques dépend évidemment de celles de la base originale, laquelle est possible à générer. La présence de défauts existants dans les documents originaux peut perturber le résultat.

Des modèles de dégradation liés à la numérisation tels que des modèles de bruits numériques d'acquisition et de flou ont été proposés dans [Baird, 1990, Kanungo *et al.*, 1993, Zhai *et al.*, 2003, Smith, 2008]. Ce type de modèles permet de générer des images binaires semi-synthétiques pour évaluer les OCR. Dans [Phillips, 1999, Liang *et al.*, 2008, Moghaddam et Cheriet, 2009, Fornés *et al.*, 2011], les dégradations sont simulées pour générer des images semi-synthétiques en niveau de gris. Ces images visent à évaluer des méthodes de restauration ou des OCR.

2.3.3.2 Génération par combinaison d'éléments extraits de documents réels

Une autre façon de générer des images semi-synthétiques est de combiner des éléments extraits de documents réels (fonte, arrière-plan, éléments de la structure) dans une nouvelle image. La figure 2.12 illustre l'ensemble de 5 étapes nécessaires à ce processus de

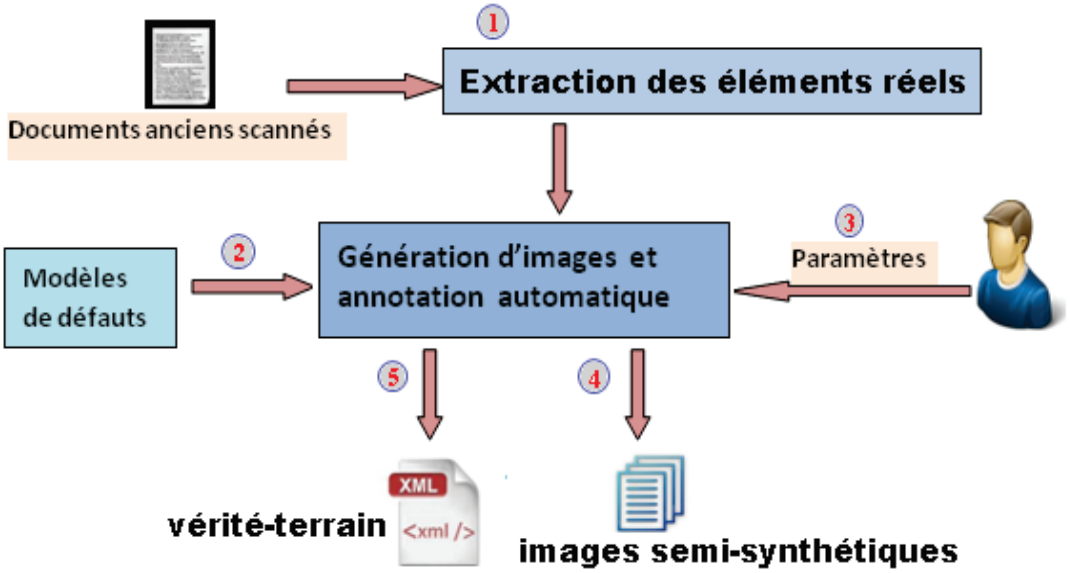


FIGURE 2.12: Schéma résumant le processus de génération de bases d'images de documents semi-synthétiques par la combinaison d'éléments réels.

génération et d'annotation automatique d'images de documents semi-synthétiques : (1) Extraction de données sources : des éléments (contenu textuel, images de fonds, illustrations, fontes, etc.) sont extraits le plus souvent manuellement et sous la forme d'imagettes, à partir de documents réels. A chaque imagette est associé un ensemble d'informations (étiquette, position et taille dans l'image d'origine, etc.). (2) Modèles de dégradation : il est possible de générer, à partir d'une image source, un ensemble de défauts simulant ceux observés dans les images réelles (transparence, courbure de page, dégradation des caractères). (3) Paramétrage des données à générer : un utilisateur peut définir de manière précise le type d'images à générer (contenu textuel, mise en page, défauts présents, niveaux de dégradations, etc.). Les étapes (4) et (5) sont celles permettant de générer les images semi-synthétiques (étape 4) et la vérité terrain associée (par exemple au format XML) (étape 5).

Il est ainsi possible de générer des images semi-synthétiques très différentes en terme de contenu physique (taille, structure, forme des éléments) et logique (style de documents, type de documents, titre, sous-titre, etc.), car l'utilisateur peut contrôler le contenu de l'image. Par conséquent, on peut ainsi générer des bases d'images de documents semi-synthétiques pour de multiples applications. Par exemple, les études [Mori *et al.*, 2000, Varga et Bunke, 2003b, Yin *et al.*, 2013] présentent des générateurs de bases d'images semi-synthétiques pour des algorithmes d'analyse d'écriture manuscrite. Les études [SLIMANE

et al., 2009, Bal *et al.*, 2009] présentent deux générateurs pour des algorithmes d'analyse de documents imprimés. Un autre générateur [Delalandre *et al.*, 2010] est appliqué sur des documents architecturaux pour évaluer des algorithmes d'analyse de symboles.

2.3.3.3 Proposition d'un générateur d'images de documents semi-synthétiques

Nous avons intégré les deux processus précédents dans notre générateur qui permet de créer des images de documents semi-synthétiques selon le fonctionnement détaillé en Figure 2.13. Deux des processus importants de notre générateur sont donc : l'extraction de fontes et de fonds (étape 1) et la génération d'images de documents semi-synthétiques (étapes 2, 3 et 4). Les étapes 2, 3, et 4 de la figure 2.13 illustrent respectivement trois façons de générer un document ancien : génération par un clavier virtuel, génération automatique à partir d'un document réel, et génération automatique par importation de fichiers texte (*.txt). Au travers de l'interface graphique, l'utilisateur renseigne plusieurs paramètres selon le rendu visuel qu'il souhaite obtenir (le texte, la fonte ancienne, le fond, la structure physique).

La figure 2.14 montre l'interface du générateur qui permet à l'utilisateur de sélectionner à la souris des caractères et des fonds d'images issus d'images de documents numérisés (étapes 1). Le générateur procède alors à l'extraction de la fonte et du fond correspondant. L'utilisateur peut ensuite générer, à l'aide de l'interface de la figure, une multitude de documents semi-synthétiques en jouant sur divers paramètres.

La suite de cette section vise à détailler l'extraction de fontes et de fonds, et la génération d'images de documents semi-synthétiques.

a. Extraction de fontes anciennes

La génération d'images semi-synthétiques et de la vérité terrain pose un double problème. Tout d'abord, il n'existe pas (ou très peu) de fontes anciennes déjà disponibles dans des éditeurs. Il n'est donc pas envisageable d'aborder le problème sous le même angle que dans le cadre de travaux sur les fontes contemporaines. Dans le même temps, il faut pouvoir créer des images reproduisant fidèlement les spécificités visuelles des documents anciens. Ainsi, la fonte n'est pas l'unique information importante. Il faut pouvoir reproduire la variabilité et la dégradation des caractères au fil des pages, le problème de l'encrage, les jeux de caractères différents pour une même lettre, les différentes dégradations du papier (taches, trous, etc.) ainsi que les déformations dues à l'étape de numérisation. Sur la base

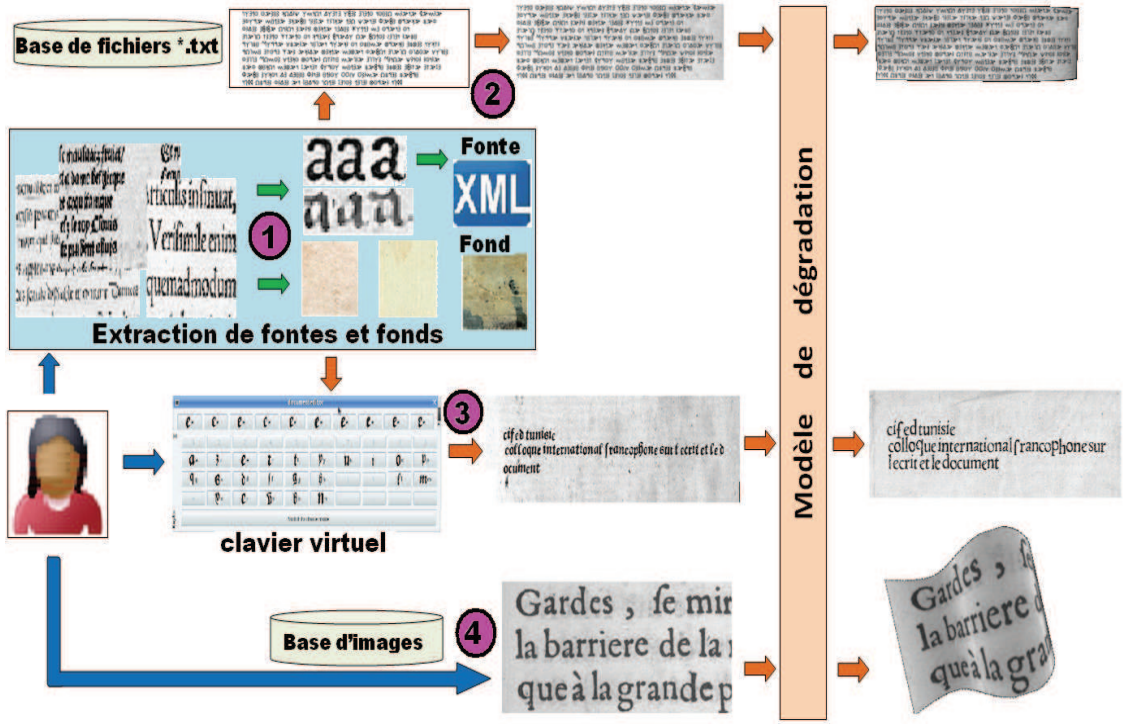


FIGURE 2.13: Schéma récapitulant le fonctionnement du processus de génération de documents anciens : l'étape (1) permet d'extraire des fontes et fonds anciens. Les étapes (2), (3), et (4) représentent les différentes approches pour la création automatique d'images de documents semi-synthétiques.

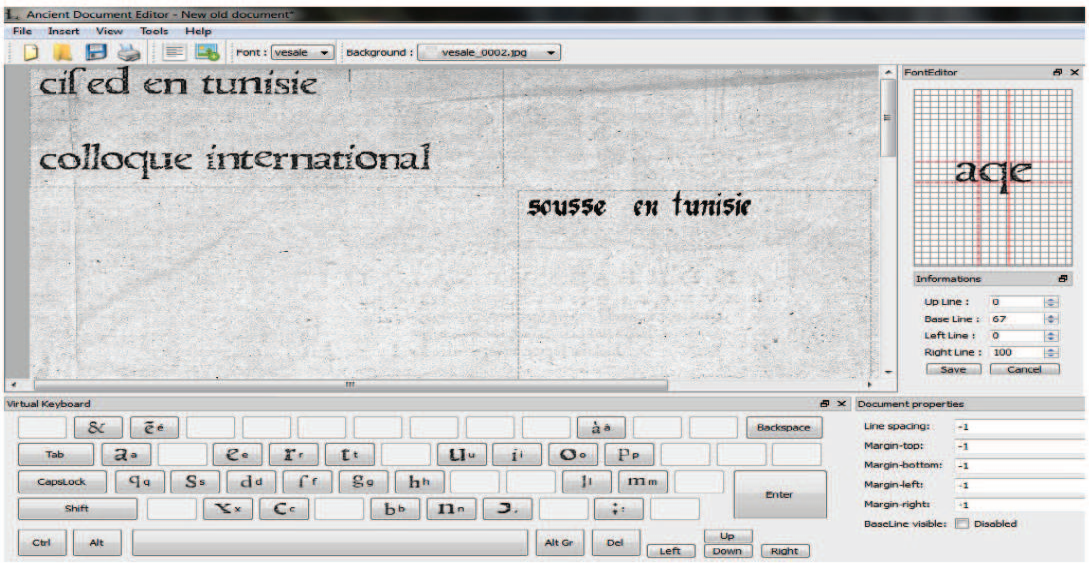


FIGURE 2.14: L'interface graphique complète de notre générateur. On y voit un document en cours de création avec le clavier virtuel.

de cette double constatation, nous avons mis en place un fonctionnement dédié à la génération d'images contenant non seulement des fontes anciennes, mais respectant également les spécificités des documents anciens.

Le générateur met à disposition de l'utilisateur une interface permettant de saisir ou de corriger à l'issue d'une phase d'extraction automatique, à la souris, des exemples de caractères de la fonte qu'il souhaite créer (l'étape 1 de la figure 2.13). Pour chaque caractère extrait, l'utilisateur indique son étiquette. On associe ainsi une étiquette à chaque forme. Lors de cette phase de saisie, une étude de la position de la composante connexe du caractère et des composantes connexes voisines permet de déterminer les points d'accroche nécessaires au futur positionnement de chaque caractère (ligne de base, placement du prochain caractère, etc.). Si les paramètres sont appris automatiquement (à partir de l'image dont sont issus les caractères), l'utilisateur peut néanmoins affiner ces différents réglages (*cf.* figure 2.15). Sur le même modèle, l'utilisateur peut extraire divers types de fonds d'images de documents anciens. Cela lui permet, par exemple, d'extraire des fonds visuellement variés (fonds bruités, fonds très clairs ou très sombres, etc.).

A la fin de l'étape 1, l'utilisateur a donc saisi pour chaque fonte ancienne qu'il souhaite générer plusieurs exemples de chaque lettre. Il a également extrait plusieurs types de fonds.

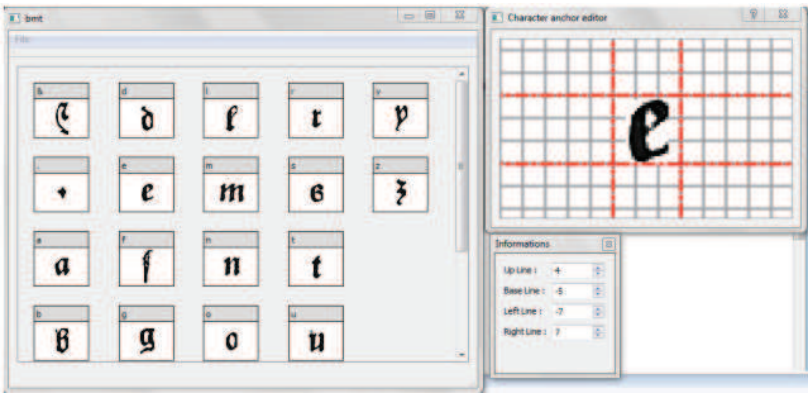


FIGURE 2.15: L'éditeur de fontes anciennes. Sur la partie de gauche sont situés les caractères déjà appris par le système avec l'aide de l'utilisateur. Sur la partie de droite se trouve la fenêtre s'ouvrant lors de la saisie d'un exemple de caractère. Les traits rouges symbolisent les 4 points d'accroche calculés automatiquement. Ces paramètres peuvent être modifiés en déplaçant à la souris les traits rouges ou en indiquant directement leurs valeurs dans les champs situés en dessous.

Après cette phase de saisie, l'ensemble des informations extraites est répertorié dans un fichier `XML` (l'étape 1 de la figure 2.13). On y retrouve, entres autres, le nom de la fonte et les caractères extraits auxquels sont associées des informations de position (points d'accroche). On trouve également dans ce fichier `XML` des informations sur les espaces inter-lignes, inter-mots et inter-caractères de la fonte extraite.

Cette phase de saisie aboutit également à la création d'un ensemble d'imagettes ; plusieurs images par caractère et une image par fond. Dans sa version actuelle, notre générateur intègre dix fontes anciennes déjà apprises, cinquante fonds en niveau de gris (*cf.* la figure 2.16), et six modèles de dégradation (les deux modèles de [Kanungo *et al.*, 1993], le modèle de transparence de [Moghaddam et Cheriet, 2009], le modèle de distorsion 2D de [Liang *et al.*, 2008], et les deux modèles de dégradation que nous proposons et qui seront détaillés dans les chapitres suivants : le modèle de bruit local et le modèle de distorsion 3D. Cela nous permet d'obtenir une base d'images de documents semi-synthétiques d'apparences variées.

b. Création d'images de documents semi-synthétiques

L'interface de la figure 2.14 propose trois façons de générer des documents anciens

FIGURE 2.16: Exemples de fonds extraits de documents réels intégrés dans notre générateur

(étapes 2, 3, et 4 de la figure 2.13) : génération *via* un clavier virtuel en utilisant des fontes et fonds intégrés au générateur, génération automatique par la dégradation d'un document réel, et génération automatique *via* des fichiers texte (*.txt). Au travers de l'interface graphique, l'utilisateur renseigne plusieurs paramètres selon le rendu visuel qu'il souhaite obtenir (le texte, la fonte ancienne, le fond, la structure physique, la présence et la localisation d'illustrations, etc.).

La première façon de générer un document ancien est d'utiliser un clavier virtuel (*cf.* figure 2.17). Pour chaque fonte, plusieurs exemples de chaque lettre sont disponibles. La figure 2.17 illustre la variété de lettres "e" pouvant exister dans une police donnée. Lors de la génération de l'image finale, chaque imagette de caractère est choisie aléatoirement parmi l'ensemble des caractères extraits en phase 1. Ce choix permet d'obtenir un résultat final d'apparence plus réaliste pour un document ancien, avec des caractères qui ne sont pas tous identiques entre eux.

La figure 2.18 illustre le résultat de la génération d'un texte selon la fonte et le fond choisis. Ces images ont été générées sans que l'humain ne corrige les points d'accroche avec l'éditeur de fontes anciennes (voir la figure 2.15). Les résultats de la figure 2.18.a-b sont donc obtenus entièrement automatiquement à partir des polices extraites. Si le visuel est

FIGURE 2.17: Clavier virtuel mis à disposition de l'utilisateur. Celui-ci peut choisir la fonte du texte et le fond de l'image. Un écran lui permet de prévisualiser la disposition des caractères en fonction des paramètres de position (points d'accroche) préalablement appris. L'utilisateur a la possibilité de choisir entre plusieurs exemples d'un même caractère (ici le dernier "e" de "exemple"), ou de laisser le système décider aléatoirement de l'imagette à utiliser.

relativement correct, on peut néanmoins remarquer que les caractères "q" et "h" ont mal été positionnés. Il faudrait donc qu'un utilisateur règle le positionnement qui a été initialement calculé avec l'éditeur de fontes anciennes afin d'obtenir un meilleur positionnement (voir la figure 2.18.c).

La deuxième façon d'obtenir une image de document semi-synthétique est de dégrader directement une image de document réel (étape 4 de la figure 2.13). Par exemple, une image de document réel est dégradée par le modèle de distorsion en 2D [Liang *et al.*, 2008] intégré dans le générateur. Si on dispose d'une vérité-terrain associée l'image réelle originale, alors on peut généralement en déduire assez facilement la vérité-terrain associée à l'image dégradée comme montré dans la figure 2.19.

Pour varier les contenus des images générées et accélérer la vitesse du processus de génération d'images semi-synthétique, dans l'étape 2 de la figure 2.13, nous avons implémenté une troisième manière de générer des images de documents semi-synthétiques à partir des fichiers de type texte. Ce processus est illustré dans le schéma 2.20. Tout d'abord, la fonction lit les fichiers `*.txt` pour retrouver les codes de caractères de la table UNICODE (*cf.* étape 1 dans le schéma 2.20). Puis, on utilise une fonte préalablement apprise pour mettre en correspondance les imagettes de caractères avec ces codes. Finalement, on applique des modèles de dégradation sur l'image de la page pour obtenir une image semi-synthétique dégradée (*cf.* l'étape 3 dans le schéma 2.20) et sa vérité-terrain associée. Les choix de fontes, de fonds, de structure physique et logique et de niveau de dégradation sont

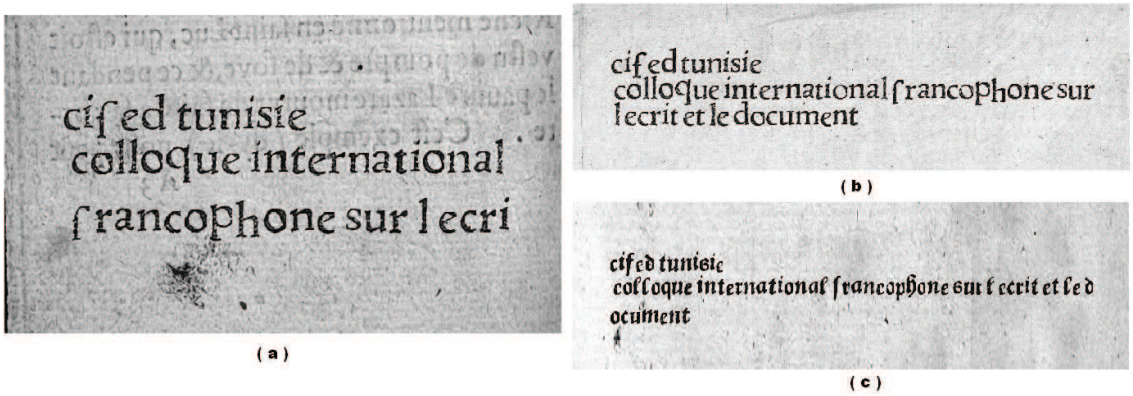


FIGURE 2.18: Images de documents anciens générés automatiquement. L'image (a) a été générée avec un fond bruité. Les deux images (b) et (c) ont été générées avec deux fontes différentes. L'image (c) est produite avec l'intervention de l'utilisateur pour corriger les points d'accroche.

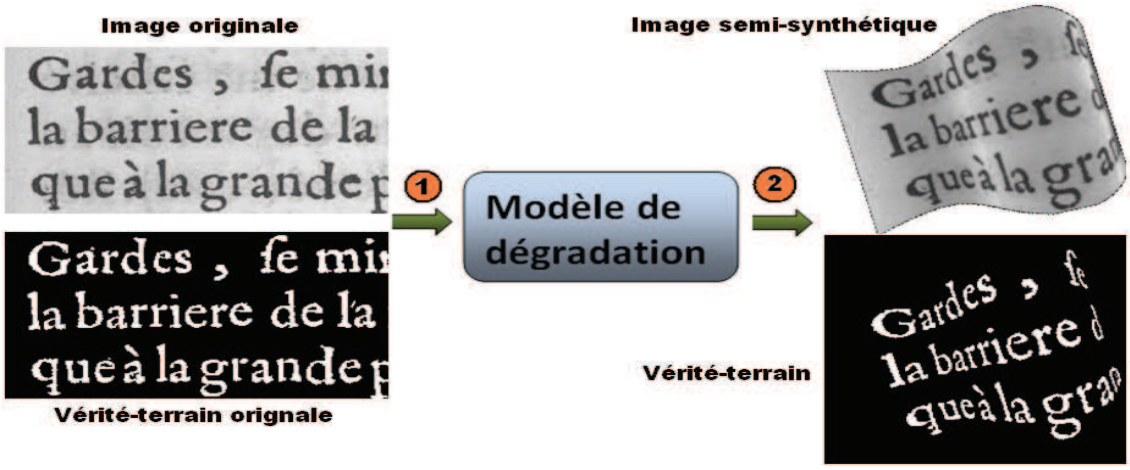


FIGURE 2.19: Le processus de génération d'images semi-synthétiques par dégradation d'une image de document réel. Ici la vérité-terrain est l'image binaire.

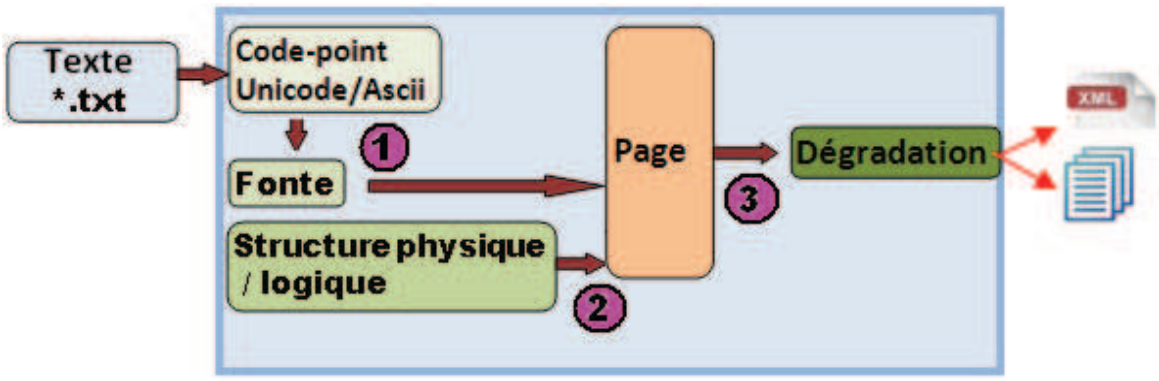


FIGURE 2.20: Génération d'images de document semi-synthétiques à partir de texte : l'étape (1) met en correspondance un caractère UNICODE dans le fichier texte au caractère correspondant dans la fonte extraite choisie, l'étape (2) met en page le texte, l'étape (3) dégrade la page générée.

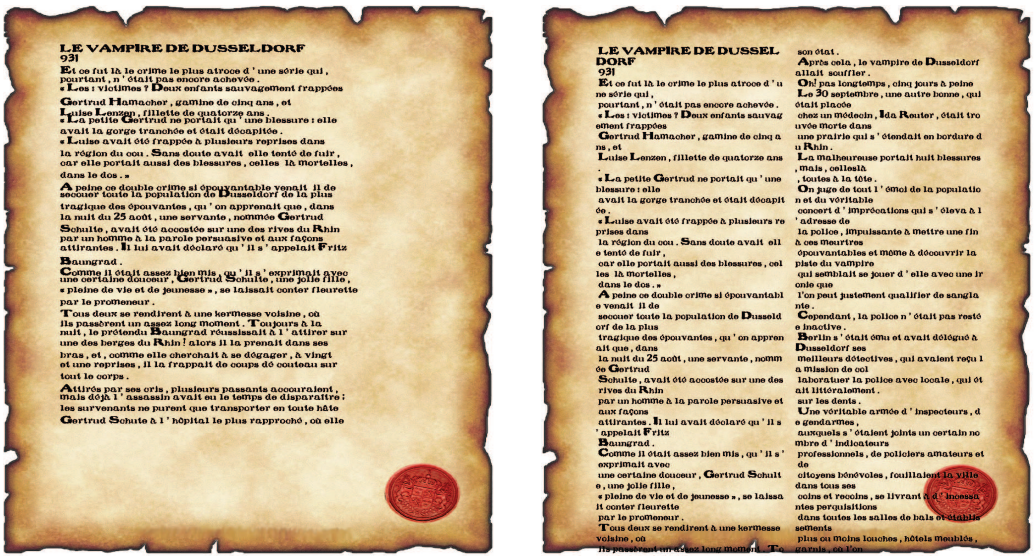
fixés par l'utilisateur au début du processus.

La figure 2.21 montre deux images de documents semi-synthétiques générées automatiquement. Elles contiennent le même texte en provenance d'un même fichier et sont générées avec la même fonte et le même fond. Le texte de l'image 2.21-a est mis en page sur une colonne, tandis que ce même texte est mis en page sur deux colonnes dans l'image 2.21-b.

2.4 Conclusion du chapitre

Nous avons présenté, au début du chapitre, le processus de constitution de bases d'images de documents réels et les difficultés associées. Ces bases d'images de documents sont utilisées pour évaluer les performances des algorithmes d'analyse de documents, mais aussi pour alimenter des modules d'apprentissage. Dans la suite du chapitre, nous avons détaillé plusieurs bases d'images de documents réels et montré que cette constitution et le travail de saisie de la vérité terrain associée est généralement fastidieux et coûteux, ceci d'autant plus que les documents sont anciens et dégradés.

Nous avons ensuite montré que la génération d'images de documents semi-synthétiques et/ou synthétiques pouvait être une alternative puisqu'elle permet de générer des images de manière rapide tout en assurant un contenu varié. Si beaucoup de travaux ont été menés en ce sens ces dernières années, peu de solutions ont été proposées pour la génération d'images de documents anciens.



(a) (b)

FIGURE 2.21: Exemple de génération automatique de documents semi-synthétiques à partir de texte : (a) mise en page en une colonne, (b) mise en page en deux colonnes.

Le travail réalisé dans cette thèse se focalise sur la génération et l'utilisation d'images de documents semi-synthétiques. Ce chapitre a permis de détailler un premier apport allant en ce sens. Nous avons développé un générateur de bases d'images semi-synthétiques adapté aux documents anciens.

Les chapitres 3 et 4 présentent des modèles de dégradation permettant d'altérer une image réelle et de générer des défauts similaires à ceux observés dans les documents anciens réels. Ces modèles sont intégrés à notre générateur présenté dans ce chapitre.

Notre ambition dans le chapitre 5 sera de montrer que ce générateur est utile pour des étapes d'évaluation de performances et/ou d'apprentissage.

Chapitre 3

Les modèles de dégradation d'images de documents

Dans le chapitre précédent, nous avons montré l'intérêt d'utiliser la génération d'images semi-synthétiques. Cette génération permet de créer un nombre conséquent d'images contenant des exemples représentatifs du contexte réel. Dans ce chapitre, nous présentons en détais plusieurs modèles de dégradation. Ces modèles, intégrés dans le générateur d'images, permettent de simuler certaines des dégradations présentes dans un document réel. Dans la section 3.1 nous commençons par présenter les dégradations les plus couramment observées. Cette première étude permet de montrer dans quelles mesures ces dégradations peuvent impacter le bon fonctionnement d'algorithmes de traitement ou d'analyse de documents.

3.1 Dégradations présentes dans les documents anciens

Les dégradations peuvent être assimilées à un processus d'ajout d'altérations ou de bruits pouvant intervenir à n'importe quel moment depuis la conception du document original, jusqu'à la phase de numérisation. Plusieurs types d'altérations peuvent être introduits dans les images de documents. Les dégradations peuvent être classées en deux catégories : les dégradations intrinsèques au document et celles dues à la numérisation. Cette classification sera détaillée dans la sous-section 3.1.1.

Les dégradations peuvent apparaître n'importe où dans une image de document. Nous avons décidé de catégoriser ces défauts selon leur localisation, leur forme, et leur rendu visuel. Nous proposons donc trois catégories de dégradations : globales, locales, et diffuses. Les dégradations globales se retrouvent sur l'ensemble des éléments du document (distorsion globale du papier, luminosité non uniforme, etc.). Les dégradations locales sont celles affectant quelques éléments du document : déformation des caractères, trous dans l'encre, taches, etc. Les dégradations diffuses se rapportent plus à une notion d'impression visuelle. Nous parlons de dégradation diffuse pour des défauts tels que la transvision (apparition du verso sur le recto par transparence). L'encre apparait de manière diffuse. De la même manière, certaines dégradations telles que des taches d'humidité apparaissent de manière diffuse. Une dégradation diffuse peut être locale ou globale. L'impact de ces trois types de dégradation sur les performances des algorithmes d'analyse d'images de documents sera présenté dans la sous-section 3.1.2.

3.1.1 Différentes catégories de dégradations

3.1.1.1 Dégradations intrinsèques au document lui même et à sa condition de conservation

Que ce soit lors de sa création (qualité du papier, de l'encre ou de la presse), de son utilisation (transport, lecture, ajout d'annotations, etc.) ou de sa conservation sur plusieurs siècles (condition de stockage et de manipulation) un ouvrage peut se retrouver altéré par plusieurs types de défauts. La photo 3.1 montre un cas extrême d'un ouvrage ancien très fortement dégradé.

FIGURE 3.1: Exemple d'un ouvrage ancien très fortement dégradé

La figure 3.2 illustre des cas moins extrêmes et plus fréquemment rencontrés de défauts locaux, présents sur les bords des pages. Ce sont des pages contenant des zones avec de petites déchirures ou noircies par l'humidité.

FIGURE 3.2: Exemples de défauts présents à la bordure des documents.

La figure 3.3 illustre les défauts de type diffus liés à l'encre. Les images 3.3.a-d présentent des larges taches sombres, dues à l'humidité, et localisées sur les zones de texte. La figure 3.3.e, présente un cas de transvision.

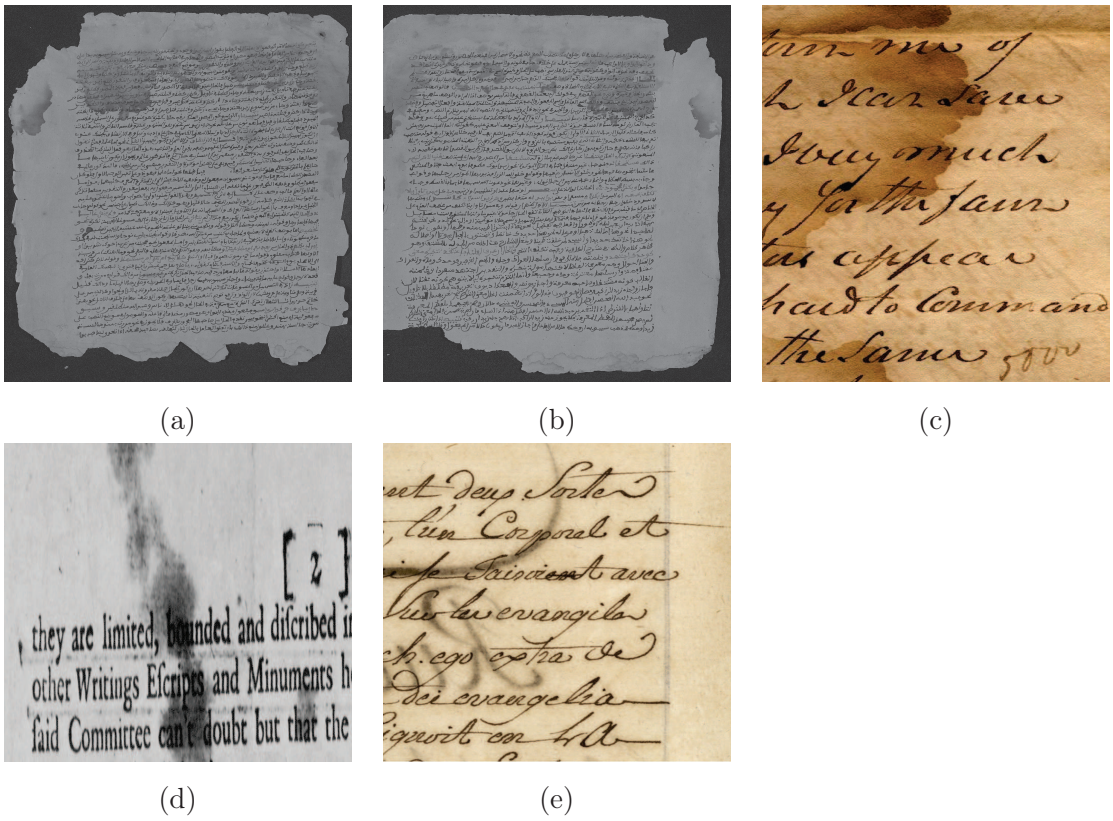


FIGURE 3.3: Des défauts liés à l'état des couleurs : (a), (b), (c) et (d) des taches d'humidité se sont créées sur le côté du livre ; (e) transparence de l'encre.

La figure 3.4 présente des documents contenant des distorsions globales (courbure du papier) et des distorsions locales (petits plis de différentes formes).

3.1.1.2 Dégradations dûes à l'étape de numérisation

Sur des étapes composant un cycle de numérisation d'un ouvrage, des défauts liés à l'acquisition peuvent apparaître. De mauvais réglages du système optique, mauvais choix de résolution et balances des blancs peuvent être à l'origine de l'apparition de défauts dans l'image numérique. De mauvaises manipulations du document physique peuvent également être à l'origine de dégradations : mauvais placement sur le scanner, mouvement pendant la numérisation. Enfin, l'évolution du contexte de numérisation peut également altérer l'image, par exemple le changement de luminosité dans la salle. Les figures 3.5 et 3.7 illustrent certains de ces défauts.

Parmi les défauts principaux dûs au processus de numérisation, nous notons :

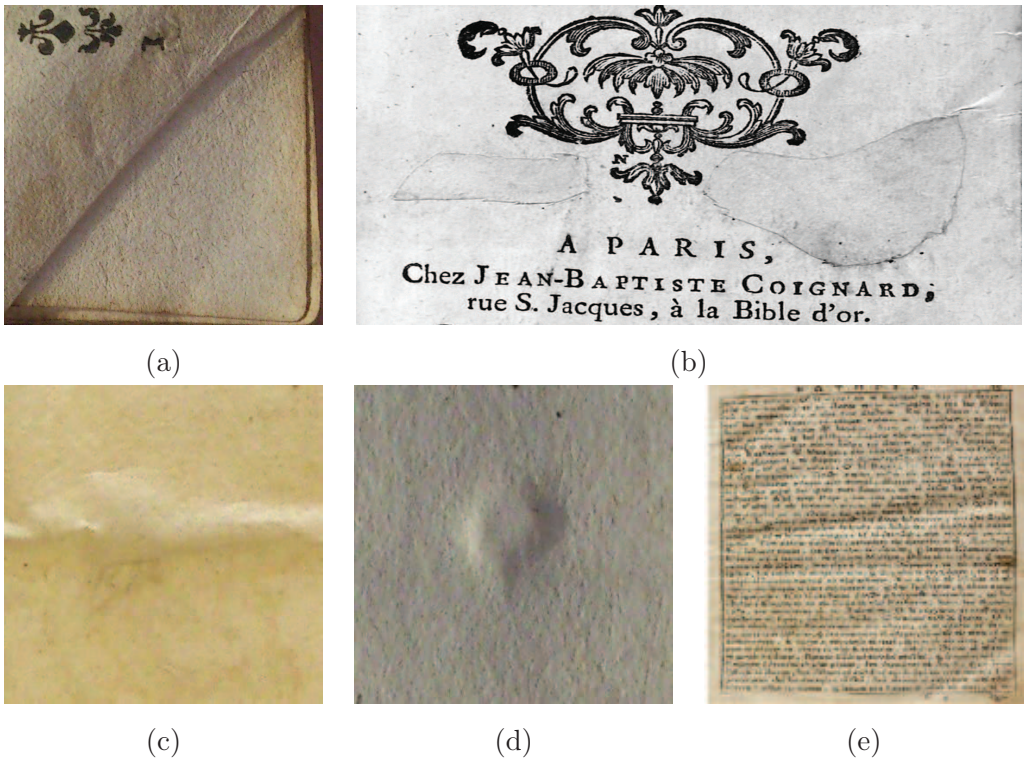
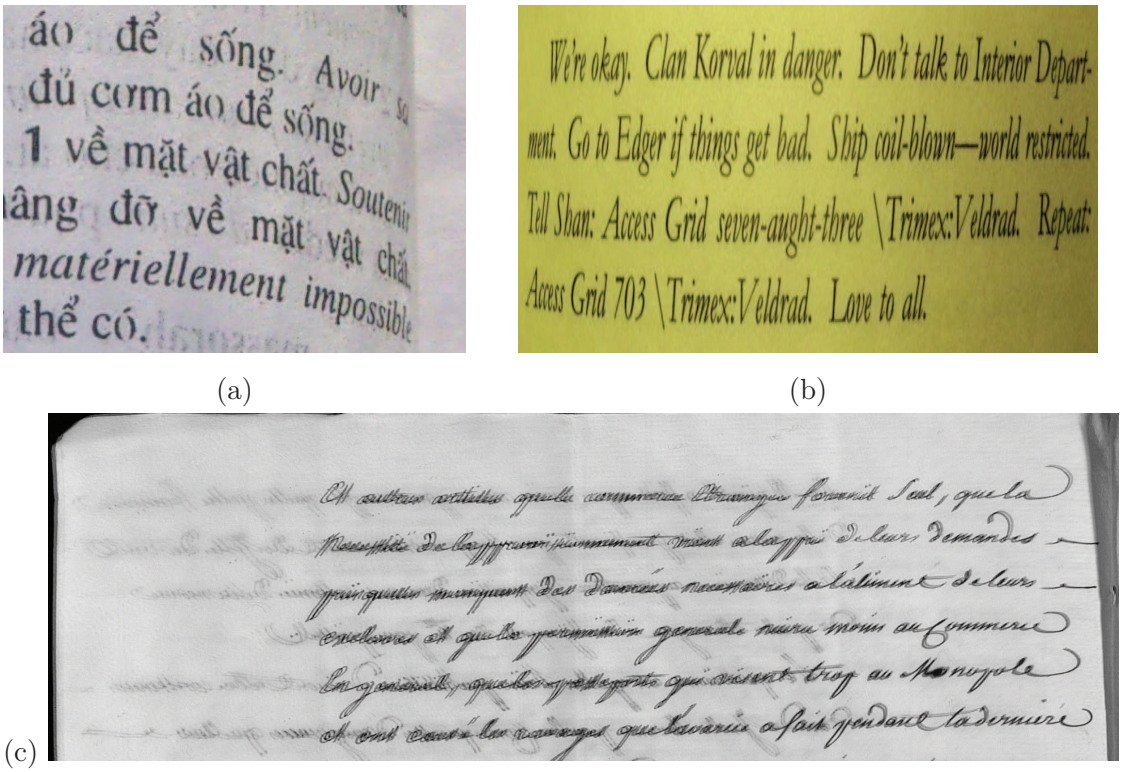


FIGURE 3.4: Des défauts liés la forme physique de la page : (a) page cornée, (b) pli transversal parcourant toute la page, (c) petite déformation convexe du papier, (d) petite déformation concave du papier, et (e) déformations globale à toute la page du support papier. Ces documents dégradés ont été aimablement donnés par la Bibliothèque Nationale de France afin d'illustrer notre propos.

- un mauvais choix de la technologie du scanner pour numériser un document avec reliure épaisse : il est difficile de les positionner "à plat" sur la vitre du scanner à plat. Il en résulte un impression visuelle de courbure de la page (figure 3.5-a).
- la distorsion de perspective : la surface de la page n'est pas totalement plane. Cela peut poser un problème au système optique (la profondeur de champs dans figure 3.5-b).
- une mauvaise gestion de la luminosité ambiante : ce défaut provient des changements de luminosité de l'éclairage ambiant autour du scanner ou lorsque la distorsion du support papier est importante. Par exemple, dans la figure 3.5-b, la luminosité au milieu est plus claire que celle du côté gauche ou droit.
- le flou : il peut être le résultat de plusieurs mauvais réglages. Par exemple, un



(a) (b)

(c)

FIGURE 3.5: Des défauts liés au processus de numérisation : (a) mauvaise gestion de reliures épaisses, (b) distorsion de perspective due au système optique, (c) flou résultant de mauvais réglages.

mauvais réglage du focus ou un mouvement du document / de la caméra pendant sa numérisation. La figure 3.5-c en montre un exemple.

– le bruit numérique : ce bruit est un signal ajouté aléatoirement à l'image de sortie. Par exemple, le bruit thermique lié à la température de capteurs ou le bruit de binarisation (*e.g.* la figure 3.6).

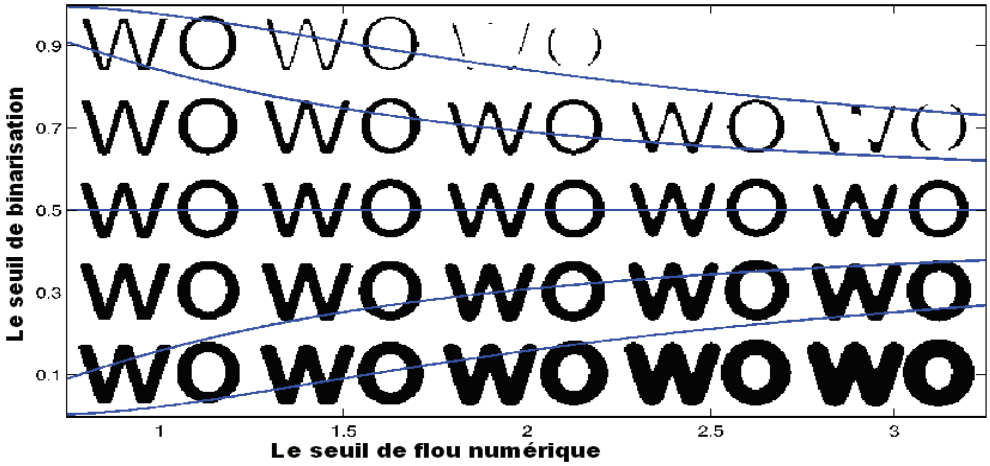
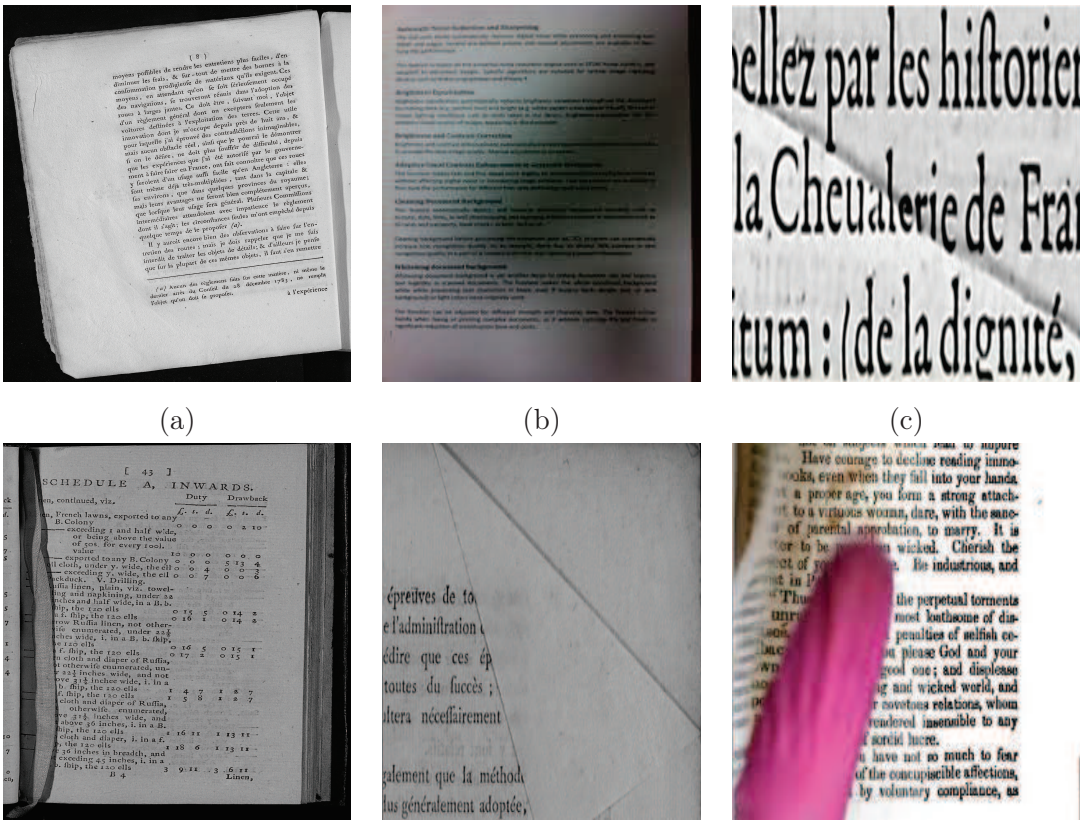


FIGURE 3.6: Exemple de bruit de binarisation et de flou numérique présentés dans l'article de [Hale et Barney Smith, 2007].

L'opérateur de numérisation peut également être à l'origine de défauts plus ou moins importants. Les plus classiques sont :

– Le mauvais positionnement. Il se concrétise par une rotation ou une translation de l'ouvrage lorsque l'opérateur fait tourner les pages de l'ouvrage (exemple figure 3.7-a).
– L'apparition d'une ombre quand l'opérateur est proche de la source lumineuse, *e.g.* la figure 3.7-b en est une illustration.
– Page pliée suite à une mauvaise manipulation. Les figures 3.7.c et 3.7.e en sont deux exemples.
– Oubli d'un objet sur l'ouvrage (figure 3.7.d)
– Présence de doigts. Ceci peut arriver lorsque l'opérateur doit mettre sa main sur la page pour corriger la distorsion (exemple figure 3.7-f).



(a) (b) (c)

(d) (e) (f)

FIGURE 3.7: Des défauts dûs à l'opérateur : (a) document mal orienté, (b) ombre de l'opérateur, (c) page pliée par l'opérateur (d) l'opérateur a oublié d'enlever le marque-page, (e) l'opérateur aurait dû décorner le document, (f) l'opérateur a oublié d'enlever son doigt. Les images ont été numérisées et fournies par la société Arkhénum.

3.1.2 Influence des dégradations sur un système d'analyse de documents

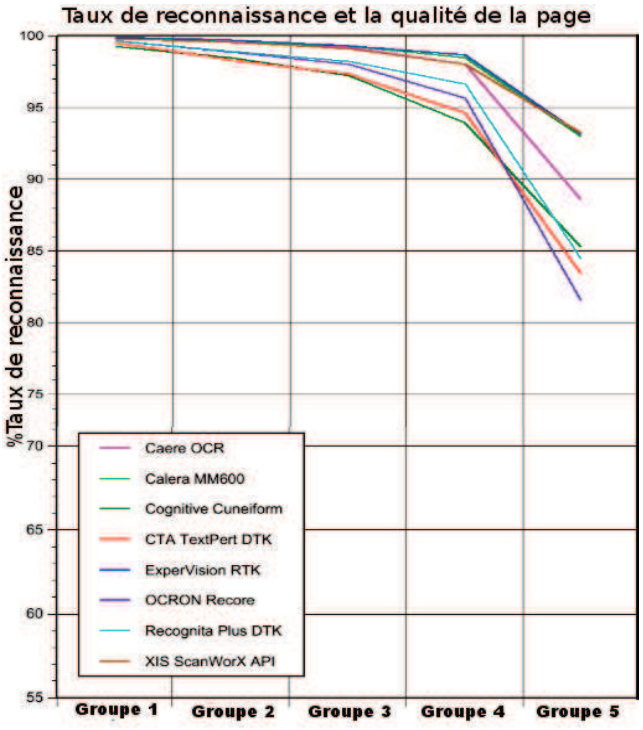
A l'issue de cette phase de numérisation du document, les images sont intégrées dans le système d'analyse de documents. Les études menées dans [Rice *et al.*, 1993, Blando *et al.*, 1995, Rice *et al.*, 1996, Kanungo *et al.*, 1998] ont montré que la qualité et la variabilité d'une image sont des facteurs donnant lieu à la baisse de performances d'un algorithme. Dans cette sous-section, nous étudions l'influence des dégradations sur la chaîne globale de traitement puis, sur chacun des maillons de cette chaîne.

3.1.2.1 Influence sur la chaîne globale

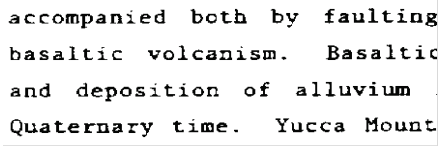
Un système d'analyse d'images de documents regroupe un ensemble d'algorithmes de traitement d'images (pré-traitement, analyse de la structure, reconnaissance, indexation cf. figure 2.1) afin d'extraire des informations le plus souvent utilisées pour l'indexation. Par exemple, un OCR enchaine les pré-traitements, la binarisation, l'extraction de régions texte et non-texte, l'extraction de caractères dans les régions de texte, et finalement un moteur de reconnaissance de caractères [Eikvil, 1993].

Plusieurs études [Rice *et al.*, 1993, Blando *et al.*, 1995, Rice *et al.*, 1996, Kanungo *et al.*, 1998, Antonacopoulos, 2011] ont permis d'évaluer les performances de nombreux OCR existants. Dans tous ces travaux, les images sont reparties selon différents niveaux de dégradation. En général, les études montrent que les performances sont sensibles, entre autre, à plusieurs types de dégradations et baissent quand le niveau de dégradation augmente.

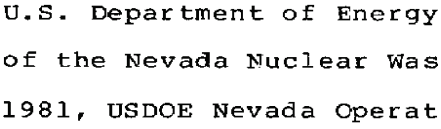
La première étude sur des systèmes OCR a été réalisée en 1993 par [Rice *et al.*, 1993]. Les auteurs ont testé 8 systèmes OCR. La figure 3.8-f donne les précisions de ces 8 systèmes sur 5 groupes d'images de documents contemporains contenant des dégradations, y compris la rotation, le flou et le bruit local (les figures 3.8.a-e montrent 5 images d'exemple de ces groupes). Le niveau de dégradation est visuellement augmenté entre le premier et le dernier groupe. C'est probablement la raison pour laquelle la moyenne du taux de reconnaissance des 8 OCR chute à partir du groupe de qualité numéro 3. Le manque d'un indicateur quantitatif du niveau de dégradation et le mélange des dégradations dans un groupe d'images ne nous permet pas de déduire les liens entre dégradations et performance d'un système de type OCR.



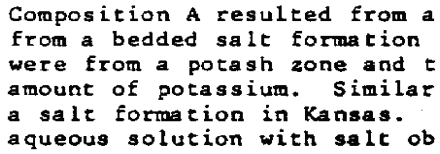
(f) Précision de 8 OCRs



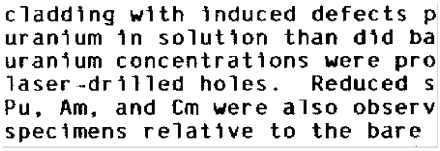
(a) groupe 1



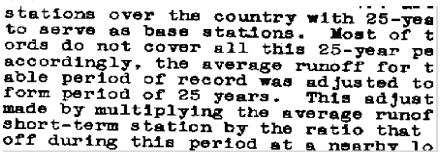
(b) groupe 2



(c) groupe 3



(d) groupe 4



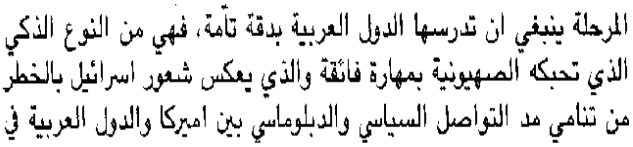
(e) groupe 5

FIGURE 3.8: Les résultats des tests obtenus par S.V Rice et al. en 1993 sur 8 systèmes OCRs [Rice *et al.*, 1993].

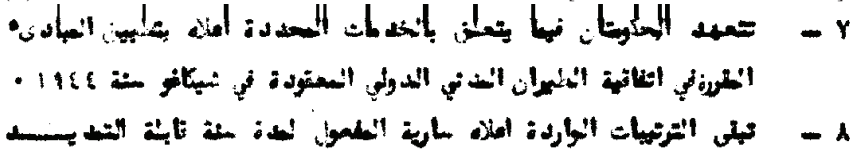
La deuxième étude réalisée par [Blando *et al.*, 1995] évalue les performances de 6 systèmes OCR. Deux bases d'images de documents contemporains, une de 460 documents scientifiques/techniques et l'autre de 200 documents de magazines, sont utilisées. Les documents issus des deux bases contiennent des taches sombres ou claires qui peuvent modifier ou couper la connectivité d'un caractère. Un classifieur essaie de trier les documents en deux catégories : bonne qualité ou mauvaise qualité selon trois heuristiques. Les deux premières heuristiques se basent sur le nombre de taches sombres connectées sur la bordure de caractères et le nombre de taches claires qui coupent la connectivité de caractères. Basé sur cette classification, l'algorithme permet de prédire les performances de six OCR. Les résultats

de cette étude montrent que le système peut prédire correctement le taux OCR pour 85% d'images dans la base de test grâce à cette classification. Cette classification peut permettre aux OCR de s'apdater aux dégradations de chacune des catégories en changeant leurs paramètres. Néanmoins, ce travail ne nous permet pas de déduire le lien entre la performance d'un OCR et chaque catégorie.

Dans l'étude de [Kanungo *et al.*, 1998], deux systèmes OCR (Sakhr OCR et OmniPage OCR) sont comparés sur leur capacité à reconnaitre du texte arabe. Ils sont testés sur une base d'images contenant du bruit local [Kanungo *et al.*, 1993]. L'intérêt de cette étude est de montrer que ces OCR sont influencés fortement par le bruit local. Par exemple, la figure 3.9-a présente une image de bonne qualité où les deux OCR ont de bonnes performances (98,08% pour Sakhr et 97,7% pour OmniPage). La figure 3.9-b montre une image de mauvaise qualité où les performances de deux OCR sont faibles (38,88% pour Sakhr et 35,79% pour OmniPage).



(a) une image de "bonne qualité" au sens de [Kanungo *et al.*, 1998]



(b) une image de "mauvaise qualité" au sens de [Kanungo *et al.*, 1998]

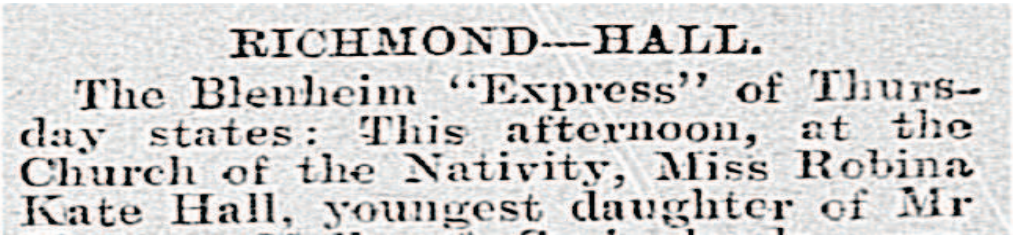
FIGURE 3.9: Deux images de deux catégories différentes dans les travaux de [Kanungo *et al.*, 1998] : (a) bonne qualité et (b) mauvaise qualité.

L'étude de [Liang *et al.*, 2008] permet d'estimer la baisse de performances du système OCR OmniPage par rapport à l'influence des distorsions globales sur des documents contemporains. Le système OmniPage est appliqué sur deux bases d'images. L'une contient 60 images de documents plats, l'autre contient 60 images de documents courbes dont les pages sont bombées (nous appelons ce type de documents : documents courbes). Les résultats présentés dans la table 3.1 montrent que les documents contenant des distorsions donnent lieu à des baisses significatives de la performance du système OCR.

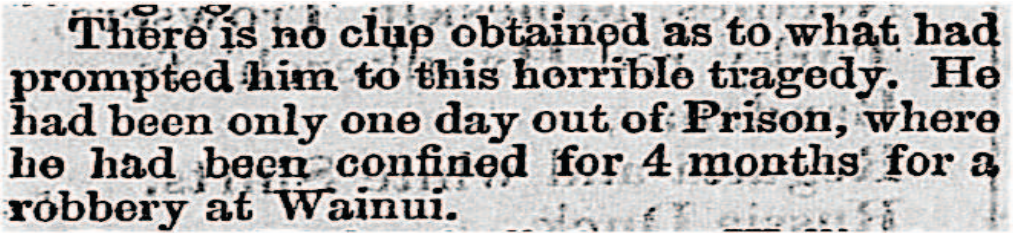
TABLE 3.1: Taux de reconnaissance du système OmniPage sur les documents plats et sur les documents déformés

Taux de reconnaissance (TR)	document plat	document courbe
TR de caractères	26,14%	23,05%
TR de mots	22,92%	14,29%

Deux autres études ont récemment été menées. La première étude, issue du projet du projet IMPACT [Antonacopoulos, 2011], présente deux expérimentations. La première expérimentation étudie la précision de l'OCR en fonction de la résolution du scanner (300dpi, 400dpi puis 500dpi). Une résolution de 300pi donne les meilleurs résultats. La seconde expérimentation, à la résolution de 300 dpi, a permis de comparer les résultats selon trois types d'images : image en couleur, image en niveaux de gris, et image binaire. Le moteur d'ABBY donne les meilleurs résultats avec les images en niveaux de gris. Cette étude donne comme consigne aux utilisateurs de choisir le bon format d'images d'entrées (images en niveaux de gris et résolution à 300dpi) pour obtenir un bon taux de reconnaissance. Néanmoins, les images utilisées contiennent plusieurs types de dégradations comme montré dans la figure 3.10. La présence de dégradations peut perturber les résultats obtenus.



(a) L'image contient des taches claires qui coupent des caractères



(b) L'image contient des taches sombres qui lient des caractères

FIGURE 3.10: Images contenant des dégradations utilisées dans l'étude [Antonacopoulos, 2011].

La deuxième étude a été réalisée par la Bibliothèque nationale de France (BnF). Il

s'agit d'étudier le comportement de systèmes OCR par rapport documents anciens (sur une période allant de 1493 à 1973). Les résultats présentés dans la figure 3.11 montrent que plus le document est ancien, plus le taux de reconnaissance est bas. Cette tendance peut être expliquée par la relation entre la date d'édition et la qualité d'image (les documents les plus anciens sont souvent les plus dégradés). Il est à noter que les dégradations ne sont pas le seul élément responsable du taux de reconnaissance d'un OCR. D'autres facteurs interviennent : le contexte d'impression, la fonte du texte, la structure, la langue, la présence d'illustration, etc.

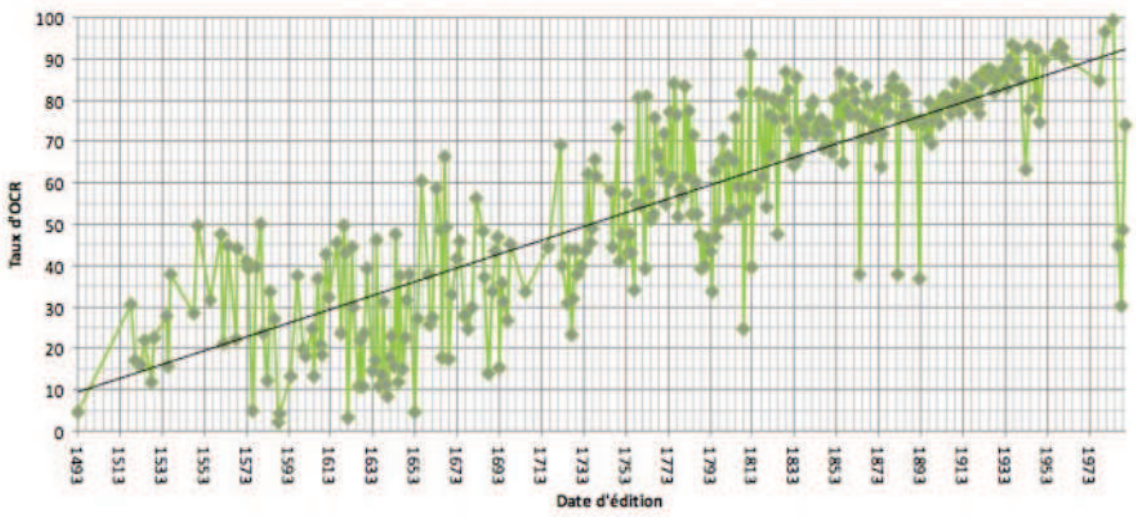
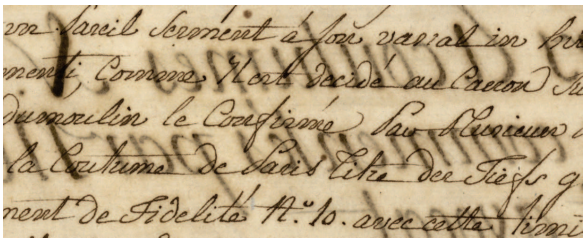


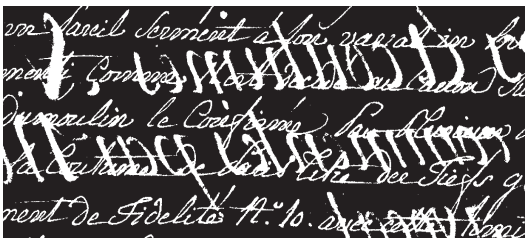
FIGURE 3.11: Taux de reconnaissance d'OCRs en fonction de la date d'édition (étude interne réalisée par la Bibliothèque Nationale de France).

Les études précédentes montrent que les dégradations sont un des facteurs responsables de la baisse de performances des OCR. Néanmoins, elles ne permettent pas de mettre en évidence un lien direct entre le niveau de dégradation et le taux d'OCR. Elles ne sont pas non plus en mesure d'évaluer l'impact successif des dégradations sur chacun des éléments de la chaîne de traitements. L'exemple dans la figure 3.12 montre un cas simple de l'influence successive des dégradations dans un processus d'OCR. La transparence dans l'image originale perturbe les résultats de la binarisation obtenus avec les méthodes d'OSTU [Otsu, 1975] et de Sauvola [Lazzara et Géraud, 2014] (figures 3.12-b et c). La transparence provoque ensuite des erreurs dans le résultat de segmentation de lignes de texte (cf. figure 3.12-d). Par conséquent, un système OCR ne peut pas reconnaitre tous les caractères dans l'image.

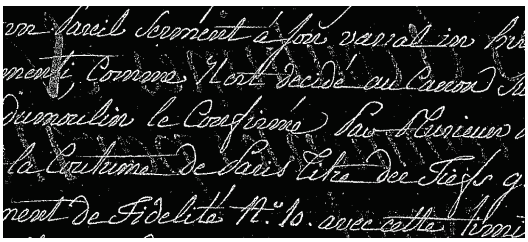
(a)Image originale



(b) Image binaire (OSTU)



(c) Image binaire (Sauvola)



(d) L'extraction de lignes de texte

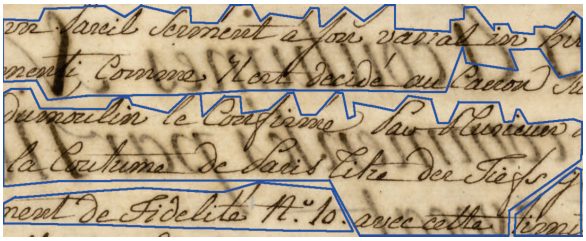


FIGURE 3.12: Exemple de l'influence successive des dégradations dans la chaîne globale.

3.1.2.2 Influence sur l'analyse de la structure de document

Dans une approche ascendante, l'analyse de la structure de documents comprend deux tâches principales : la segmentation et la catégorisation du contenu. La segmentation permet de décomposer des pixels en deux catégories : les régions/zones de texte et les régions non-texte (l'arrière-plan, les illustrations, les lettrines, les schémas, etc.). La catégorisation des différents éléments en couches d'information permet de séparer des composants de même type en plusieurs sous-types, par exemple les paragraphes, les lignes, les mots, les caractères. Ils sont considérés comme des éléments de la structure physique du document [Mao *et al.*, 2003]. L'analyse de la structure de documents se base généralement sur l'extraction de caractéristiques typographiques et typologiques du document telles que l'espace interligne, l'espace inter mots ou inter lettres. De manière générale, la variabilité du document a un impact important sur l'analyse de la structure. Ainsi, la plupart des algorithmes classiques

fonctionnent correctement avec des documents imprimés contemporains dont la structure est homogène. Par contre, la structure dans les documents anciens/historiques imprimés est plus variable. Cela peut poser des difficultés pour ces algorithmes (*e.g.* la figure 3.13 montre 4 erreurs d'extraction de la structure de documents dans les études [Kise *et al.*, 1998, Jain et Yu, 1998]).

(a) (b) (c) (d)

FIGURE 3.13: Erreurs d'extraction de la structure physique : (a) la segmentation est sensible à l'espace entre les mots qui varie entre les paragraphes [Kise *et al.*, 1998]. (b) un trait séparant les blocs de textes mène à une sur-segmentation [Kise *et al.*, 1998], (c) la lettre "I" majuscule engendre une erreur de classification [Jain et Yu, 1998], (d) la reliure a été classée comme illustration [Jain et Yu, 1998].

En plus de la variabilité du document, les dégradations , elles aussi ont une influence sur les performances des algorithmes d'analyse de structure [Mao *et al.*, 2003, Lee *et al.*, 2000, Wenyin et Dori, 1997, Diem *et al.*, 2013]. Les études menées dans [Dutta *et al.*, 2010, Fornés *et al.*, 2011, Su *et al.*, 2012a] montrent des impacts des distorsions en 2D (rotation, courbure, translation) et de bruits (bruit de [Kanungo *et al.*, 1993], taches blanches ou noires, diffusion de l'encre, etc.) sur l'extraction de lignes dans les documents musicaux. La figure 3.14 illustre quelques exemples de ce type de dégradations sur des documents musicaux. Sur ces images, un algorithme d'extraction de lignes a été appliqué. Les résultats de la table 3.2 montrent que les distorsions globales et hétérogènes produisent des taux d'erreur importants sur l'extraction de lignes, par exemple les taux d'erreur sur la Y-translation (3,64% en niveau moyen, 4,58% en niveau haut de dégradation) ou sur la translation de blocs (2,09%). Les

distorsions régulières ou les bruits locaux comme la courbure (1,43%), la rotation (1,65%) ont un impact moindre.

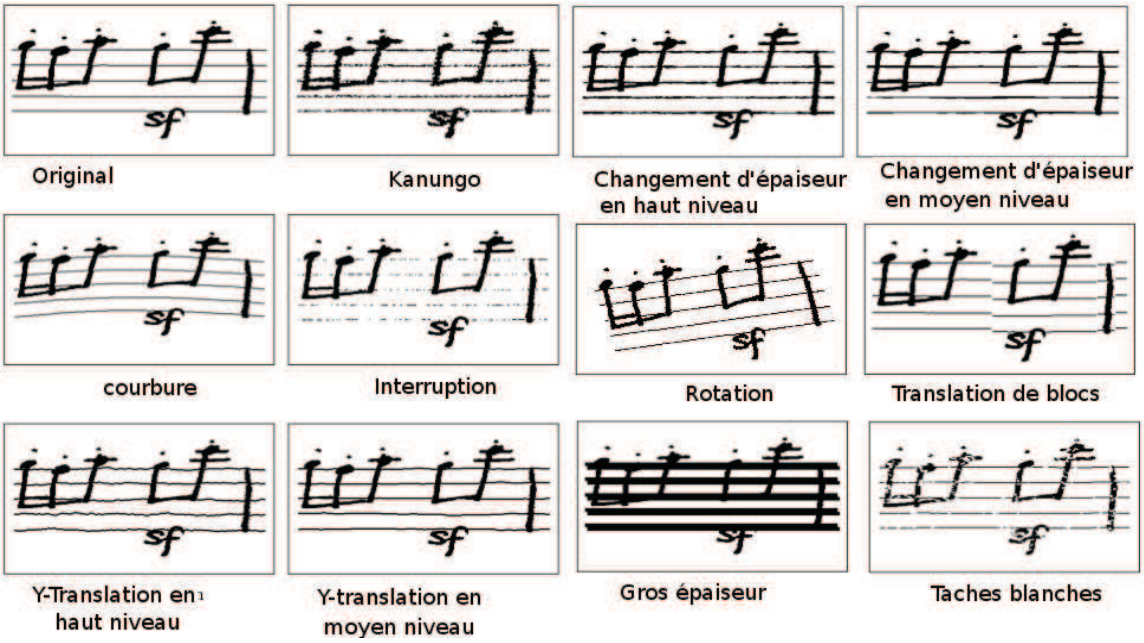


FIGURE 3.14: Défauts apparaissant dans des documents musicaux.

TABLE 3.2: Taux d'erreur suite à l'extraction de lignes dans des document musicaux [Su *et al.*, 2012a] par rapport à la présence de dégradations de la figure 3.14

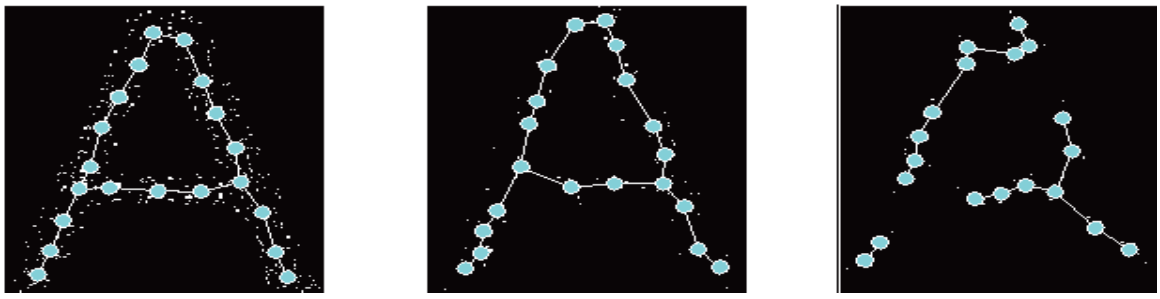
Défaut	Taux d'erreur	Défaut	Taux d'erreur
Sans distorsion	1,33%	Courbure	1,43%
Interruption	1,02%	Bruit de Kanungo	2,84%
Rotation	1,65%	ligne épaisses	3,62%
Epaisseur de ligne moyenne	2,89%	Taches blanches	1,37%
Y-Translation de niveau moyen	3,64%	Y-Translation élevée	4,58%
		Translation de blocs	2,09%

3.1.2.3 Influence sur l'extraction des caractéristiques

Certains algorithmes de reconnaissance ou d'indexation de documents nécessitent d'extraire des caractéristiques des images de documents. L'extraction de ces caractéristiques

suivent généralement un processus enchaînant les étapes de binarisation, de segmentation pour extraire les caractéristiques géométriques et statistiques (*e.g.* taille, localisation, alignement, distance relative, forme, orientation) et caractéristiques de type image (couleur, texture, etc.) [Jung, 2004]. Les dégradations peuvent générer des erreurs lors de ces différentes étapes. Par exemple, l'étude menée dans [Fujita, 2013] montrent visuellement l'influence des distorsions, du bruit local et de la présence de tâches claires lors de l'extraction de caractéristiques (cf. figure 3.15). Selon cette étude, les distorsions ne produisent qu'un impact limité sur les caractéristiques extraites. Le bruit local tel que le bruit de [Kanungo *et al.*, 1993] peut amener la génération d'un squelette de caractère coupé (cf. dernière image de la figure 3.15-b). De même, les taches claires peuvent couper le squelette ou séparer ce squelette en deux. Par exemple, la première image de la figure 3.15-c montre une tache claire à l'intérieur du caractère "a" et l'autre qui touche la bordure du caractère. La première tache sépare le squelette en deux et la seconde change la direction du squelette. L'image au milieu de la figure 3.15-c montre une tache claire qui coupe le trait du caractère "e". Cette tache coupe également le squelette du caractère. Ces dégradations peuvent introduire des erreurs importantes lors de la reconnaissance des caractères.

(a) Distorsions des caractères



(b) Bruits locaux

(c) Taches claires

FIGURE 3.15: L'impact des dégradations sur des algorithmes d'extraction de squelette des caractères : (a) les distorsions dégradent les squelettes des caractères, (b) et (c) les taches blanches coupent les caractères et leurs squelettes.

Les analyses précédentes montrent que les dégradations et la variabilité des images de documents réels influent sur la performance des algorithmes. Néanmoins, le lien entre le niveau de chaque dégradation et la performance d'un algorithme n'est pas montré clairement, car il est difficile de mesurer le niveau d'une dégradation dans une image de document réel où plusieurs dégradations sont présentes en même temps. De plus, comme énoncé précédemment, collecter un nombre conséquent d'images réelles suffisamment représentatives des défauts existants est un travail complexe et fastidieux. La création de modèles permettant de simuler des dégradations offre donc une alternative intéressante. Un modèle de dégradation permet de générer une dégradation, et de mesurer le niveau de cette dégradation en fonction du jeu de paramètres. Cela nous permet d'étudier les comportements des algorithmes

par rapport aux différentes formes de dégradations. La section suivante présente plusieurs modèles de dégradation que nous avons regroupés selon qu'ils modélisent des dégradations globales ou locales.

3.2 Modèles de dégradations existants

La proposition de modèles de dégradation nous permet d'intégrer des dégradations ciblées dans les images de documents pour en mesurer l'impact sur des algorithmes. Une dégradation générée peut l'être localement sur quelques pixels ou globalement sur tous les pixels d'une image. Par conséquent, on peut catégoriser les modèles de dégradation existants en deux groupes : les modèles de dégradation globaux et les modèles de dégradation locaux.

3.2.1 Modèles de dégradation globale

Un modèle de dégradation globale permet de générer des dégradations altérant tous les pixels d'une image. Dans la sous-section suivante, nous présentons des modèles de dégradation globaux proposés dans le domaine de l'analyse de documents.

3.2.1.1 Modèle de bruit global de Baird-1990

L'apparition de bruits dans une image de documents lors de sa numérisation ou de son impression est un phénomène couramment observé. Dès les années 90, Baird a étudié ce type de bruit pour en identifier son origine puis trouver un moyen de le simuler. Baird a proposé un modèle de dégradation basé sur le changement des paramètres physiques du matériel de numérisation [Baird, 1990]. Il a proposé un modèle basé sur onze valeurs liées à la fois au contenu de l'image mais également au contexte de numérisation : la taille du texte normalisée, le taux d'échantillonnage, le degré de rotation, le décalage en translation, le facteur d'échelle, le taux d'instabilité, le flou, le bruit libre, et le seuil de binarisation.

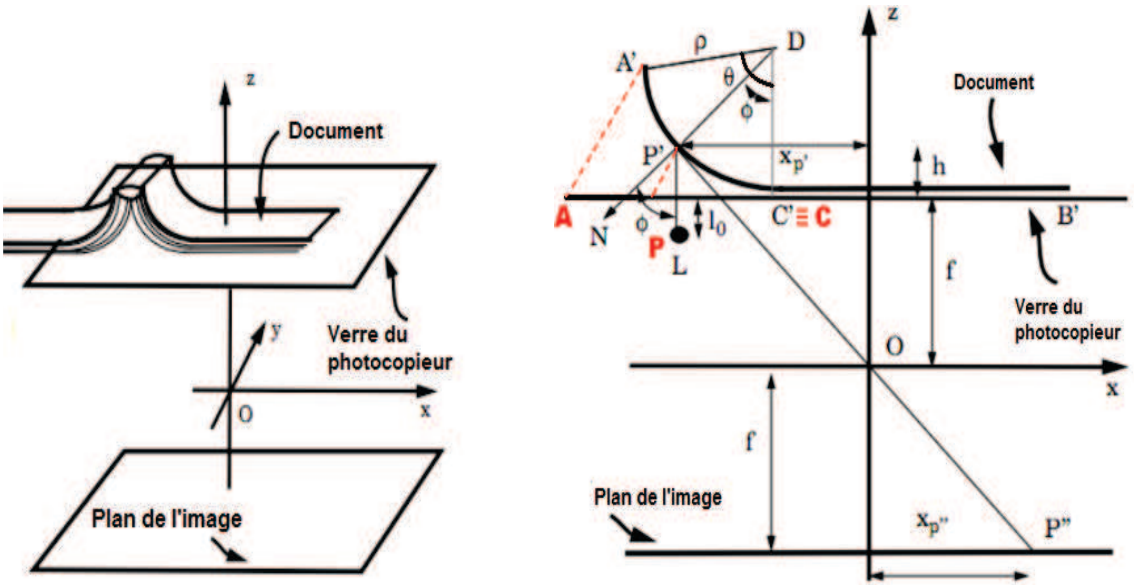
FIGURE 3.16: Exemple de différentes dégradations appliquées sur le caractère "R" en utilisant le modèle de [Baird, 1990].

Le changement de ces 11 paramètres dégrade les images de différentes façons. En général, il génère quatre types de dégradations (voir la figure 3.16). Le premier est une distorsion 2D (rotation et translation). Le deuxième est la résolution de l'image en jouant sur le taux d'échantillonnage. Le troisième est l'ajout d'un bruit déformant les contours des caractères et qui dépend du taux d'instabilité, du facteur d'échelle, du seuil de binarisasion, et du paramètre *sens*. Le flou est le quatrième type de dégradation que le modèle peut générer. Les images générées ont été utilisées pour évaluer la performance d'un système OCR.

Ce modèle a été proposé dans les années 90 pour évaluer les performances d'OCR [Ho et Baird, 1995] en fonction des dégradations générées par les scanners de l'époque. De nos jours, les scanners se sont améliorés et les défauts modélisés par [Ho et Baird, 1995] sont maintenant devenus marginaux. De plus, ce modèle s'applique sur des images binaires alors que les images de documents anciens sont souvent en niveaux de gris ou en couleur.

3.2.1.2 Modèle de déformation globale de Kanungo *et al.* 1993

Numériser à plat un ouvrage possédant une reliure épaisse produit une image déformée (courbure du papier). [Kanungo *et al.*, 1993] ont proposé un modèle reproduisant cette déformation globale.



(a) numérisation à plat d'un ouvrage (b) modélisation de la courbure et du système optique

FIGURE 3.17: Modélisation de la déformation globale de la bordure d'un document [Kanungo *et al.*, 1993].

La figure 3.17-(a) représente un document épais posé sur le verre du photocopieur. Le centre de la perspective du système optique est à O. La figure 3.17-(b) illustre comment est modélisée la déformation intervenant lorsque le bord du document est courbé. Selon le modèle géométrique, si le document est plat, tout le document devrait se situer sur le verre. Ceci est modélisé par la ligne $APCB$. Par contre, si le document est épais, APC devient l'arc $A'P'C$. Dans ce cas, on suppose qu'une partie du document est pliée selon les angles ρ et θ (l'arc $A'P'C$ est considéré comme une partie d'un cercle).

Suppose que ϕ est l'angle opposé à l'arc $P'C$. La valeur ϕ prend sa valeur entre 0 ($P' \equiv C$) et θ ($P' \equiv A'$). Grâce à ce modèle, nous pouvons calculer les nouvelles coordonnées physiques de tous les points se trouvant sur la partie pliée du document à photocopier/scanner. Tous les points dans le reste du document ne changent pas de coordonnées. A partir des coordonnées physiques calculées, il est nécessaire de trouver les nouvelles coordonnées des points projetés sur le plan d'image dans le système optique du photocopieur/scanner. La figure 3.17-b détaille le processus de projection d'un point $P' \in$ segment $[A'C]$ pour obtenir le point P''.

En supposant que la lentille se trouve au centre de la perspective du système de coordonnées et que cette lentille est séparée par un verre d'épaisseur f sur lequel est posé

l'ouvrage. A travers le système optique, le point $P' \in$ segment $[A'C]$ équivaut au point P''. Pour tous les points se trouvant dans le segment $[CB]$, $P'' \equiv P' \equiv P$. Donc, une version inverse du document se trouve sur le plan d'image. À la fin de cette étape, nous connaissons les nouvelles coordonnées sur le plan d'image de tous les points dans le document.

Finalement, le modèle prend en compte également le changement de luminosité due à la déformation de la perspective en recalculant le niveau de gris de chaque pixel en fonction des nouvelles coordonnées de P'' (voir l'article de [Kanungo *et al.*, 1993] pour avoir plus de détails).

Ce modèle comporte six paramètres : $\triangle$, f, ρ, θ, k, l_0 qui permettent de simuler principalement la distorsion globale du document épais. Ce modèle peut introduire par ailleurs l'effet de la luminosité grâce à l'équation du flou Gaussien.

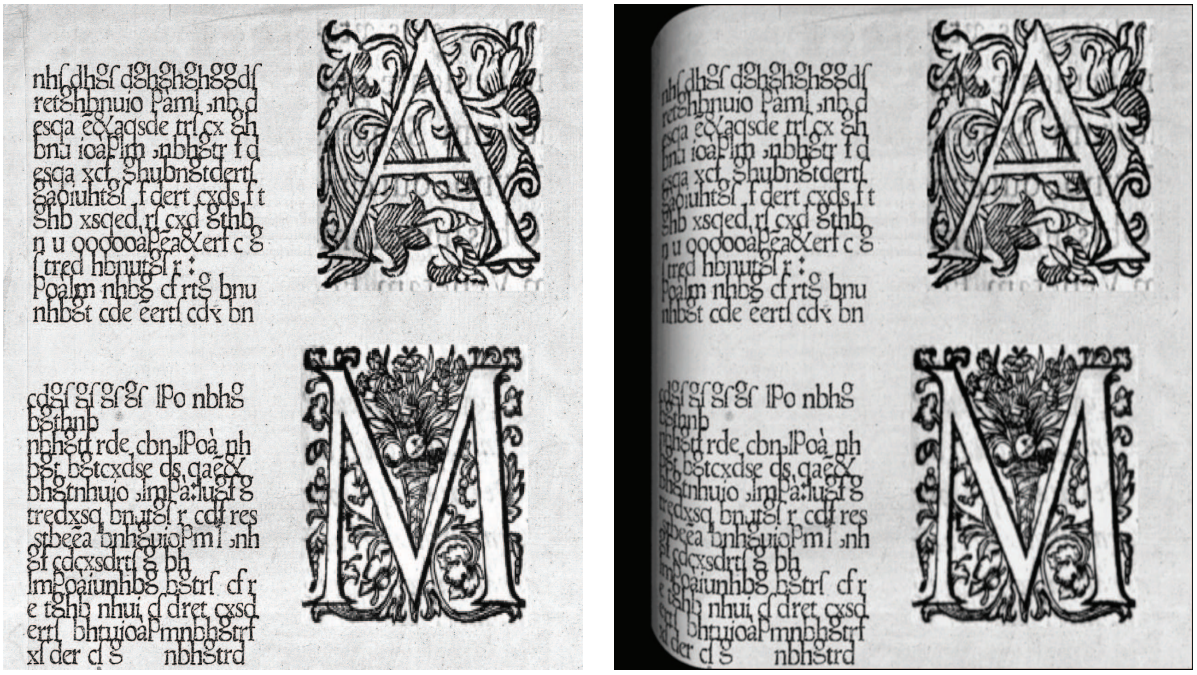


FIGURE 3.18: Exemple de résultat du modèle : (a) image originale, (b) image dégradée avec les valeurs de paramètres : $\triangle = 10$, $f = 150$, $\rho = 1200$, $\theta = 10$, $k = 10$, $l_0 = 20$.

La figure 3.18-(b) présente une image générée par ce modèle. Le pli apparaît sur la bordure du document. La zone pliée est plus foncée. La distorsion sur cette zone est homogène, car le modèle génère une courbure selon un arc du cercle.

Dans la réalité, une courbure est rarement aussi parfaitement homogène. Les documents anciens contiennent des zones dégradées visuellement plus variées. Ce modèle appliqué

sur des documents anciens produit des distorsions non réalistes. De même, la gestion de la luminosité sur la zone déformée présente le même défaut : elle est homogène.

3.2.1.3 Modèle de distorsion global de Liang *et al.* 2008

[Liang *et al.*, 2008] ont proposé un générateur d'images de documents semi-synthétique basé sur modèle de distorsion global afin d'évaluer leur méthode de restauration d'images de documents courbés. Ce modèle peut générer des distorsions (rotation, translation, courbure) ainsi que des défaut de d'éclairage. Le modèle comporte trois étapes résumées dans la figure 3.19 : la distorsion en 2D, la distorsion en 3D, et la distorsion perspective.

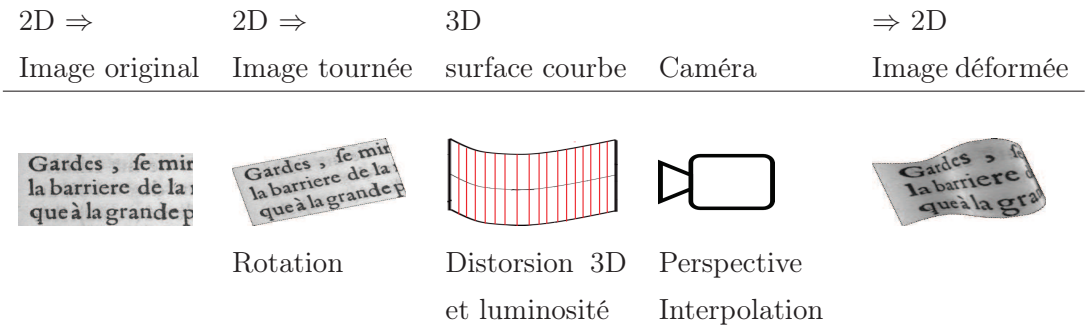


FIGURE 3.19: Modèle de distorsion globale de Liang *et al.* (2008).

L'étape de distorsion en 2D permet d'appliquer la rotation ou la translation sur l'image d'entrée.

L'étape de distorsion en 3D prend l'image transformée pour la plaquer sur une surface développable dans l'espace 3D. Cette étape est composée de deux processus : le processus de projection et celui de plaquage. Le processus de projection permet de déplier une surface développable sur le plan d'image en 2D. Une surface développable est construite en déplaçant une ligne courbe dans l'espace 3D. Par exemple, la figure 3.20-a montre une forme de type cylindrique créée en déplaçant la ligne z = F(x). Cette forme cylindrique est divisée en n facettes comme dans la figure 3.20-c. Une facette est considérée approximativement plate. La longueur de chacune est égale à $\triangle x$. Chaque facette est projetée sur le plan d'image en 2D pour obtenir une facette correspondante comme dans la figure 3.21. Puis, le processus de mapping met l'image transformée sur le plan de l'image pour calculer les valeurs des pixels dans les facettes du plan d'image. Chaque facette du plan d'image est plaquée sur sa facette correspondante dans la surface cylindrique. Ce processus est illustré dans la figure 3.21.

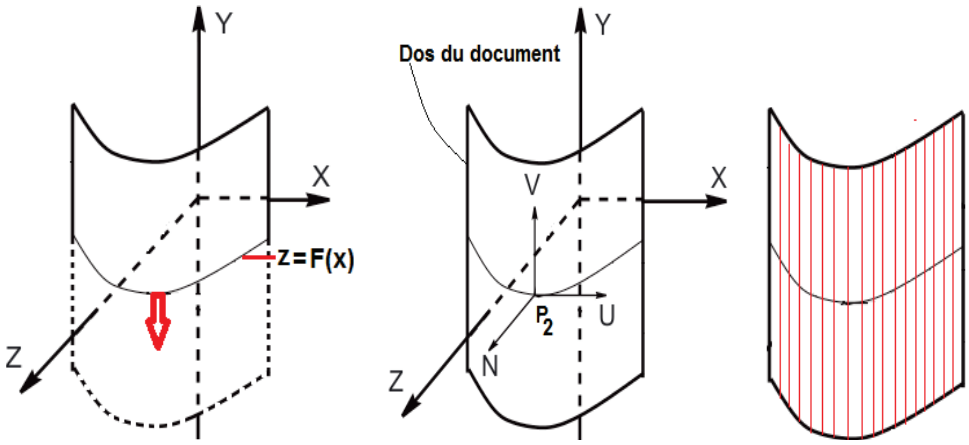


FIGURE 3.20: La forme du document est considérée comme une surface cylindrique dans l'espace 3D.

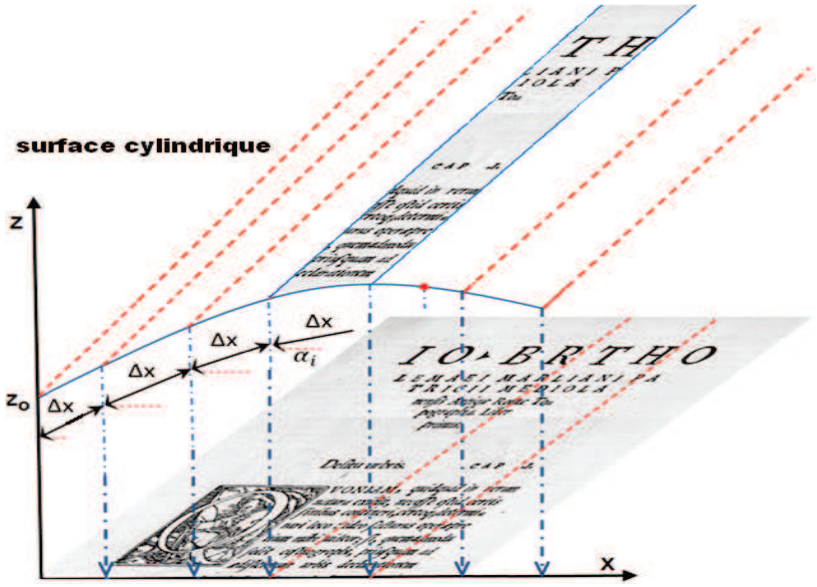


FIGURE 3.21: La surface cylindrique divisée en plusieurs facettes dans l'axes OXZ.

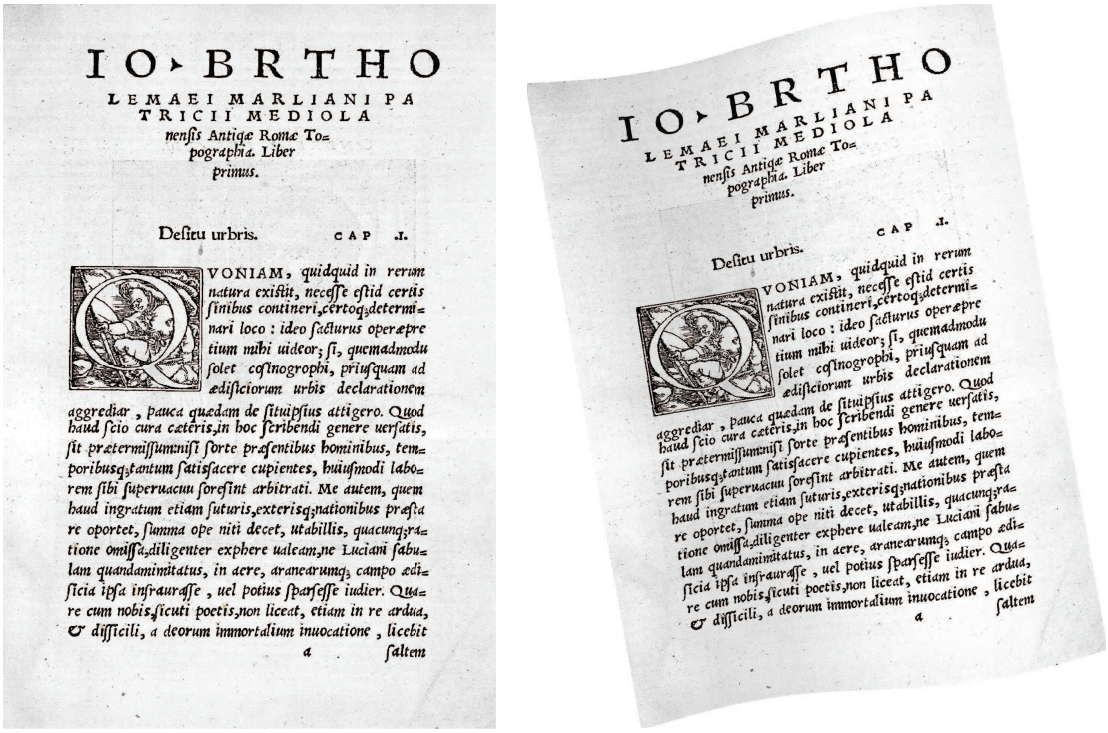
L'étape de distorsion de perspective permet de projeter l'image plaquée dans l'espace 3D sur le plan de l'image de sortie de la caméra. Cette étape se passe comme le processus de capture d'une image réelle par la caméra. Les sources lumineuses sont présentes dans l'espace de la caméra. Supposons qu'on ait K sources lumineuses, chacune est caractérisée par leur rayon R (le rayon de la zone luminaire) et un vecteur de direction D. Supposons que I

soit la valeur en couleur au point P_3, elle est calculée ainsi :

$$I = I_0 \times \sum_{i=1}^{K} (R_i \times {D^T}_i \times N) \tag{3.1}$$

Où I_0 est la fonction de distribution bidirectionnelle de la réflectance et $\overrightarrow{N}$ est le vecteur normalisé de la surface au point P_3 (cf. figure 3.20-b). La valeur $\overrightarrow{N}$ est calculée ainsi : $\overrightarrow{N} = \overrightarrow{U} \times \overrightarrow{V}$ où $\overrightarrow{U}$ est le vecteur tangent de la surface autour P_3, $\overrightarrow{V}$ est le vecteur suivi de l'axe OY.

La figure 3.22-b présente une image dégradée en utilisant ce modèle. La surface utilisée est une surface sinusoïdale. L'image originale (cf. figure 3.22-a) est tournée avec un angle de $\varphi = 15$ dégrées avant d'être plaqué sur cette surface. Les sources lumineuses sont mises au dessus du document. Par conséquent, l'effet de la luminosité est plus clair dans la courbure à droit (plus proche de la source lumineuse) et plus sombre dans celle à gauche de la figure (plus éloignée de la source lumineuse) 3.22-b.



(a) (b)

FIGURE 3.22: Image dégradée par le modèle de Liang et al. : (a) image originale, (b) image dégradée avec la rotation de $\varphi = 10^{\circ}$, la fonction de la surface $Z = sin(15 * X)$.

La figure 3.23 montre deux exemples d'images générées en changeant la fonction Z. Pour générer l'image 3.23-a, $Z = sin(5 * X)$, on peut voir 5 plis dans l'image. Avec Z $sin(0.5 * X)$, cela génère 12 plis dans l'image 3.23-b. Quand on diminue le coefficient de la fonction sin, le nombre de plis augmente.

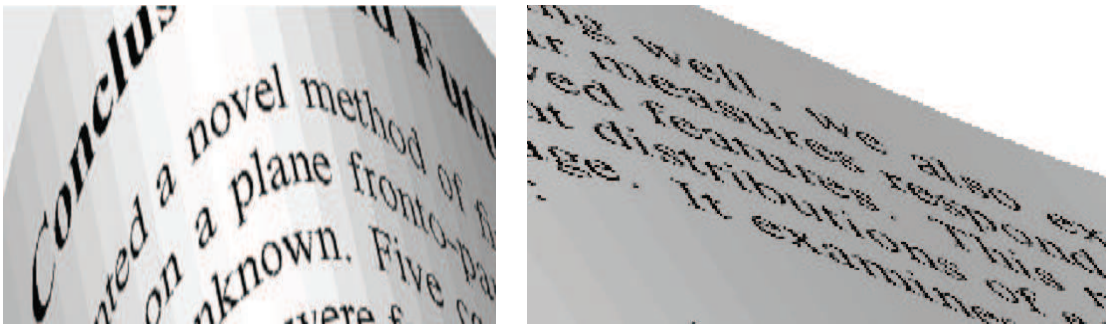
(a) (b)

FIGURE 3.23: Image dégradée par le modèle de Liang et al. : (a) image dégradée avec $\varphi = 10^{\circ}$, $Z = sin(5 * X)$; (b) image dégradée avec $\varphi = 10^{\circ}$, $Z = sin(0.5 * X)$.

Ce modèle a été utilisé pour générer une base de 120 images de documents déformés. Cette base est restaurée par la méthode de distorsions proposée dans [Liang *et al.*, 2008]. Pour évaluer la performance de la méthode, les auteurs utilisent le moteur OCR OmniPage version 12.

Ce modèle soulève deux problèmes tendant à produire des images non réalistes. Ces problèmes viennent du choix du paramètre $\triangle x$ et l'autre vient de l'étape d'interpolation. La figure 3.24 en illustre deux exemples. La figure 3.24-a montre que la surface du document n'est pas lisse quand la valeur $\triangle x$ est grande. La figure 3.24-b montre que les valeurs de pixels du premier plan sont dégradées à cause de la double application de l'interpolation. Le premier problème peut être résolu en diminuant la valeur du paramètre $\triangle x$, c'est-à-dire qu'on divise la surface de la page en de très nombreuses facettes. Cela va conduire

à augmenter le temps de calcul. Le deuxième problème peut être amélioré en choisissant une méthode d'interpolation qui s'adapte à la résolution d'image. Cependant, la méthode d'interpolation prend beaucoup de temps.



(a) (b)

FIGURE 3.24: Images exemples illustrant les limites du modèle de Liang et al. (images extraites de l'article [Liang *et al.*, 2008]).

De plus, les courbures dans les images de documents dégradées par ce modèle sont homogènes. Ce modèle ne permet pas de reproduire une grande partie des défauts rencontrés dans les images de documents anciens et qui ont tendance à être localisés dans certaines parties de l'image.

3.2.1.4 Modèle de transparence du verso sur le recto de Moghaddam 2009

La transparence ou l'apparition de l'encre du verso sur le recto (cf. la figure 3.25 à droit) est souvent présente dans les documents anciens. Ce défaut a évidemment des effets sur des algorithmes d'analyse de documents. En 2009, il a été simulé en utilisant la fonction de diffusion détaillée dans l'article [Moghaddam et Cheriet, 2009] (cf. la figure 3.25 à gauche).

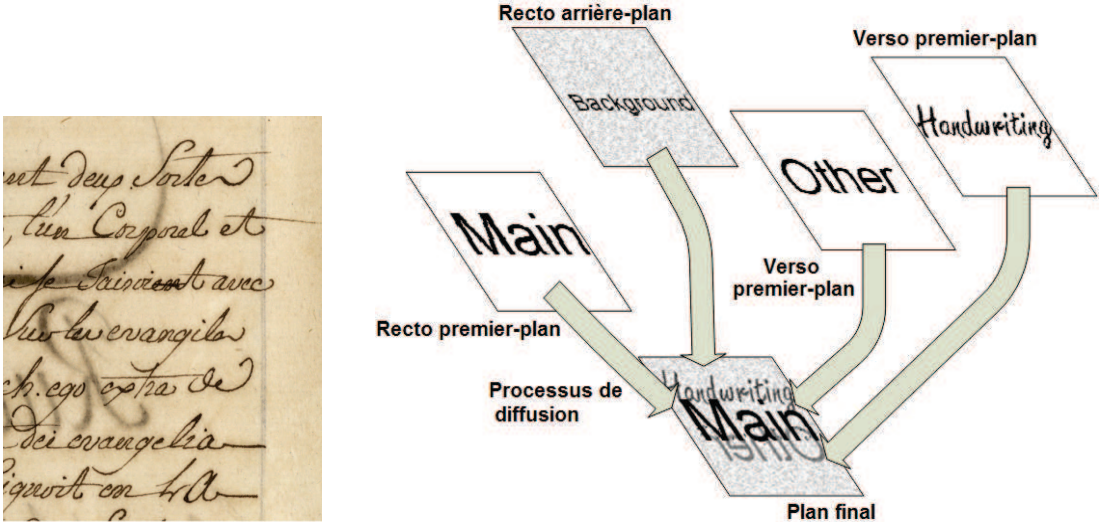


FIGURE 3.25: Un exemple d'apparition du verso sur le recto dans un document réel à droit et le processus de diffusion de l'encre de [Moghaddam et Cheriet, 2009] à gauche.

La fonction DIFF(u, s, c) où la valeur s est la source de diffusion (ex. verso), la valeur u est la cible (ex. recto), et la valeur c est le coefficient de diffusion permet de modéliser ce phénomène.

$$\frac{\partial u}{\partial t} = DIFF(u, s_{recto}, v_{recto}) + DIFF(u, s_{bg}, c_{bg}) + DIFF(u, s_{verso}, v_{verso}) \tag{3.2}$$

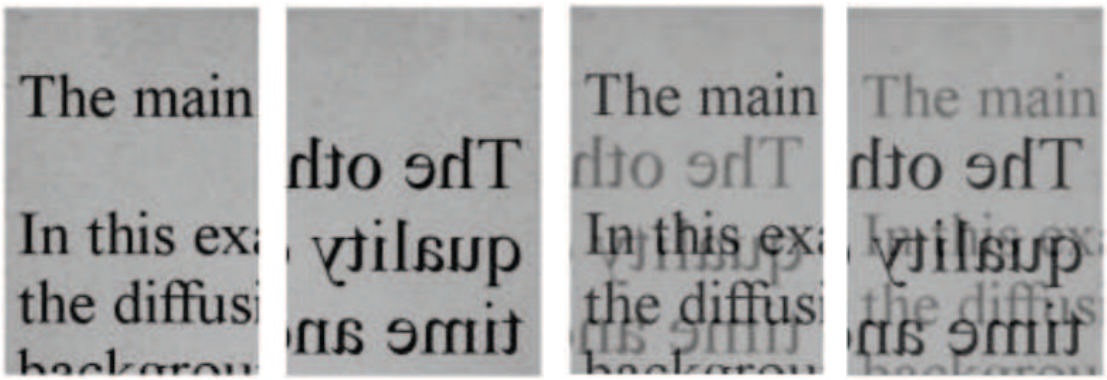
Dans l'équation (3.2), $u(x,\ y,\ t)$ est la valeur de niveau de gris du pixel $(x,\ y)$ à l'instant t. Les entrées s_{recto}, s_{verso} correspondent à l'image du recto et celle du verso. L'entrée s_{bg} est un document réel sans texte/dessin. Elle est considérée comme l'arrière-plan de l'image de sortie. Les trois coefficients de diffusion sont positifs et plus petits que 1. Ils sont calculés ainsi :

$$\begin{cases} c_{recto} = \frac{1}{1+(u_t/\sigma_{recto})^2} \\ c_{bg} = d_{bg} \times (1 + \tanh(u_t - s_{bg} - \delta_{bg})/\sigma_{bg} \\ c_{verso} = \frac{d_{verso}}{1+(s_{verso}-u_t)^2/(\sigma_b)^2} \times \frac{1}{1+s^2_{verso}/\sigma^2_{ink}} \end{cases} \tag{3.3}$$

Puisque la différence entre l'image de sortie u_t et l'arrière-plan s_{bg} est grande, les trois paramètres d_{bg}, δ_{bg}, et σ_{bg} sont ajoutés pour assurer que le texte ne change pas. Ils sont plus petits que 1. Dans les expérimentations, les auteurs ont choisi $\delta_{bg} = 0{,}2$ et $\sigma_{bg} = 0{,}3$. Le paramètre d_{verso} est le ratio entre la diffusion du verso et la diffusion du recto. Le

paramètre σ_{ink} est un paramètre général qui contrôle la diffusion de l'encre. Il permet de fixer la diffusion de l'encre du verso vers le recto (ou inversement).

La figure 3.26 montre deux exemples (c) et (d) générés avec ce modèle de dégradation de l'apparition de l'encre du verso (a) sur le recto (b). Ces deux résultats sont différents en fonction du ratio d_{bg}/d_{verso}. La transparence générée par ce modèle est visuellement très réaliste et reproduit bien toutes les nuances de niveaux de gris observables dans les images réelles possédant de la trasparence.



(a) recto d'origine (b) verso d'origine (c) recto synthétique (d) verso synthétique

FIGURE 3.26: Exemples de l'apparition de l'encre du verso sur le recto : (a) et (b) sont le recto et le verso, (c) et (d) sont deux images générées par ce modèle [Moghaddam et Cheriet, 2009].

La plupart de modèles précédents sont dédiés aux contextes des images binaires [Baird, 1990, Kanungo *et al.*, 1993, Zhai *et al.*, 2003, Smith, 2008]. Selon nous, seul le modèle de Moghaddam [Moghaddam et Cheriet, 2009] génère des résultats visuellement similaires à ceux observés dans les documents anciens.

3.2.2 Modèles de dégradation locale

Un modèle de dégradation locale permet de générer des dégradations qui sont effectuées sur quelques pixels d'une image. Par conséquent, un tel type de dégradation impacte uniquement quelques éléments du document : des caractères, certaines parties du fond. Dans la sous-section suivante, nous listons des modèles existants qui peuvent être appliqués sur des images de documents.

3.2.2.1 Modèle de bruit de Kanungo *et al.* 1993

Le modèle de dégradation locale proposé par [Kanungo *et al.*, 1993] permet de générer le bruit au niveau du pixel dans une image binaire. En général, ce bruit peut modifier la courbure des caractères. De fait, ce modèle permet d'inverser aléatoirement des valeurs de pixels (un pixel du premier-plan devient l'arrière-plan et vice-versa). Cette inversion se produit indépendamment sur chaque pixel selon une fonction de probabilité dépendant de la distance de transformation de ce pixel par rapport au bord d'un caractère.

La distance de transformation est la distance minimale (d) d'un pixel par rapport au bord du caractère. Cette distance est calculée en 4-connectivité ou 8-connectivité (voir la figure 3.27). La probabilité pour qu'un pixel de l'image soit inversé du premier-plan (noir) à l'arrière-plan (blanc) et réciproquement est calculée selon la formule suivante :

$$\begin{cases} p(0|1, d, \alpha_0, \alpha) = \alpha_0 \times e^{-\alpha d^2} & \text{Si le pixel est noir} \\ p(1|0, d, \beta_0, \beta) = \beta_0 \times e^{-\beta d^2} & \text{Si le pixel est blanc} \end{cases} \tag{3.4}$$

Avec $p(0|1,\ d,\ \alpha_0,\ \alpha)$ la probabilité pour inverser un pixel d'encre (1) en un pixel de fond (0). Et $p(1|0,\ d,\ \beta_0,\ \beta)$ est la probabilité pour inverser un pixel de fond (0) en un pixel d'encre (1). Les paramètres du modèle sont α_0, α, β_0, β.

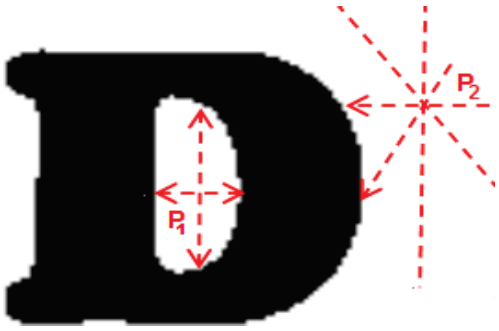


FIGURE 3.27: La distance minimale pour un pixel en 4-connectivité (P1) et en 8-connectivité (p2).

Un pixel sera inversé si sa probabilité est supérieure au seuil r. Le seuil r est fixé aléatoirement et indépendamment pour chaque pixel selon une distribution uniforme. Finalement, une fermeture morphologique est appliquée pour connecter le contour du caractère.

Les paramètres α et β permettent de contrôler le nombre de pixels inversés. Le nombre de pixels inversés est augmenté selon la distribution exponentielle lorsque les deux paramètres diminuent (cf. Figure 3.28).

Originale $\alpha = \beta = 86$ $\alpha = \beta = 50$ $\alpha = \beta = 18$ $\alpha = \beta = 3$ $\alpha = \beta = 1,2$

FIGURE 3.28: Images dégradées par le modèle de bruit local de Kanungo où $\alpha_0 = \beta_0 = 1$.

La figure 3.28 montre les images dégradées du caractère "D" en diminuant les valeurs de deux paramètres α et β. Ce bruit peut couper la connectivité du caractère ou modifier le contour du caractère. Il influence effectivement les méthodes d'analyse et de reconnaissance de caractères. Le gros défaut de ce modèle est qu'il est applicable uniquement aux images binaires.

3.2.2.2 Modèle de diffusion de l'encre de Curtis *et al.* 1997

Les auteurs de l'article [Curtis *et al.*, 1997] ont proposé un modèle de diffusion de l'encre reproduisant la diffusion à travers le papier d'une goutte d'eau. Cette diffusion peut produire plusieurs types de défauts dans un document, par exemple des grandes taches dans l'arrière-plan, des caractères abimés ou se connectant artificiellement deux caractères proches (cf. exemples 3.3.a-e).

L'encre se compose de molécules qui adhèrent à la page après l'action d'écriture. Physiquement, quand une goutte de liquide tombe sur la page, ces molécules se déplacent aléatoirement dans la zone de la goutte. Après le liquide s'évapore ou pénètre la page, les molécules se diffusent dans la zone et produisent des défauts. Ce processus physique est modélisé en trois étapes résumées dans la figure 3.29, et détaillées dans les paragraphes suivants.

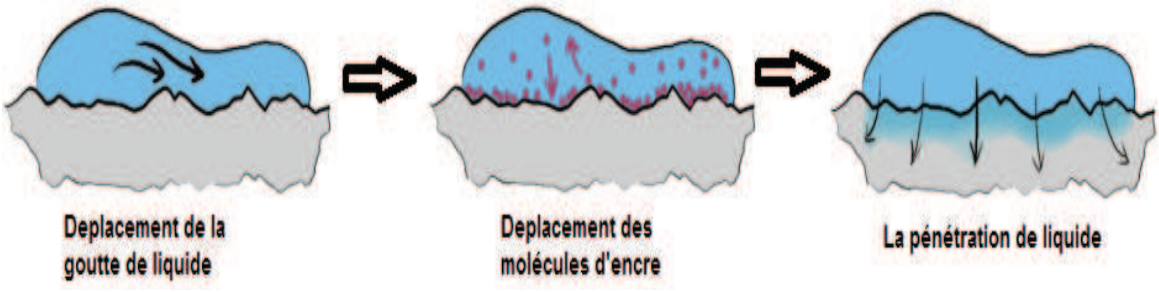


FIGURE 3.29: Les trois étapes du modèle de diffusion de Curtis et al. 1997 (images extraites de [Curtis *et al.*, 1997]).

Les paramètres initiaux de la première étape sont les suivants : l'état de la page (M : 1-humide, 0-non humide), la vitesse de déplacement de la goutte (v_x, v_y), la pression de la goutte (p), le niveau de rugosité de la page ($\triangle h$), la concentration g^k de la molécule k, la viscosité μ et la traînée visqueuse K qui sont fixées μ= 0,1 et K = 0,01. La suite du processus est modélisée selon des propriétés physiques basées sur la vitesse, la densité, la convergence de coloration et la granularité. Ces propriétés modélisent le phénomène lié à l'absorption du liquide par la page. Le phénomène de pénétration du liquide dépend de deux paramètres principaux : la saturation en eau et la capacité de rétention de l'eau. Le liquide se diffuse à cause de la saturation. Les molécules liquides se fixent après l'évaporation/pénétration. La figure 3.30 montre deux taches couleur générées par ce modèle avec deux densités différentes.

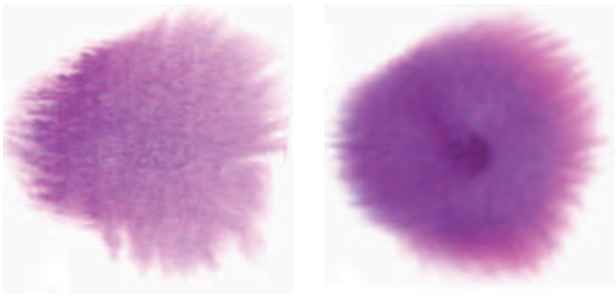


FIGURE 3.30: Deux d'exemple de génération de tache par diffusion en utilisant deux paramètres de densité différents.

Ce modèle a été utilisé pour de la génération d'images en couleur et l'animation d'images en couleur (*e.g.* il est intégré dans le logiciel Photoshop). Néanmoins, son application sur des images de documents reste limitée car il nécessite une forte paramétrisation et une intervention manuelle conséquente de la part de l'utilisateur. Une adaptation de ce modèle dédiée aux documents permettrait de générer à grande échelle et facilement des dégradations de type diffusion d'encre.

3.2.2.3 Modèle de bruit "hard pencil noise" de Jian zhai *et al.* 2003

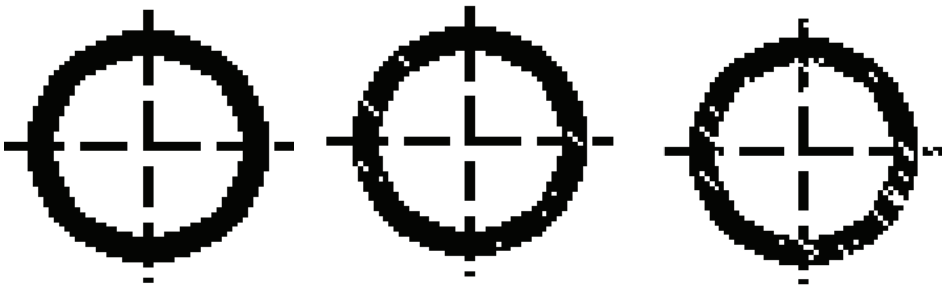
Les auteurs de l'article [Zhai *et al.*, 2003] ont proposé un modèle de bruit appelé "hard pencil noise" qui est lié au problème de l'encre lors de l'utilisation d'un stylo. Ce stylo est posé sur la page pour dessiner un trait, mais quelques fois l'encre du stylo ne sort plus. Cela produit de petits espaces blancs le long du tracé.

Les auteurs définissent un seuil L entre 0 à L_{MAX} où L_{MAX} est la longueur maximale du contour d'une composante connexe. Au pixel $P(i, j)$, une valeur R est aléatoirement

fixé avec la formule :

$$R = 13 \times \frac{rand() \times L_{MAX}}{RAND_{MAX}} \quad \text{Où } RAND_{MAX} \text{ est une constante} \tag{3.5}$$

Pour ce pixel, une ligne de bruit est générée si et si seulement le chiffre $R < L$. La longueur de la ligne est choisie entre 0 et $\frac{L+5}{3}$. La valeur du seuil est donc directement liée au niveau de dégradation générée.



Originale (L=0) Image dégradée (L=5) Image dégradée (L=10)

FIGURE 3.31: Images dégradées par le modèle de Jian Zhai *et al.* (Les sources d'images viennent de l'article [Zhai *et al.*, 2003]).

La figure 3.31 montre deux images dégradées par ce modèle avec différents niveaux de dégradation L qui contrôle le nombre de lignes de bruit. Ce bruit permet de modifier le contour ou de couper la connectivité des composantes connexes. Comme le bruit Kanungo [Kanungo *et al.*, 1993], ce bruit est binaire, il est donc difficile à appliquer aux images de documents anciens en niveaux de gris.

3.2.2.4 Modèle de bruit local d'Elisa Smith

Dans l'article de [Smith, 2008], l'auteur a présenté un modèle de bruit pour générer des défauts reproduisant ceux générés par un scanner. Le processus est modélisé de la façon suivante :

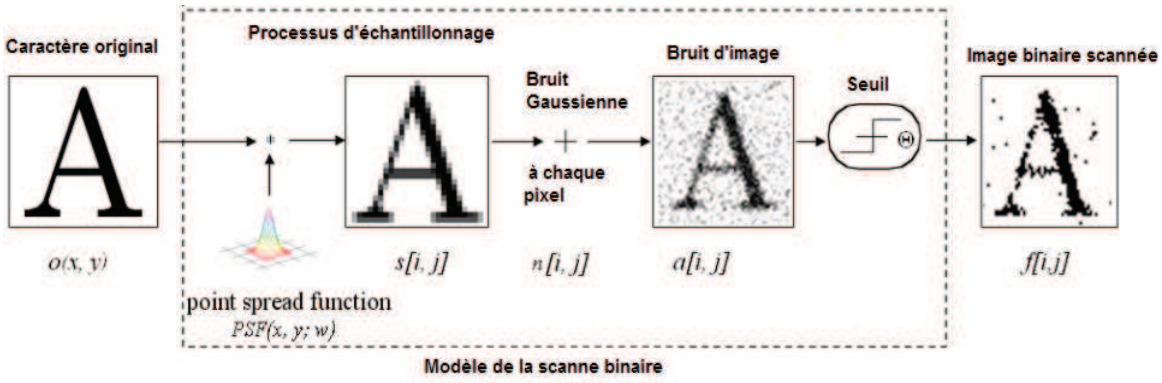


FIGURE 3.32: Schéma du modèle de bruit scanné d'Elisa B. Smith.

Le scanner dispose des capteurs linéaires qui permettent de capturer l'illumination. La température des capteurs augmente pendant la numérisation et perturbe la valeur lumineuse obtenue. Par conséquent, un capteur peut, pour deux pixels normalement identiques, produire deux valeurs différentes alors que les réglages du scanner sont les mêmes. Après binarisation du document, ce phénomène a tendance à générer un bruit de type "poivre et sel". L'impact de la luminosité $s[i, j]$ lors de l'acquisition de chaque pixel est modélisé par la fonction de diffusion $PSF(x, y)$ (Point Spread Function). Un bruit Gaussien indépendant $n[i, j]$ est ajouté par la suite. La valeur d'intensité finale d'un pixel est donc calculée selon la formule suivante :

$$a[i, j] = s[i, j] + n[i, j] \tag{3.6}$$

Un seuil global θ est défini pour fixer la valeur binaire d'un pixel (i, j). Ce pixel devient noir (1) si sa valeur $a(i,j)$ est supérieure à θ ; sinon il est blanc (0).

L'article [Barney Smith, 2000,Hale et Barney Smith, 2007] détaille comment le seuil de binarisation et les valeurs de la fonction PSF (Point Spread Function) du modèle sont estimés. Une validation visuelle est réalisée sur la base d'une comparaison entre caractères scannés *via* deux scanners différents et les caractères synthétiquement déformés. Néanmoins, cette validation ne permet pas de mesurer un lien entre le niveau de dégradation et les performances OCR. Autre problème, comme le bruit Kanungo [Kanungo *et al.*, 1993], ce bruit ne s'applique que sur les images binarisées.

3.3 Conclusion

Dans ce chapitre nous avons examiné les dégradations les plus couramment observées dans les documents et plus particulièrement dans les documents anciens. Nous proposons de catégoriser ces dégradations selon qu'elles soient intrinsèques au document ou dues à la numérisation. Nous avons montré que la présence de ces dégradations influence les performances d'algorithmes d'analyse de documents. Dès lors, nous avons présenté plusieurs modèles de dégradation visant à les reproduire de manière synthétique. Ces modèles permettent d'intégrer dans une image originale une dégradation (ou une combinaison de dégradations) dans la base de test pour évaluer les performances d'algorithmes ou générer des données d'apprentissage.

La plupart de modèles de dégradation s'appliquent sur les images binaires de documents. Ils sont rarement adaptés aux documents anciens qui sont en niveaux de gris. De plus, une majorité des ces modèles de dégradation ne permettent pas de générer des distorsions globales hétérogènes. Les approches locales ne permettent pas non plus de reproduire des déformations locales telles que les plis, les trous, etc.

Le chapitre suivant détaille deux de nos contributions liées à la modélisation de défauts. Le premier permet de produire, en niveaux de gris, les défauts apparaissant dans les caractères ou à leur proximité. Le second modèle permet de reproduire très finement les déformations réelles du papier.

Chapitre 4

Proposition de modèles de dégradation d'images de documents

La plupart des modèles de dégradation existants présentés dans le chapitre précédent ont été pensés pour reproduire des défauts observés dans les documents contemporains. Ils s'appliquent généralement sur des images binaires et reproduisent des défauts inhérents aux scanners utilisés dans les années 80. Il est donc nécessaire de proposer de nouveaux modèles permettant de synthétiser des modèles de dégradation adaptés aux spécificités des documents anciens. Nous proposons ainsi deux nouveaux modèles de dégradation qui permettent de reproduire les distorsions du papier dans la section 4.1 et la dégradation de l'encre dans la section 4.2.

4.1 Contribution 1 : Proposition d'un modèle de distorsion 3D du papier

Comme vu dans le chapitre précédent, le papier sur lequel est imprimé un ouvrage peut présenter plusieurs types de défauts. Nous les classons en deux groupes : les distorsions globales (longe courbure, rotation) et les distorsions locales (pli, trou, petite courbure, etc.). Les distorsions globales peuvent être simulées par des modèles de distorsion en 2D qui utilisent des transformations mathématiques comme la rotation, la translation, ou la déformation de la perspective [Kanungo *et al.*, 1993, Liang *et al.*, 2008, Fornés *et al.*, 2011]. En plus de générer des défauts visuellement très synthétiques, ces modèles ne permettent pas de générer des distorsions locales. C'est la raison pour laquelle nous proposons un modèle de distorsion en 3D qui permet non seulement de générer des distorsions globales, mais également de simuler des distorsions locales. Ce modèle se compose de trois étapes successives détaillées dans les sous-sections suivantes.

4.1.1 Présentation du modèle de distorsion en 3D

L'idée générale est d'acquérir la forme géométrique naturelle de documents anciens, et puis d'y plaquer une image 2D pour générer des images semi-synthétiques. Par conséquent, des distorsions très précises peuvent être reproduites. Ce modèle de distorsion en 3D comprend trois étapes principales : (A) génération de maillages, (B) calcul de coordonnées de texture du maillage, et (C) plaquage d'une image sur le maillage. Les trois étapes sont présentées dans la figure 4.1.

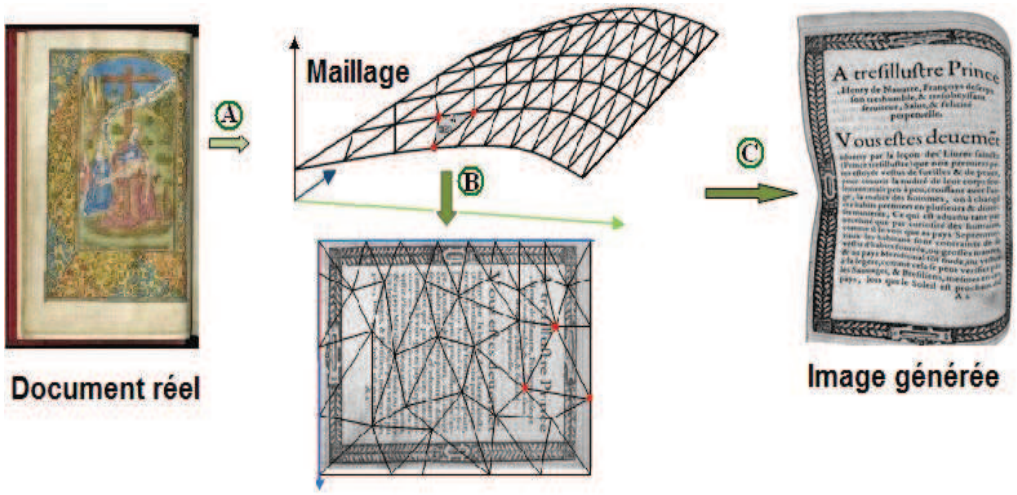


FIGURE 4.1: Schéma du modèle de distorsion en 3D : (A) numérisation 3D d'un document ancien pour générer un maillage, (B) calcul de coordonnées de texture du maillage, et (C) plaquage d'une image sur le maillage pour générer une image de document dégradée.

4.1.1.1 Étape A : Génération de maillages

Cette étape permet d'acquérir la surface de documents sous la forme de maillages. La numérisation *via* scanner 3D permet de capturer la forme d'un document avec une précision remarquable. Nous utilisons un Scanner 3D Kréon, modèle Aquion doté d'une précision de 60 μm. Chaque page est scannée par le haut en posant le verso de l'ouvrage sur une vitre (cf. figure 4.2 à gauche). Afin de garder les distorsions naturelles du papier, aucune pression n'est exercée sur l'ouvrage. A l'aide d'un opérateur, un laser est projeté sur la surface de la page et génère un grand nombre de points dont les coordonnées (3D) sont celles de la surface des deux pages. Tous ces points acquis permettent de reconstruire la forme de la page. Le résultat est un maillage qui contient des triangles (cf. figure 4.2) dans l'espace 3D. Ce processus nécessite environ 15 minutes de travail par page.

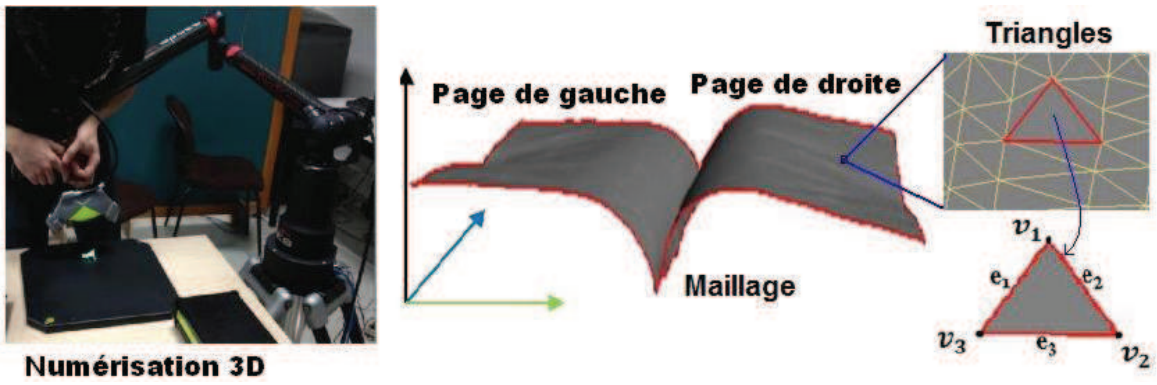


FIGURE 4.2: Processus de numérisation 3D d'un document.

Après l'étude menée dans le chapitre précédent, nous avons sélectionné 12 ouvrages présentant des déformations du papier représentatives et diverses [1]. Ces documents contiennent des pages présentant des distorsions telles que des plis de différentes tailles et formes, de petites déformations convexes ou concaves, des ondulations en forme de vagues, etc. Les tailles et épaisseurs de documents choisies sont diverses. Ces documents sont classés en deux types : l'un contient des distorsions globales, l'autre contient des distorsions locales. Nous avons finalement généré 12 maillages : 4 maillages contiennent des distorsions globales, 8 autres contiennent des distorsions locales. Les 12 maillages sont utilisés en entrée de l'étape suivante de calcul de coordonnées de texture.

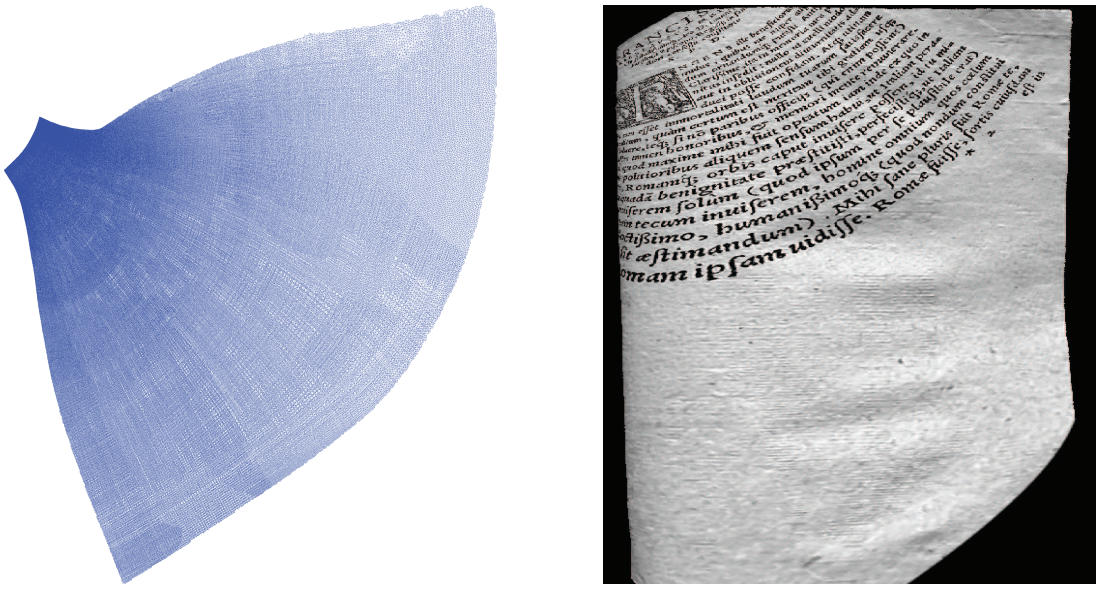
Ces 12 maillages judicieusement choisis nous ont permis de mener à bien de nombreux tests (*cf.* chapitre suivant). De plus, les maillages générés peuvent toujours être déformés pour produire de nouveaux maillages (*e.g* on peut déplacer des points dans un maillage pour en construire un nouveau).

4.1.1.2 Étape B : Calcul de coordonnées de texture du maillage

La forme d'un objet complexe peut être dépliée sur le plan 2D pour plaquer n'importe quelle image sur la surface de cet objet. L'étape de dépliage est bien connue en modélisation 3D. Elle revient à extraire de coordonnées de texture. La plupart des méthodes proposées sont dédiées au plaquage de texture sur des objets complexes dont la surface contient des occlusions et beaucoup de variations [Lienhardt, 1988, Shinagawa *et al.*, 1991, Lévy *et al.*, 2002, Pal *et al.*, 2014]. Ces méthodes nécessitent une étape de segmentation manuelle qui permet de diviser la surface de l'objet en plusieurs sous-parties simples à déplier sur un plan

1. Nous remercions la société Arkhenum pour l'aide apportée à cette étape de sélection d'ouvrages issus de leurs collections

2D. Il existe, par exemple, des logiciels professionnels comme 3Ds max qui permettent de déplier manuellement un maillage.



(a) Les coordonnées de texture sur le plan 2D

(b) une image 2D plaquée sur le maillage en utilisant les coordonnées de texture (a)

FIGURE 4.3: Résultat d'un dépliage effectué avec la méthode présentée dans [Lévy *et al.*, 2002]

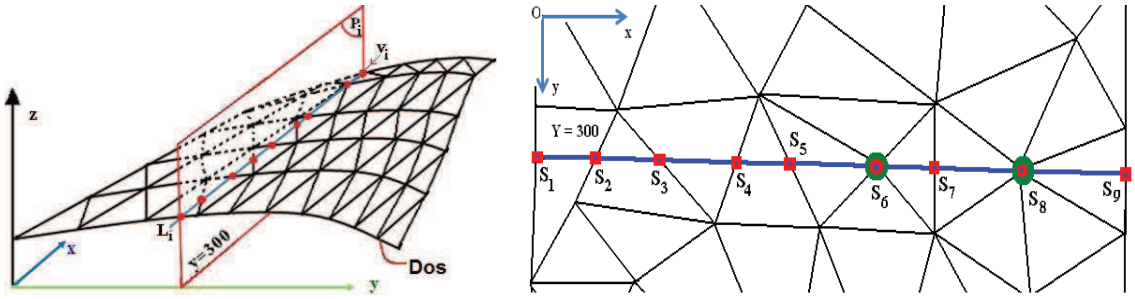
Dans notre cas, la surface du document (le maillage) ne se trouve pas être aussi complexe que les surfaces sur lesquelles travaille la communauté de modélisation 3D. Par conséquent, nous pensons que la qualité du dépliage de la surface peut être réalisée sans étape de segmentation manuelle du maillage. Nous souhaitons donc pouvoir réaliser un processus de dépliage automatique. Pour cela, nous avons adapté le processus de dépliage présenté dans [Lévy *et al.*, 2002] (code source publié sur le site [2]). Cette adaptation nous permettra, *in fine* de plaquer automatiquement une image 2D sur le maillage acquis en 3D. La figure 4.3 montre un résultat obtenu sur l'un de nos maillages en utilisant un méthode de plaquage basé sur une segmentation automatique. Les coordonnées de texture calculées *via* cette méthode génèrent une distortion 4.3-a. Par conséquent, l'image finale est mal plaquée sur le maillage d'entré (voir la figure 4.3-b). C'est la raison pour laquelle nous proposons notre propre méthode d'extraction des coordonnées de texture pour nos maillages de documents anciens.

2. http ://alice.loria.fr/index.php/publications.html ?Paper=lscm@2002

Nous proposons donc ici une approche originale qui, au lieu de chercher à déformer une image 2D pour imiter les déformations réelles, va chercher à plaquer une image 2D sur une surface 3D acquise sur un document réel.

L'idée de cette étape est de pouvoir reproduire algorithmiquement ce que l'on ferait à la main en dépliant/étirant un document légèrement déformé (en 3D) pour l'appliquer sur une surface plane (2D). Toute la difficulté est de trouver comment faire correspondre un point dans le document (2D) avec un point du document déformé (3D).

La figure 4.4 illustre la première étape de ce calcul de coordonnées. Soit le maillage de la figure 4.4-a. Supposons qu'un plan P (en rouge sur la figure) soit perpendiculaire au plan Oxy et que l'intersection de ce plan P avec le maillage passe au moins par un sommet des triangles composant ce maillage. Dans l'exemple figure 4.4-b le plan P coupe le maillage par une ligne droite (dans l'espace 3D) L = S_1, S_2, S_3, ..., S_9 où S_i est une intersection entre P et un triangle du maillage. Dans cet exemple, S_6 et S_8 correspondent à deux sommets du maillage.



(a) (b)

FIGURE 4.4: Le processus de calcul des coordonnées de texture.

L'idée de l'étape suivante est illustrée sur la figure 4.5), elle consiste à projeter cette ligne (en 3D) sur le plan 2D que représente une image de document. Cela permet de calculer les coordonnées de texture des sommets du maillage d'origine. Pour réaliser cette projection, l'algorithme proposé s'appuie sur les propriétés visuelles d'une image de document qui se trouve être une image rectangulaire avec principalement des composantes de texte alignées les unes par rapport aux autres. Notre algorithme reporte la longueur entre deux points successifs (3D) de la ligne $[L_iV_i]$ sur une ligne (2D) de l'image de document. Les caractéristiques géométriques sont ainsi conservées et cela minimise la déformation du plaquage de texture. Les points d'intersection sont donc projetés sur le plan tel que la distance entre ces points soit maintenue. Supposons que L' soit la ligne de projection sur

le plan 2D. Cette ligne contient des points projetés = S'_1, S'_2, S'_3, ..., S'_9 dans lesquels , S'_6 et S'_8 sont deux points correspondant aux deux sommets S_6 et S_8 avec comme contraintes $S'_1S'_2 = S_1S_2$, $S'_2S'_3 = S_2S_3$, $S'_3S'_4 = S_3S_4$ ...etc. Par conséquent, les coordonnées de S'_6 et S'_8 sont les coordonnées de texture des deux sommets S_6 et S_8.

Cette étape de mise en correspondance entre sommets du maillage 3D et de l'image 2D est réitérée pour chaque plan P_i. Le nombre de plans P_i dépend du nombre de sommets du maillage. Le plan P_i glisse successivement entre le premier et le dernier sommet pour calculer toutes les coordonnées de texture. Il y a donc au maximum autant de plan P_i que de sommets. Si, comme dans la figure 4.5, une ligne passe par deux sommets, cela réduit le nombre de plans P_i à analyser.

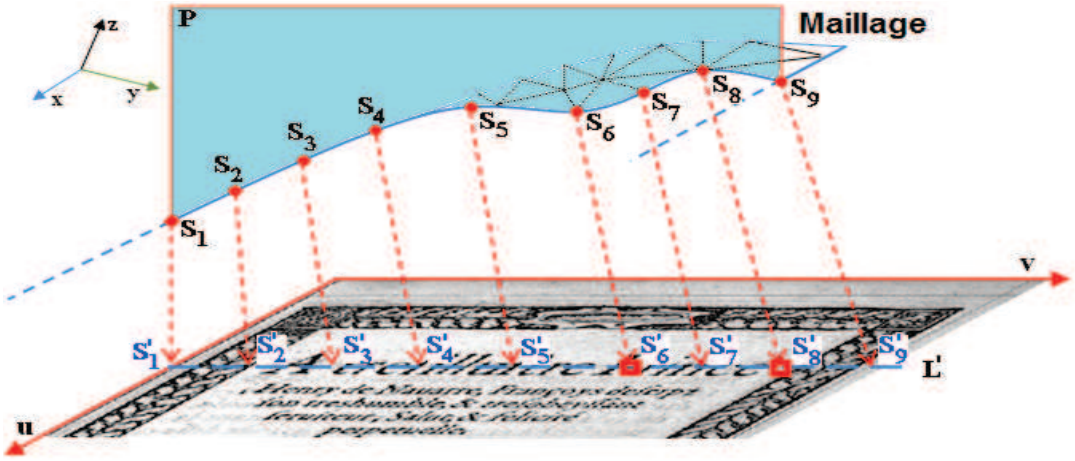


FIGURE 4.5: La projection d'une ligne d'intersection en 3D sur le plan 2D d'une image de document.

La figure 4.6 montre un résultat obtenu à l'aide de notre méthode. Les coordonnées de texture extraites ne sont pas dégradées. La forme d'une page est dépliée sur le plan 2D comme une page plate (voir la figure 4.6-a). L'extraction de ces cordonnées permet de plaquer correctement une image 2D sur un maillage (voir la figure 4.6-b).

(a) Les coordonnées de texture sur le plan 2D

(b) une image 2D plaquée sur le maillage en utilisant les coordonnées de texture (a)

FIGURE 4.6: Résultat du dépliage de notre méthode sans besoin d'une segmentation manuelle du maillage d'origine

4.1.1.3 Étape C : Plaquage d'une image 2D sur le maillage

Cette étape consiste à prendre en entrée les coordonnées de texture calculées à l'étape précédente (correspondance entre sommets 3D et pixels de l'image). Pour plaquer la texture sur le volume, nous utilisons la bibliothèque OpenGL. Elle nous permet de réaliser ce plaquage en réalisant une interpolation bi-linéaire pour corriger les pixels blancs dans l'image résultat.

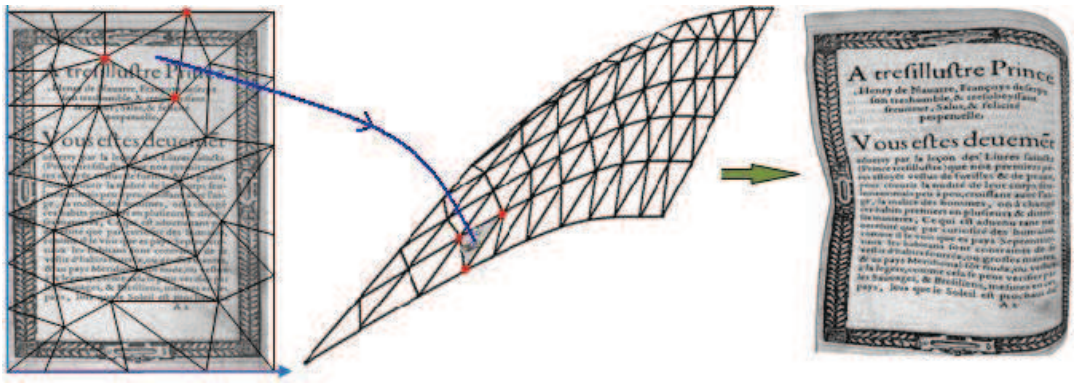


FIGURE 4.7: Processus de plaquage d'une image sur un maillage pour produire une image déformée.

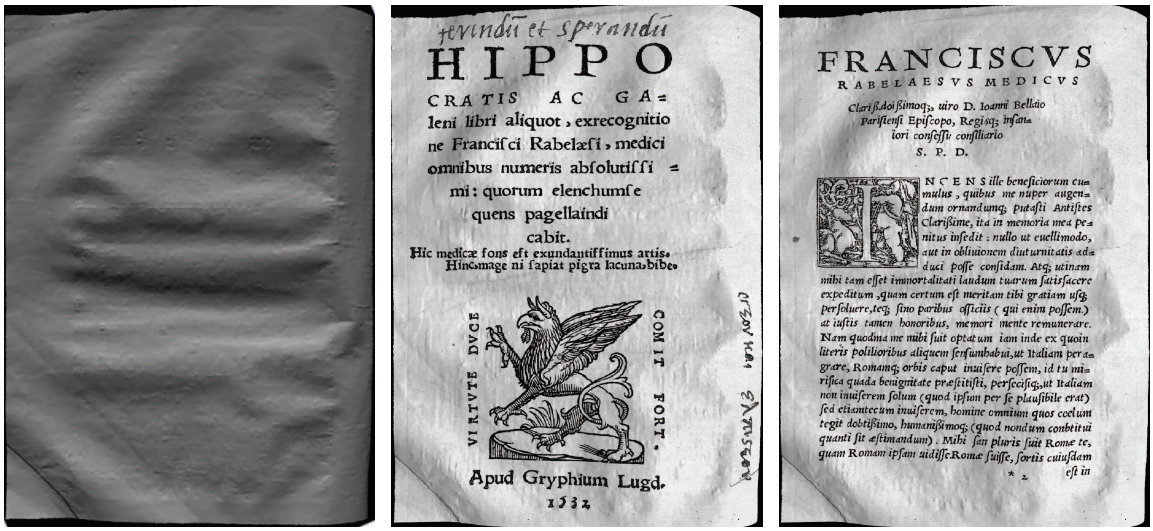
La figure 4.7 résume le travail réalisé à cette étape. A chaque triangle (2D) de l'image va correspondre à un voxel du maillage 3D. Pour générer l'image déformée la plus réaliste possible, le modèle de réflexion de [Phong, 1975] est utilisé pour simuler la luminosité projetée sur la page. Ainsi selon la déformation, les zones de la page seront plus ou moins illuminées.

4.1.2 Évaluation du modèle

Dans cette section, nous essayons de montrer que les distorsions générées sont réalistes. Par la suite nous proposons une technique permettant d'évaluer le niveau de distorsion global et local de chaque maillage.

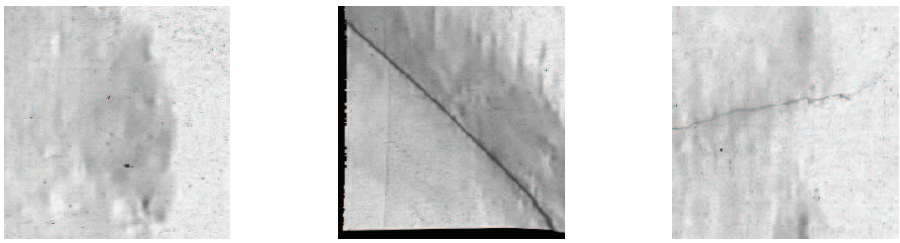
4.1.2.1 Évaluation qualitative

Nous générons des images déformées à partir de 12 maillages afin de montrer notre capacité à générer des images de documents reproduisant des distorsions réelles. La figure 4.8 montre deux images générées en utilisant un même maillage. L'image 4.8-b présente exactement les mêmes distorsions que l'image 4.8-c. On peut y voir dans ces images des distorsions concaves ou convexes, des pliures de tailles différentes ou de petites déchirures, etc. La figure 4.9 permet de voir plus nettement certaines zones extraites de la figure 4.8-b-c. L'image 4.8-b est plus claire que celle de la figure 4.8-c, car les deux textures sont différentes donnant lieu à l'étape d'ajout de luminosité du modèle de [Phong, 1975].



(a) un maillage original (b) image déformée autre image déformée (c)

FIGURE 4.8: Images déformées par le modèle proposé : (a) un maillage original, (b) et (c) deux images déformées en utilisant le maillage (a).

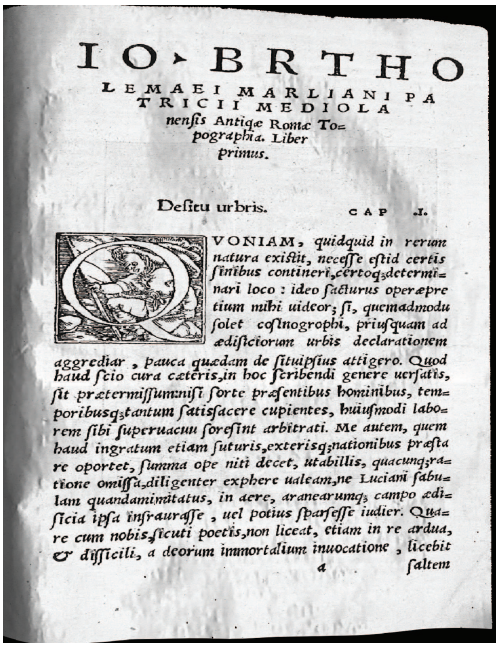


(a) une concavité (b) un pli (c) légère déchirure du papier

FIGURE 4.9: Des distorsions sont reproduites par le modèle proposé.

La luminosité calculée par le modèle de Phong permet de rendre encore plus réaliste les déformations générées. En effet, cette luminosité diffère en fonction du type de distorsions. Par exemple, la figure 4.10-b montre une troisième image dégradée avec le même maillage que dans la figure 4.8 dont l'effet de lumière est différent des précédentes générations. La figure 4.11 illustre localement ce phénomène.

(a) un maillage

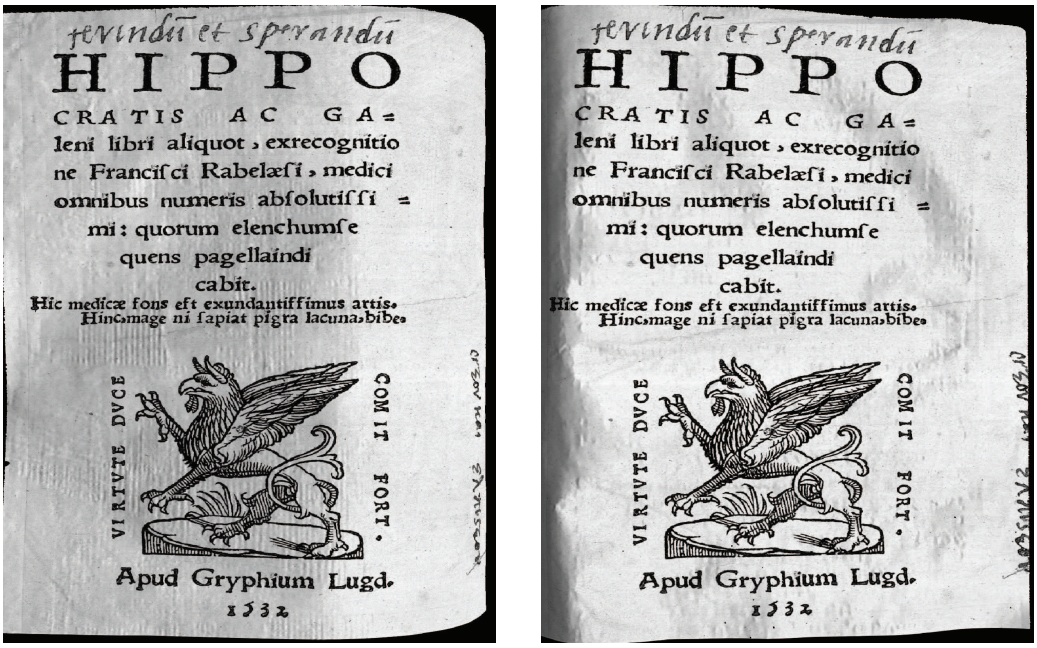


(b) une image générée en utilisant le maillage (a)

FIGURE 4.10: Effet de la luminosité dans des images générées par notre modèle.

FIGURE 4.11: Impact de l'ajout de luminosité sur deux zones dégradées similaires .

A ce stade, nous avons décidé de classer nos maillages selon leur déformation principale : globale et/ou locale. L'image 4.12-a présente un maillage que nous avons considéré comme contenant principalement une déformation globale (ici déformation due à la reliure). Selon cette classification, l'image 4.12-b contient un ensemble de déformations locales réparties sur toute la page. Cette classification des maillages sera utilisée dans le chapitre suivant pour les tests d'évaluation de performances.

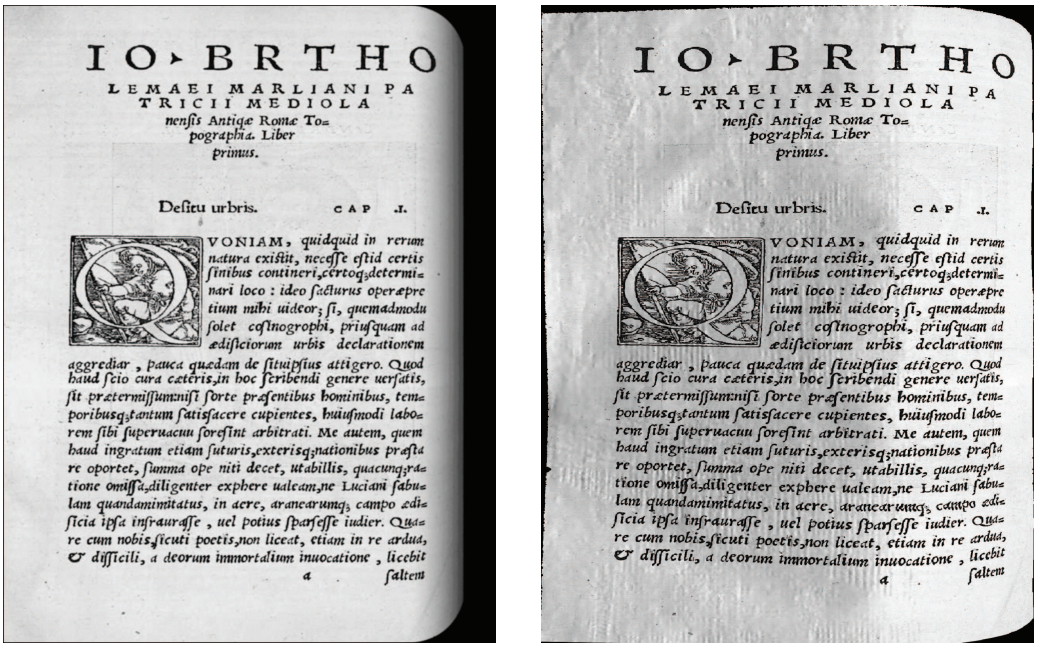


(a) Image générée contenant une distorsion globale

(b) Image générée contenant des distorsions locales

FIGURE 4.12: Déformation globale ou locale générée par deux types différents de maillages.

Nous avons également généré des images différentes pour montrer que le modèle proposé peut générer des distorsions plus réalistes que celles de modèles existants. La figure 4.13-a montre une image générée par le modèle de [Kanungo *et al.*, 1993] qui permet de simuler la courbure générée à cause d'une reliure épaisse. Cette courbure a la caractéristique visuelle d'être homogène du haut en bas de la page. Aucune "nuance" n'apparait dans la déformation générée lui donnant de ce fait un visuel très synthétique. En revanche, l'image générée par notre modèle (cf. Figure 4.13-b) apporte plus de nuances sur les déformations générées rendant plus naturelle l'image générée.

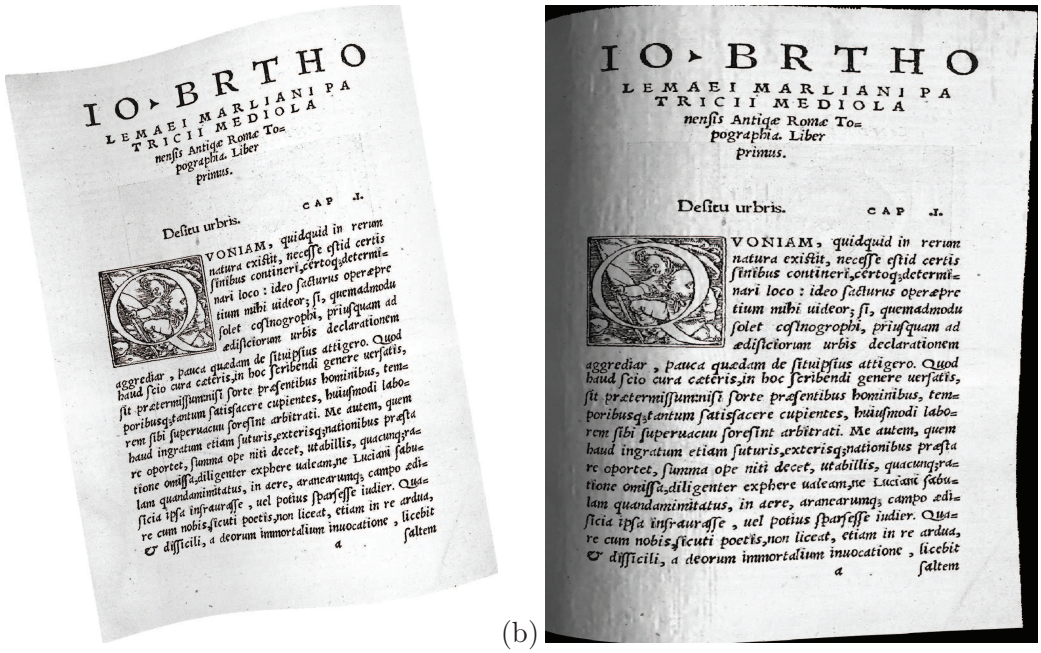


(a) Image générée par le modèle de Kanungo [Kanungo *et al.*, 1993]

(b) Image générée par notre modèle

FIGURE 4.13: Comparaison entre l'image générée par le modèle de Kanungo (a) et celle de notre modèle en 3D (b).

Enfin, la figure 4.14-a montre une image générée par le modèle de [Liang *et al.*, 2008]. Un peu comme avec le modèle de Kanungo, la courbure de cette image est trop régulière et répétitive pour être réaliste visuellement. La figure 4.14-b générée avec notre modèle nous semble posséder une variété de déformations rendant le tout moins synthétique qu'avec le modèle de Liang.



(a) (b)

FIGURE 4.14: Comparaison entre l'image générée par modèle de Jiang (a) et celle de notre modèle en 3D (b).

Dans l'annexe A, nous présentons 12 maillages qui contiennent des distorsions globales et locales très représentatives. Á côté de chaque maillage, une image dégradée d'exemple est générée par notre modèle.

4.1.2.2 Catégorisation du maillage

Dans la sous-section précédente, nous montrons visuellement des exemples d'images générées par notre modèle. Ces images contiennent des distorsions globales et locales très réalistes par rapport aux images générées par les modèles de [Kanungo *et al.*, 1993] et de [Liang *et al.*, 2008]. En ce qui concerne la vérité terrain associée à chaque maillage, nous l'avons réalisée en associant manuellement une étiquette indiquant si le maillage contient (ou pas) des dégradations locales ou globales. De même, nous attribuons manuellement un qualificatif (faible, intermédiaire, fort) relatif à l'importance de chaque dégradation. Si cette solution nous a permis de réaliser plusieurs tests (cf chapitre suivant), nous avons étudié la possibilité de calculer automatiquement des caractéristiques sur ce maillage nous permettant de catégoriser finement le type et l'importance de chaque dégradation qui y est présente. Nous souhaitons, par exemple, sur un maillage comme celui montré figure 4.15, pouvoir automatiquement détecter la présence de déformations globales et locales dans le document

afin de s'en servir comme vérité terrain. *In fine*, notre objectif est de pouvoir sélectionner certains types de maillage pour étudier la robustesse de logiciels tels que des OCR.

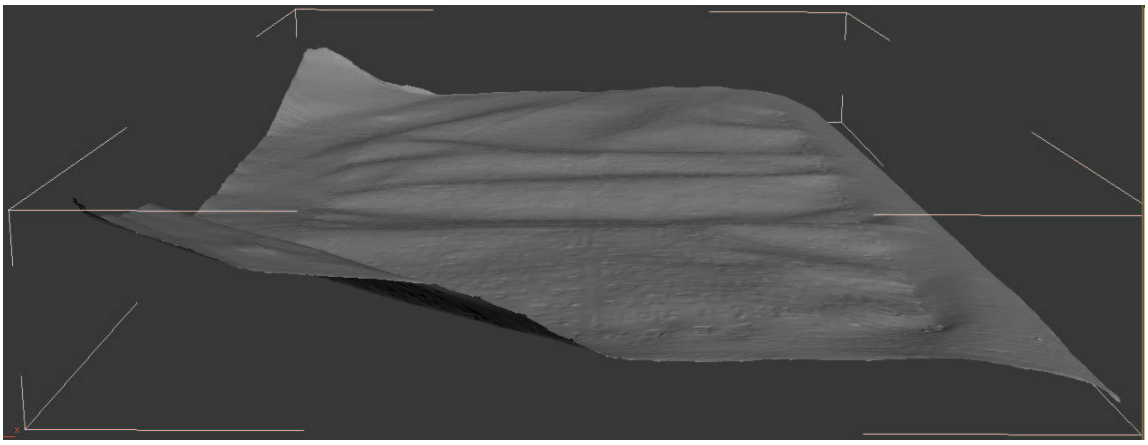


FIGURE 4.15: Exemple d'un maillage contenant des distorsions globales et locales

Pour réaliser cette objectif de catégorisation automatique du maillage, nous proposons d'utiliser une méthode d'analyse géométrique afin de savoir si le maillage contient des régions de distorsions globales et/ou locales. Soit un plan P parallèle avec le plan Oxy. Ce plan glisse de bas en haut du maillage. A chaque itération, ce plan coupe le maillage avec ce que nous appelons des lignes d'intersection (cf. figure 4.16). Ces lignes représentent des régions de distorsions pour chaque plan P.

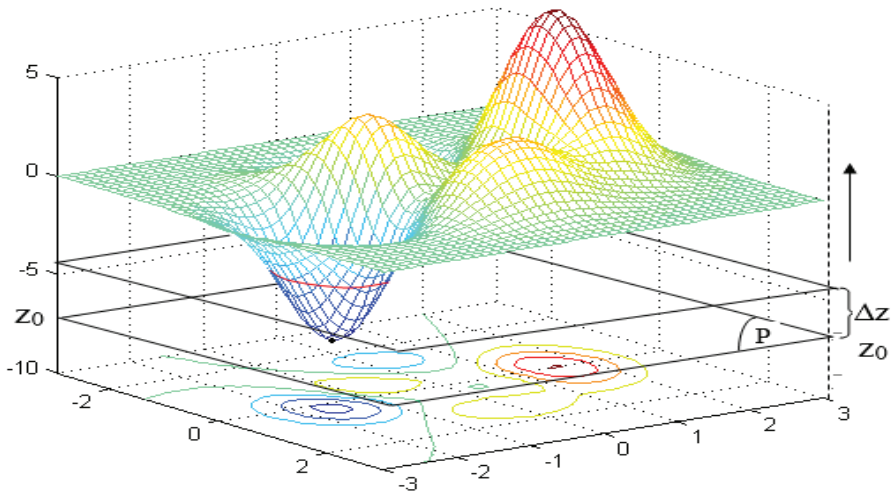
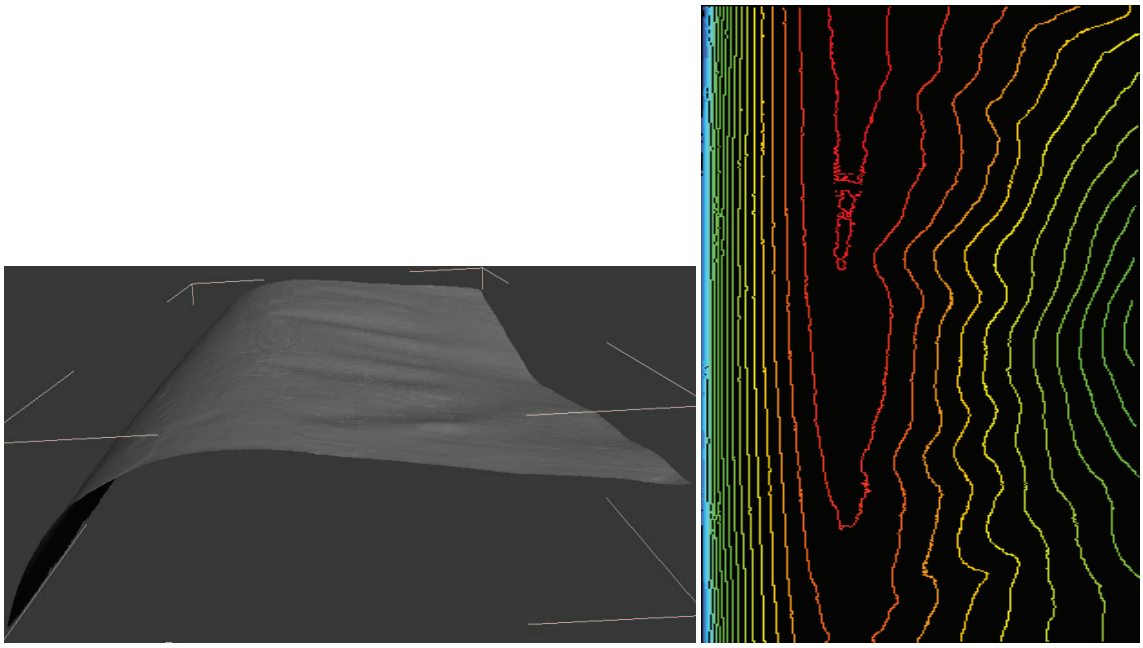


FIGURE 4.16: Maillage 3D analysé selon plusieurs plans P. Les lignes d'intersection sont projetées sur le plan 2D.



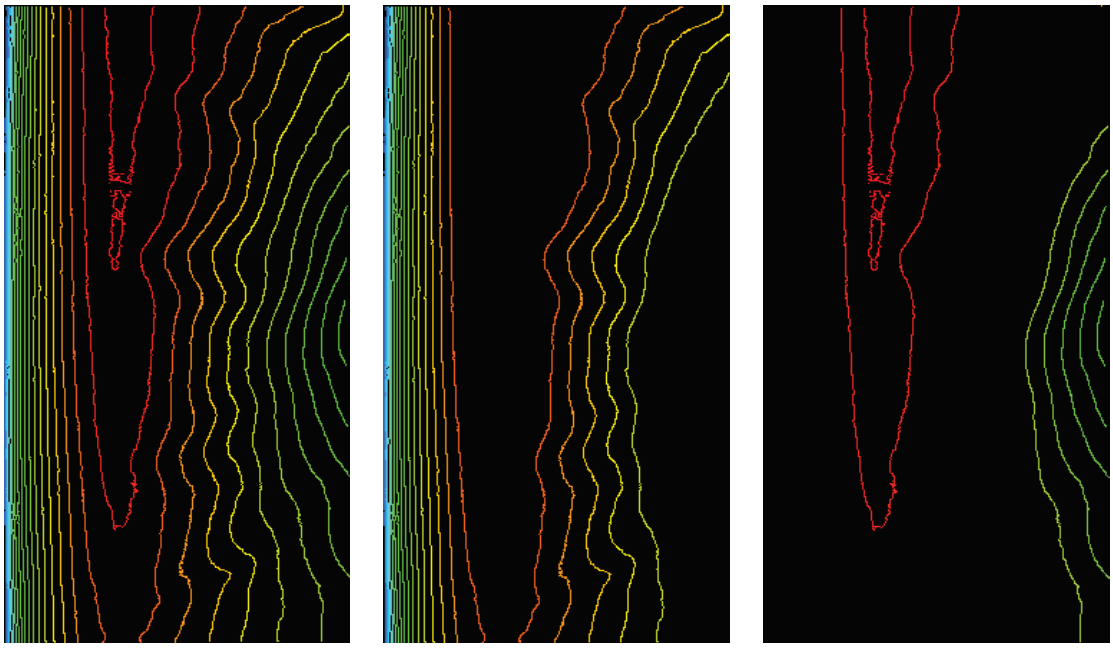
(a) Maillage M2 en 3D

(b) lignes d'intersection représentant les régions de distorsion du maillage M2

FIGURE 4.17: Le maillage M2 (a) est coupée par un plan P pour produire des lignes d'intersection. Ces lignes d'intersection sont projetées sur le plan 2D (b).

La figure 4.17 montre les lignes d'intersection calculées à partir du maillage M2. Les lignes d'intersection sont projetées sur le plan 2D comme le montre la figure 4.17-b. La hauteur de chaque ligne est représentée par la couleur de cette ligne. Le bleu correspond aux points (3D) les plus bas et en rouge les plus haut. De longues lignes représentent les régions de type distorsion globale (par exemple la courbure à gauche du document) tandis que de courtes lignes représentent des régions possédant des distorsions locales (petites ondulations sur la droite du document).

Ce sont ces lignes que nous analysons pour déterminer quel type de déformation est présente dans un document. Nous séparons d'abord en deux catégories les lignes : celles qui traversent entièrement de haut en bas une page d'un document (figure 4.18-b) et qui correspondent à des distorsions globales ; et celles qui ne traversent pas entièrement la page de haut en bas (figure 4.18-c) et qui correspondent à des distorsions locales.



(a) Les lignes d'intersection du maillage M2

(b) lignes correspondant à des distorsions globales

(c) lignes correspondant à des distorsions locales

FIGURE 4.18: Les lignes d'intersection du maillage M2 (a) les distorsions sont séparées en deux catégories : globales (b) et locales (c).

Après avoir séparé les régions, nous analysons deux caractéristiques géométriques de chacune d'entre elles : le nombre de régions de type distorsion locale (nbR_{local}) et une proportion correspondant à la surface totale des régions de type distorsion globale et la surface totale des distorsions de type locales ($S_{global/local}$). Une surface est calculée à partir de la boîte englobante d'une ligne. Si deux boîtes englobantes ont au moins un certain pourcentage de surface commune, alors une seule composante est comptabilisée. Nous avons fixé pour nos tests ce pourcentage à 90%.

Ensuite, nous utilisons l'information de hauteur de chaque ligne (représentée par des couleurs sur la figure 4.18). Nous calculons la moyenne et l'écart-type des lignes de type déformation globale ($\overline{\Delta h_{global}}$, σ_{global}) et la hauteur moyenne et l'écart-type des lignes de type déformation locale ($\overline{\Delta h_{local}}$, σ_{local}). Nous calculons également la hauteur maximale des lignes de déformation de type locale.

La table 4.1 indique les valeurs des 7 caractéristiques géométriques calculées sur l'ensemble de nos maillages (cf visuel des maillages en annexe A).

TABLE 4.1: Les 7 caractéristiques géométriques sont extraites à partir de 12 maillages : le nombre de régions de type distorsion locale (nbR_{local}), la proportion de la surface régions globales / régions locales ($S_{global/local}$), la hauteur moyenne et écart-type des lignes globales ($\overline{\Delta h_{global}}$, σ_{global}), la hauteur moyenne et écart-type des distorsions locales ($\overline{\Delta h_{local}}$, σ_{local}), la hauteur maximale de distorsions locales Max(h_{local}).

Maillage	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$S_{global/local}$	**3.00**	**3.90**	1.09	1.91	0.46	0.38	0.48	0.96	0.81	0.45	0.39	0.6
$\overline{\Delta h_{global}}$	**565**	**772**	294	249	160	213	174	309	261	83	120	141
σ_{global}	**2.28**	**2.77**	0.72	0.58	0.39	0.52	0.52	0.75	0.64	0.48	0.32	0.7
nbR_{local}	2	2	6	6	7	5	6	**8**	3	3	7	**8**
$\overline{\Delta h_{local}}$	105	103	145	177	200	163	**268**	**206**	198	196	109	204
σ_{local}	1.15	1.12	0.63	0.79	0.80	0.57	**2.30**	0.81	0.63	0.40	0.33	0.64
Max(h_{local})	222	227	193	231	274	223	**548**	**372**	253	246	196	267

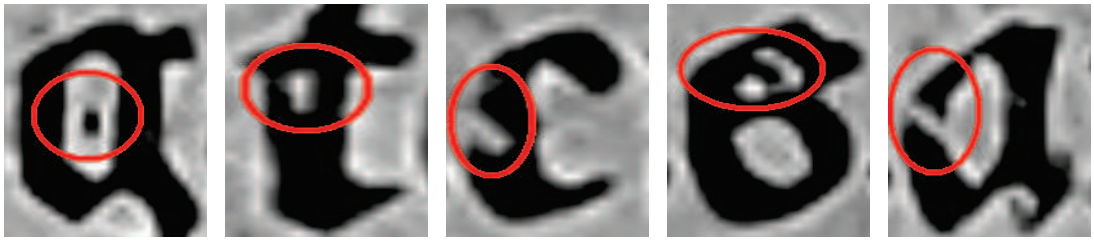
L'intérêt du calcul de ces caractéristiques est que l'on peut désormais proposer à un utilisateur de sélectionner un maillage selon divers critères : peu ou beaucoup de chacune des dégradations, des déformations plus ou moins fortes, de forts écarts au sein du même type de dégradation, etc. Il suffit simplement de trier les maillages selon l'un des 7 critères calculé. Par exemple, les maillages 1 et 2 sont ceux contenant des déformations globales les plus variées en termes de hauteur de déformation. Le maillage 12 est celui contenant le plus de déformations locales etc.

4.2 Contribution 2 : Proposition d'un modèle de bruit local

Cette section détaille notre modèle de bruit local dont l'objectif est de simuler les taches sombres et claires. Ces taches sont fréquemment présentes dans le premier-plan des images (*e.g.* éléments graphiques, illustrations, caractères, etc.). Ces taches sont dues au processus d'impression (taches d'encre qui tombent/bavent à coté des tampons), à l'âge du document, etc. Pour simplifier le propos, ous allons étudier le cas de caractères d'images de documents, particulièrement de documents anciens. Cette proposition du modèle est motivée par deux réalités. Premièrement, ces taches modifient la forme, le contour, ou la connectivité des caractères. Cela pose des difficultés aux algorithmes d'analyse de documents (voir la sous-section 3.1.2 et la figure 3.15). Deuxièmement, la majorité de modèles de bruit

existants s'applique sur des images binaires [Baird, 1990, Kanungo *et al.*, 1993, Zhai *et al.*, 2003, Smith, 2008]. Par conséquent, ils ne sont pas adaptés aux images de documents anciens en niveaux de gris.

Nous proposons de diviser en trois groupes les types de défauts apparaissant sur ou à proximité des caractères : les taches isolées, les taches connectées, les déconnexions. Les taches isolées sont sombres (la figure 4.19-a) ou claires (la figure 4.19-b). Elles sont généralement localisées sur le fond de l'image. Dans ce modèle, nous nous intéressons uniquement aux taches isolées qui sont à proximité (à l'intérieur ou à l'extérieur) de la bordure d'un caractère sans jamais le toucher. Les taches connectées sont celles qui touchent la bordure d'un caractère, mais ne créent pas de discontinuité (si la tache est claire comme dans la figure 4.19-c) ou ne fusionnent pas deux caractères (si la tache est sombre comme dans la figure 4.19-d). Les défauts de type déconnexion sont uniquement les taches claires qui coupent la connectivité du caractère (figure 4.19-e).



(a) (b) (c) (d) (e)

FIGURE 4.19: Trois types de bruit dans les documents réels : (a), (b) Taches isolées ; (c), (d) Taches connectées ; et (e) Déconnexions.

4.2.1 Présentation du modèle de bruit local

Les trois types de bruit (cf. figure 4.19) apparaissent à côté de la bordure des caractères. Cette caractéristique a été à l'origine de notre modèle, c'est cette spécificité visuelle que nous avons cherché à reproduire synthétiquement. Pour cela, l'idée principale de notre modèle est, pour chaque pixel d'une image, de définir s'il devient ou non un point de dégradation en se basant (entre autres) sur sa distance avec le contour du caractère le plus proche.

Notre modèle de dégradation s'appuie sur trois étapes résumées dans la figure 4.20 : (A) sélection de points de dégradation, (B) classification des points de dégradation en trois types, et (C) génération de bruit local.

Dans un premier temps, un algorithme pseudo-aléatoire permet de placer les points où seront localisées les centres des futures dégradations (*cf.* figure 4.20.A). Ces points sont appelés des "points de dégradation". Dès lors, une ellipse est dessinée avec des dimensions calculées en fonction de la position du point par rapport aux contours du caractère le plus proche et une direction déterminée en fonction du gradient local au point de dégradation (figure 4.20.B). Nous avons choisi la forme elliptique dont la forme dépend de deux axes : axe semi-majeur et axe semi-mineur). Le changement de ces deux valeurs peut produire des formes très variables. Enfin, chaque pixel à l'intérieur de l'ellipse est modifié pour donner l'impression que cette zone est dégradée (figure 4.20.C).

Ce processus de génération permet à un utilisateur de fixer lui-même une quantité de dégradation à générer. Ceci revient à définir un nombre de points de dégradation sur lesquels le modèle sera appliqué. Une catégorisation permettra d'appliquer de générer des dégradations différentes selon la localisation des points générés. Cette catégorisation représente le souhait de l'utilisateur en termes de génération de types de défauts. Pour fixer facilement les paramètres dans ce modèle, nous définissions n_{user} comme la quantité de dégradation désirée par l'utilisateur. Elle est spécifiée `via` un seul paramètre (à fixer entre 0 et 100). Les trois pourcentages I, O, D correspondent au pourcentage de taches isolées, de taches connectées, et de déconnexions.

Nous proposons donc un modèle simple à paramétrer et qui tend à reproduire fidèlement les défauts liés à l'encre.

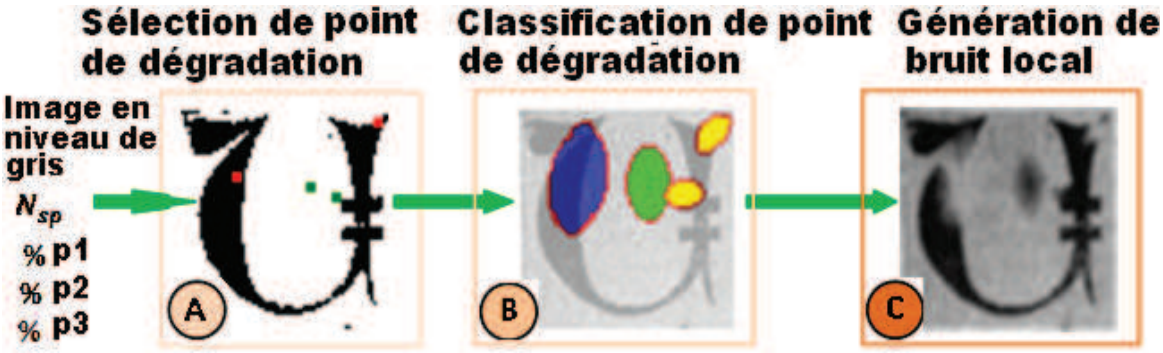


FIGURE 4.20: Les trois grandes étapes constituant notre modèle de dégradation de l'encre à proximité ou sur les caractères

4.2.1.1 Étape A : sélection des points de dégradation

Dans cette étape, nous utilisons un algorithme pseudo-aléatoire qui permet de sélectionner des points de dégradation en fonction de leur distance minimale à la bordure des caractères. D'abord la probabilité w_{p_i} de chaque pixel p_i (x, y) est calculée en fonction de sa distance minimale à un bord d'une composante connexe. Puis, un ensemble de N pixels qui ont les plus grandes probabilités de devenir un point de dégradation est défini. Toute la difficulté de cette étape est de choisir les points de dégradation permettant de générer des dégradations réalistes. Le nombre de N pixels est expérimentalement fixé $n = (1 + I)\times n_{user}$ où I est le pourcentage de taches isolées et n_{user} est la quantité de bruit souhaité par l'utilisateur. Cet algorithme est résumé dans l'algorithme 4.2.1.

Algorithm 4.2.1: SelectionPointsDeDegradation$(n_{sp_{user}}, I, D)$

$N \leftarrow \{\phi\}\,; n \leftarrow 0; BI \leftarrow$ *tous les pixels d'image binarisee*
for each *pixel* : $p_i \in BI$
do
- $d_i \leftarrow distance(p_i$ **to** *le plus proche bordure d'un caractre*$)$;
- $\alpha \sim U[0, +\infty]; \beta \sim U[0, +\infty]$;
- **if** $(p_i$ **is** *un pixel du background*$)$
  - **then** $w_{p_i} \leftarrow (e^{-\beta d_i} - D)$;
  - **else** $w_{p_i} \leftarrow (e^{-\alpha d_i})$;

while $(n < ((1 + I) \times n_{sp_{user}}))$
do
- $p \leftarrow max\_proba(BI)$;
- $BI \leftarrow BI \setminus \{p\}\,;$
- $N \leftarrow \{N; p\}\,;$
- $n \leftarrow n + 1$;

 # max_proba(BI) retourne le pixel qui a la plus grande probabilite w_{pi} dans BI
return (N)

Les deux paramètres α et β sont aléatoirement générés selon une distribution uniforme. Le pourcentage D de déconnexions est ajouté dans la formule de probabilité des pixels de l'arrière-plan $w_{p_i} = (e^{-\beta d_i}$ - $D)$ afin de contrôler le ratio de points de dégradation de l'arrière-plan (qui deviendront les taches noires) et ceux du premier-plan (qui deviendront les taches blanches).

Par exemple, la figure 4.21 montre 4 images avec des points sélectionnés avec comme valeurs différentes $D = 10\%$, 30%, 80%, et 100%. Dans le cas où $D = 100\%$, tous les points de dégradation deviendront des déconnections (taches blanches). Ils sont donc obligatoirement à l'intérieur des caractères (points rouges) comme dans la figure 4.21-d. Lorsque D=100%, les probabilités de tous les pixels de l'arrière-plan (à l'extérieur de caractères) $w_{p_i} = (e^{-\beta d_i}$

- 1) sont inférieures à 0. Dans ce cas-là, ces pixels ne sont jamais sélectionnés. L'algorithme sélectionne donc seulement des pixels à l'intérieur de caractères.

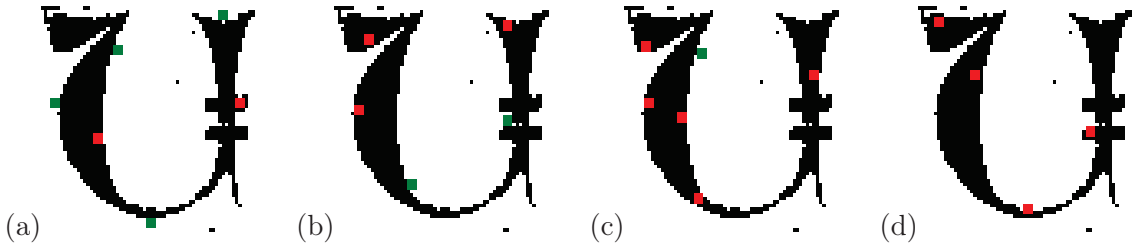


FIGURE 4.21: Les points de dégradation sélectionnés aléatoirement quand le pourcentage D diminue ($n_{sp_{user}}$ est fixé = 4) : (a) I=50%, D=10% , (b) I=50%, D=30%, (c) I=10%, D=80%, (d) I=0%, D=100%

Puisque notre objectif principal est de générer des taches dans les caractères réels qui apparaissent à côté de la bordure des caractères, notre algorithme est créé pour choisir les points de dégradation proches de la bordure de caractères. Dans cet algorithme, la probabilité d'un pixel de devenir point de dégradation est inversement proportionnelle à la distance minimale de ce point à la bordure du caractère le plus proche. Par conséquent, plus les pixels sont proches de la bordure, plus leurs probabilités d'être transformés sont grandes.

L'algorithme 4.2.1 sélectionne N points dont leurs probabilités d'être des points de dégradation est la plus grande. La figure 4.22 montre 4 images de points de dégradation sélectionnés avec les mêmes paramètres et la même image en entrée. Les points choisis sont proches de la bordure.

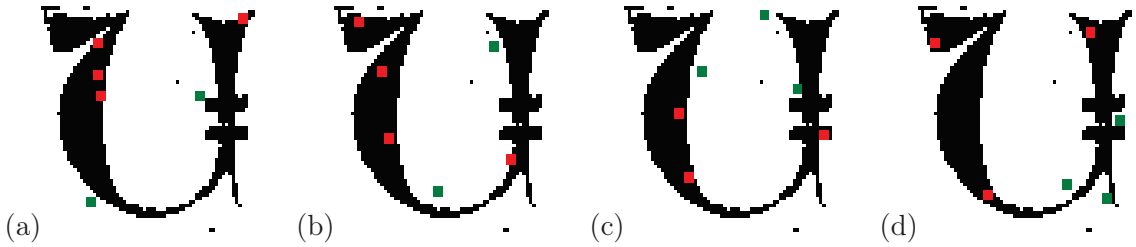


FIGURE 4.22: Les points de dégradation sélectionnés aléatoirement 4 fois avec les mêmes paramètres $n_{sp_{user}}$ = 4, I=50%, D=20% : (a) et (b) contiennent 4 points de dégradation à l'intérieur (rouge) et 2 autres à l'extérieur (vert), (c) et (d) contiennent 3 points de dégradation à l'intérieur et 3 autres à l'extérieur

4.2.1.2 Étape B : classification des points de dégradation en trois types

Le but principal de cette étape est d'affecter à chaque point de dégradation l'une des trois catégories de dégradation possible : les taches isolées, les taches connectées, et les taches de type déconnexion. La difficulté est d'attribuer au bon point de dégradation, le type de défaut à générer. Une mauvaise affectation peut impliquer des résultats visuellement trop synthétiques (et donc peu réalistes). Par exemple, se servir d'un point de dégradation éloigné d'un caractère pour synthétiser une tache noire de type connectée (la figure 4.23.a), a pour conséquence de générer une tache noire trop grosse (non-réaliste).

(a)  (b) 

FIGURE 4.23: (a) Une tache noire de type connecté non-réaliste, (b) une tache noire de type connecté avec le même point de dégradation mais plus réaliste.

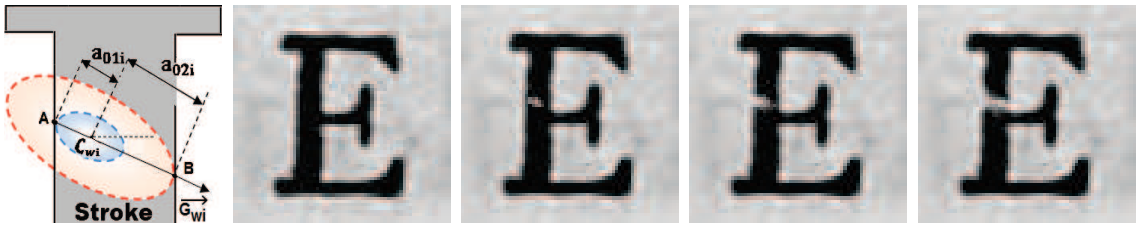
En sortie de première étape, nous disposons de l'ensemble des points de dégradation N. Certains sont des points situés à l'intérieur d'un caractère (un point d'encre devient une tache claire) et d'autres sont des points localisés à l'extérieur d'un caractère (un point clair du fond devient une tache sombre). Ces points correspondent aux ensembles C_w et C_b dont C_{wi} et C_{bi} sont respectivement un point de dégradation à l'intérieur d'un caractère et l'autre à l'extérieur d'un caractère. Chaque point sera à l'origine d'une tache de forme elliptique : le point C_{wi} va produire une tache claire tandis que C_{bi} va produire une tache sombre. Le type de chaque point de dégradation dépend de deux distances a_{01} et a_{02} à partir de ce point aux deux points d'intersection le plus proche. Les deux points d'intersection correspondent aux deux plus proches bords de la forme dans la direction du gradient. L'affectation de chaque point de dégradation est l'un des trois types est réalisé dans l'algorithme 4.2.1.

Avant de détailler l'algorithme permettant d'attribuer chaque point de dégradation à chacune des trois catégories de défauts, nous détaillons ci-dessous la relation entre la valeur de a_{wi} correspondant à la taille du grand axe de l'ellipse (forme utilisée pour dégrader l'image) et le type de dégradation généré.

Supposons que le point de dégradation soit situé dans un caractère (un pixel d'encre va générer une zone claire). Soit $\overrightarrow{G_{wi}}$ le vecteur gradient au point de dégradation C_{wi}. Le

vecteur $\overrightarrow{G_{wi}}$ coupe les deux plus proches bordures du caractère en A et B. On a donc $a_{01i} = AC_{wi}$ et $a_{02i} = C_{wi}B$ avec comme contrainte $a_{01i} \leq a_{02i}$. Selon la valeur donnée à a_{wi} (soit la longueur du grand axe de l'ellipse) la dégradation générée ne sera pas la même. La direction du grand axe de l'ellipse est donnée par le vecteur gradient $\overrightarrow{G_{wi}}$ (cf. 4.24-a). Nous avons discerné 3 cas de figure différents :

- C_{wi} est une tache isolée et claire si $1 \leq a_{wi} < a_{01i}$ (e.g. 4.24-c)
- C_{wi} est une tache connectée et claire si $a_{01i} \leq a_{wi} < a_{02i}$ (e.g. 4.24-d)
- C_{wi} est une tache déconnexion et claire si $a_{02i} \leq a_{wi}$ (e.g. 4.24-e)

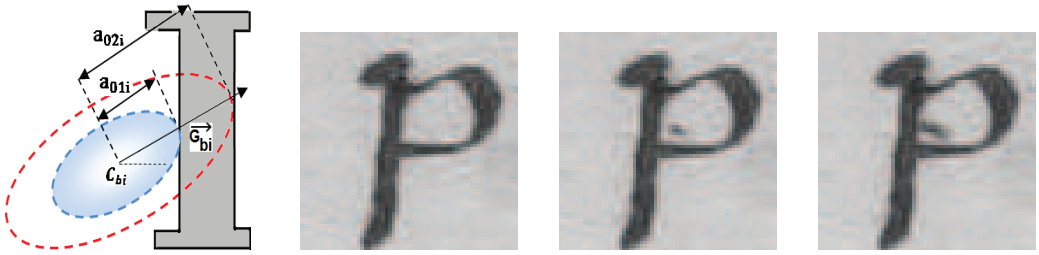


(a) (b) (c) (d) (e)

FIGURE 4.24: Exemple d'un point de dégradation situé à l'intérieur du caractère "E" (b) : (c) tache isolée, (d) tache connectée, (e) tache générant une déconnexion.

De la même manière, supposons que le point de dégradation soit situé à côté d'un caractère (un pixel de fond va générer une zone sombre). Soit $\overrightarrow{G_{bi}}$ le vecteur gradient au point de dégradation C_{bi}. Le vecteur $\overrightarrow{G_{bi}}$ coupe les deux plus proches bordures de caractères par A et B. On a donc $a_{01i} = AC_{bi}$ et $a_{02i} = C_{bi}B$ où $a_{01i} \leq a_{02i}$. Supposons que a_{bi} soit la longueur du grand axe de la tache elliptique à ce point. Cet axe suit la direction du vecteur gradient $\overrightarrow{G_{bi}}$ (cf. 4.25-a). Selon la dimension choisie pour a_{bi} la dégradation générée est différente :

- C_{bi} est une tache isolée (sombre) si $1 \leq a_{bi} < a_{01i}$ (e.g. 4.25-c)
- C_{bi} est une tache connectée (sombre) si $a_{01i} \leq a_{bi} < a_{02i}$ (e.g. 4.25-d)



(a) (b) (c) (d)

FIGURE 4.25: Exemple d'un point de dégradation à l'extérieur du caractère "P" (b) : (c) tache isolée, (d) tache connectée.

Que ce soit pour générer une dégradation sombre ou claire, c'est la valeur choisie du grand axe de l'ellipse qui va conditionner la forme de la tache générée. Afin de respecter les proportions des différentes taches que souhaite générer l'utilisateur tout en produisant des dégradations visuellement réalistes, notre proposition est de classer chaque point de dégradation en l'un des trois types. Pour cela, nous nous appuyions sur les deux distances a_{01} et a_{02} de chaque point. Supposons que N_{is}, N_{os}, N_{ds} soient respectivement l'ensemble des points de dégradation du type taches isolées, l'ensemble des taches connectées, et l'ensemble des taches de type déconnexion. Le nombre total de points de dégradation des trois ensembles doit être égal au nombre donné par l'utilisateur : $|N_{is}| + |N_{os}| + |N_{ds}| = n_{user}$. Le processus de classification est au final implémenté comme décrit dans l'Algorithme 4.2.2. Ce processus de classification est résumé ci-dessous :

- Choisir $|N_{ds}|$ points de dégradation dans N, dont la distance a_{02} est minimale pour devenir des taches de type déconnexion.
- Puis, choisir $|N_{os}|$ points de dégradation dans le reste de N, dont la distance a_{01} est minimale pour devenir des taches de type connectées.
- Finalement, choisir $|N_{is}|$ points de dégradation dans le reste de N, dont la distance a_{01} est maximale pour devenir des taches isolées.

Algorithm 4.2.2: ClassificationDesPointsDegradation($n_{sp_{user}}, I, O, D$)

$N \leftarrow$ SelectionPointsDeDegradation($n_{sp_{user}}, I, D$);
$N_{is}, N_{os}, N_{ds} \leftarrow \{\phi\}\,; sp1, sp2, sp3 \leftarrow \{\phi\}\,;$

while $(|N_{ds}| < (D \times n_{sp_{user}}))$ **do**
- $sp1 \leftarrow min\_a02(N)$;
- **if** ($sp1$ *est un pixel d encre*)
  - **then**
    - $N \leftarrow N \setminus \{sp1\}\,;$
    - $N_{ds} \leftarrow \{N_{ds}; sp1\}\,;$

\# $min\_a02(N)$ *retourne un pixel ayant le plus petit* a_{02} *dans* N

while $(|N_{os}| < (O \times n_{sp_{user}}))$ **do**
- $sp2 \leftarrow min\_a01(N)$;
- $N \leftarrow N \setminus \{sp2\}\,;$
- $N_{os} \leftarrow \{N_{os}; sp2\}\,;$

\# $min\_a01(N)$ *retourne un pixel ayant le plus petit* a_{01} *dans* N

while $(|N_{is}| < (I \times n_{sp_{user}}))$ **do**
- $sp3 \leftarrow max\_a01(N)$;
- $N \leftarrow N \setminus \{sp3\}\,;$
- $N_{is} \leftarrow \{N_{is}; sp3\}\,;$

\# $max\_a01(N)$ *retourne un pixel ayant le plus grand* a_{01} *dans* N

return (N_{is}, N_{os}, N_{ds})

Ce processus de sélection nous permet d'obtenir la meilleure affectation possible afin de générer des dégradations réalistes. La figure 4.26-a illustre l'importance qu'il y a à bien choisir le type de défaut associé à chaque point. Supposons que l'utilisateur ait besoin, pour les 4 points de dégradation, d'une tache déconnexion, d'une tache connectée et de 2 taches isolées. Puisque l'attribution du type déconnexion est déterminée en premier, et que les points de dégradation de ces taches sont à l'intérieur du caractère, C_1 et C_3 sont sélectionnés en premier comme potentiels points de type déconnection. La distance a_{02} de C_1 est plus petite que celle de C_3. Seul le point C_1 devient donc une tache de type déconnexion. Notre algorithme évite de choisir des points tel que C_3 devienne une tache de type déconnexion. Ceci évite de génèrer une tache grosse et visuellement peu réaliste comme illustré sur la figure 4.26-c.

Ensuite, le point qui va devenir une tache connectée sera choisi dans la liste de points restants C_2, C_3, et C_4. Selon la position de ces trois points, la distance a_{01} de C_2 est inférieure à celles de C_3 et C_4. Par conséquent, C_2 devient une tache connectée. Si C_3 ou C_4 avaient été choisies pour devenir des taches connectées au caractère, cela aurait généré des résultats peu réalistes comme sur la figure 4.26-d.

Finalement, C_3 et C_4 sont choisis pour devenir deux taches isolées. La figure 4.26-e montre le résultat obtenu.

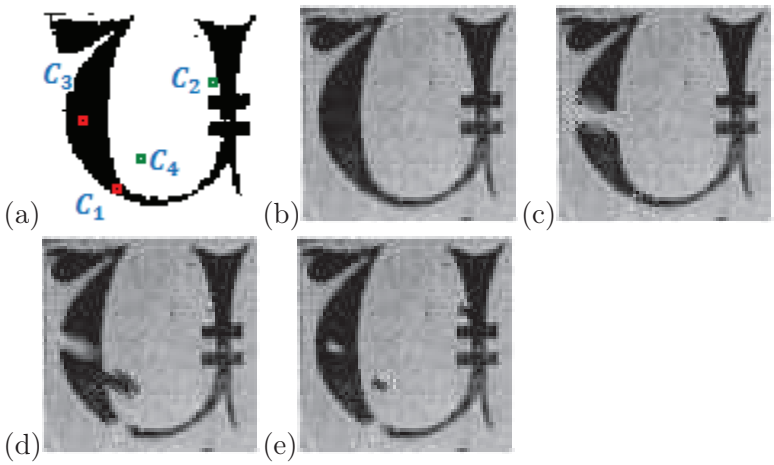


FIGURE 4.26: Exemple d'un processus de classification de 4 points de dégradation du caractère "U".

4.2.1.3 Étape C : génération de bruit

Après avoir choisi des points de dégradation et les avoir classés en trois types, cette ultime étape permet de générer des régions de bruit de forme elliptique dont les dimensions dépendent du type de bruit affecté à chaque point de dégradation lors de l'étape précédente. Cette étape se compose de deux tâches principales : définition de la taille d'une région de bruit et puis génération aléatoire des nouvelles valeurs de niveaux de gris des pixels dans cette région. La difficulté de cette étape est de générer ces valeurs de pixels à l'intérieur d'une région de bruit pour qu'elle ressemble à une région de bruit réel.

La taille d'une région elliptique au point C_i consiste à calculer les longueurs de l'axe principal a_i et du petit axe b_i de cette région elliptique. La taille de l'axe a_i est calculée comme défini dans l'équation 4.1 où a_{01i} et a_{01i} sont les distances du point C_i aux deux plus proches bordures du caractère (cf. 4.24-a et 4.25-a). La taille de l'axe $b_i = a_i \times (1 - g_i)$ où g_i est le facteur d'aplatissement de la région elliptique.

$$a_i = \begin{cases} a_{01i} \times \mu & \text{Si } C_i \in N_{is} \\ a_{01i} + (a_{02i} - a_{01i}) \times \mu & \text{Si } C_i \in N_{os} \\ a_{02i} + 1 & \text{Si } C_i \in N_{ds} \end{cases} \tag{4.1}$$

Dans l'équation 4.1, la valeur μ est choisie aléatoirement entre 0 et 1. Le paramètre g_i permet de jouer sur la forme (aplatissement) de l'ellipse. Par exemple, une tache de type déconnexion est d'autant plus fine que la valeur g_i est proche de la valeur 1 (*e.g* la figure

4.27-c). En revanche, une tache isolée est d'autant plus ronde que la valeur g_i est proche de 0 (*e.g* la figure 4.27-a). Afin de varier les formes générées, nous avons décidé de fixer la valeur de g_i aléatoirement (selon une loi uniforme U). Précisément, $g_i = U[0, \frac{1}{3}]$ si le point de dégradation C_i est une tache isolée, $g_i = U[0, 1]$ si le point de dégradation C_i est une tache connectée, et $g_i = U[\frac{2}{3}, 1]$ si le point de dégradation C_i est une tache de type déconnexion.

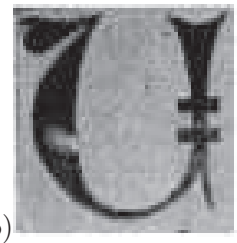
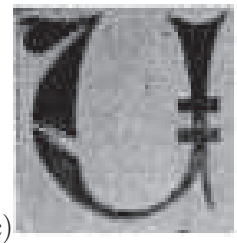
(a)  (b)  (c) 

FIGURE 4.27: L'impact du facteur d'aplatissement g_i selon les types de taches : (a) une tache isolée est ronde avec $g = 0.23$, (b) une tache connectée d'épaisseur moyenne $g = 0{,}46$, (c) une tache déconnectée fine avec $g = 0{,}75$

Puisque nous avons défini la taille et la direction de chaque région de bruit, il ne reste plus qu'à détailler comment les pixels situés à l'intérieur de l'ellipse sont modifiés. Afin de rendre réaliste la dégradation générée, la nouvelle valeur de chaque pixel doit satisfaire les deux conditions suivantes :

- le nouveau niveau de gris d'un pixel ne doit pas être inférieur au niveau de gris minimum et ne pas être supérieur au niveau de gris maximal des pixels situés dans la zone elliptique avant modification.
- L'écart (en niveau de gris) entre de deux pixels adjacents ne doit pas être plus grand que l'écart maximal entre deux pixels situés dans la zone elliptique avant modification.

De ce fait, on calcule les nouvelles valeurs de chaque pixel de la région elliptique comme indiqué sur la figure 4.28. D'abord, la nouvelle valeur du niveau de gris c_i au pixel du centre C_i est définie par une fonction aléatoire $N_c(\mu_c, \sigma_c^2)$ suivant une distribution normale où σ_c est égal à l'écart type des niveaux de gris de la zone avant modification, et μ_c est calculé comme suit :

- μ_c = la moyenne des niveaux de gris de tous les pixels appartenant à l'arrière-plan (le fond) si C_i est à l'intérieur d'un caractère.
- μ_c =la moyenne des niveaux de gris de tous les pixels appartenant au premier-plan (l'encre) si C_i est à l'extérieur d'un caractère.

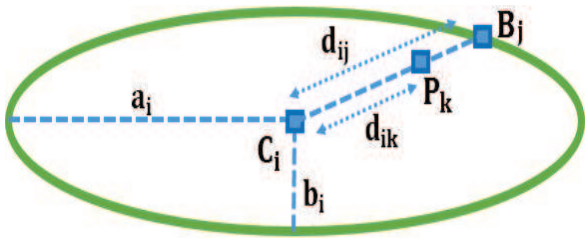


FIGURE 4.28: Dégradation d'une région elliptique.

La nouvelle valeur b_j du pixel B_j situé sur la bordure de la région elliptique est égale à la moyenne de gris de ses 8 voisins. Pour chaque pixel P_k appartenant au segment C_iB_j (cf. Figure 4.28), sa nouvelle valeur de gris est aléatoirement générée par une fonction Gaussienne $N_k(\mu_k, \sigma_k^2)$ où σ_k est égal à l'écart de gris moyen de la zone elliptique avant modification, et μ_k est calculé comme suit :

$$\mu_k = c_i + (b_j - c_i) \times \frac{d_{ik}}{d_{ij}} \quad \text{Où } d_{ik} = C_iP_k \text{ et } d_{ij} = C_iB_j \tag{4.2}$$

Finalement, un filtre Gaussien est appliqué sur la région dégradée pour lisser les différences entre pixels adjacents dont les niveaux de gris diffèrent trop.

4.2.2 Évaluation du modèle

Il est difficile de prouver qu'un modèle peut générer une dégradation réaliste qui s'approche d'une dégradation réelle. De notre point de vue, ce problème peut être résolu de trois manières différentes :

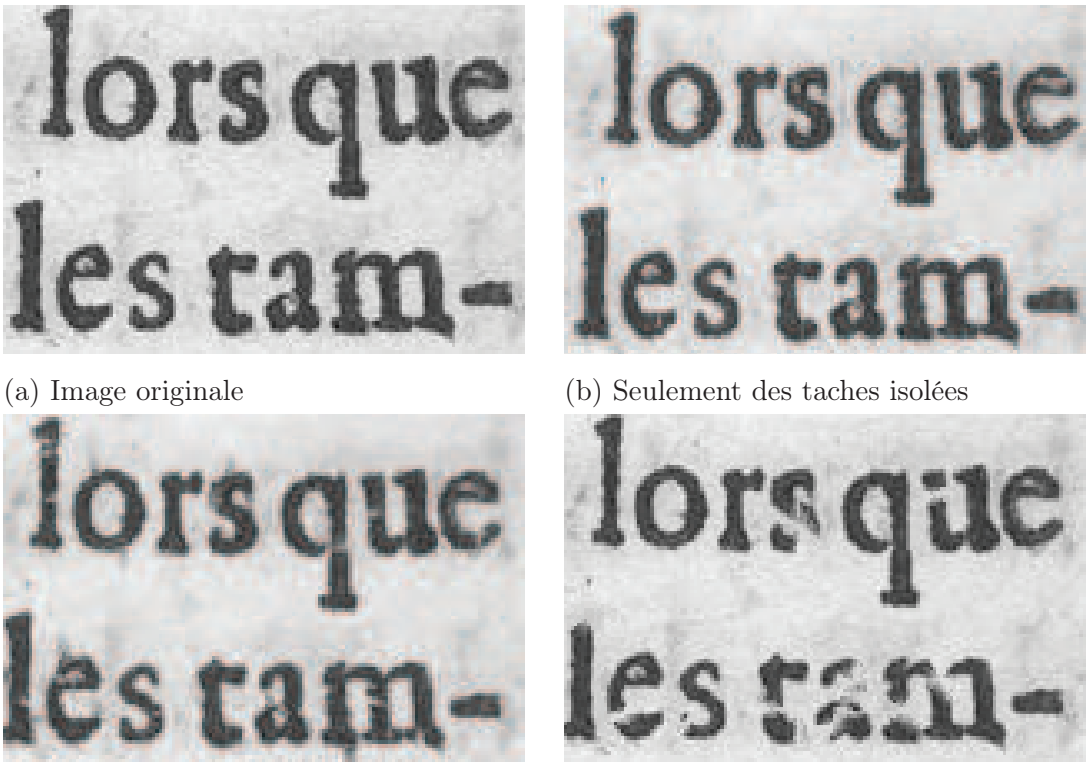
- L'évaluation qualitative [Li *et al.*, 1996] consiste, par exemple, à montrer des images synthétiques à des utilisateurs (dont il faut établir un panel) et de leur demander un retour sur ce qu'il voient (est-ce réaliste ? Font-ils la différence avec des images réelles ?) ;
- La validation automatique [Li *et al.*, 1996] consiste à comparer les performances de logiciels d'analyse ou de reconnaissance de documents (par exemple un OCR) sur des images possédant des dégradations générées synthétiquement par rapport à des images contenant uniquement de dégradations issues de documents réels ;
- La validation statistique [Kanungo *et al.*, 2000] permet de prouver de façon quantitative que la distribution d'une dégradation générée est similaire à celle observée dans la réalité ;

L'évaluation statistique proposée par [Kanungo *et al.*, 2000] a nécessité un travail complexe de constitution d'une base d'images contenant des défauts réels. Le modèle teste l'hypothèse de similarité de distribution entre un défaut réel et l'autre défaut généré au niveau du pixel. Cette validation a été possible, car les documents réels ne contenaient que le défaut reproduit synthétiquement. Dans notre contexte du document ancien, les images contiennent un grand nombre de dégradations différentes (taches claires ou sombres, taches isolées ou connectées, etc.) et très variées les unes par rapport aux autres. Il nous a donc été impossible de constituer une base importante d'images réelles contenant uniquement plusieurs exemples différents de défauts sur les caractères. C'est la raison pour laquelle nous proposons de valider notre modèle de manière visuelle (montrer des images contenant des défauts synthétiques qui ressemblent de défauts réels) et automatique (montrer que les performances d'un OCR sur des défauts synthétiques sont similaires avec ceux obtenus avec des images contenant des défauts réels). Le chapitre suivant propose tout un ensemble de tests (évaluation de performances ou ré-apprentissage) réalisés en collaboration avec des chercheurs d'autres universités et qui permettent de compléter la validation qui est présentée ci-dessous.

4.2.2.1 Évaluation qualitative

Dans cette section, nous proposons deux tests dans lesquels nous faisons varier à la fois le type et la quantité de bruit généré. L'objectif est de pouvoir visualiser les documents synthétiques générés en fonction des différentes valeurs fixées par un utilisateur.

Dans un premier test, nous générons des images qui contiennent seulement un type de bruit. Les figures 4.29-b, c, et d montrent trois images dégradées à partir de l'image originale 4.29-a. Grâce à la classification et à l'étape de génération aléatoire des niveaux de gris, la diffusion de l'encre sous la forme d'une tache sombre ou claire est réaliste. Plus précisément, la diffusion de l'encre du centre à la bordure de deux régions de bruit (cf. deux taches connectées et sombres) dans le caractère "O" de la Figure 4.29-c se concrétise par un dégradé de niveaux de gris évitant un fort contraste (peu réaliste) avec l'encre. De même, la dégradation coupant le haut du caractère "U" (cf. Figure 4.29-d) est rendue réaliste, car les nouveaux niveaux de gris de la zone ne génèrent pas de forts contrastes avec le fond de l'image.

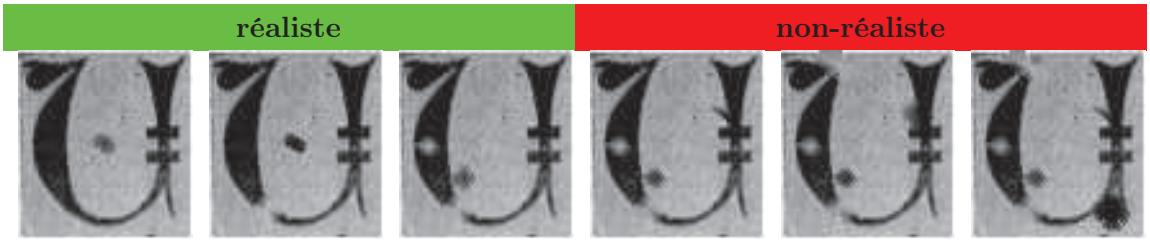


(a) Image originale (b) Seulement des taches isolées

(c) Seulement des taches connectées (d) Seulement des taches déconnexion

FIGURE 4.29: Test 1 : génération sélective des différentes dégradations que nous sommes en mesure de générer. L'image (b) contient seulement des taches isolées ; l'image (c) contient des taches connectées ; et l'image (d) ne contient que des taches de type déconnexion.

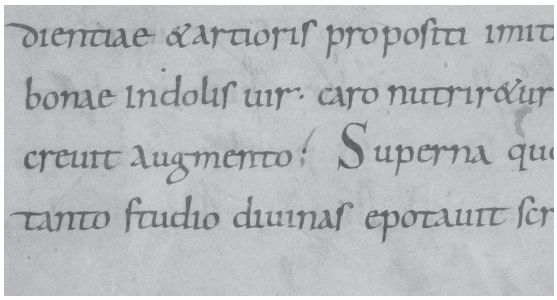
Le deuxième test consiste à générer toujours la même répartition de type de dégradation ($I = 15\%$, $0 = 65\%$, $D = 25\%$) tout en faisant augmenter le nombre total de dégradations. Les figures 4.31-b, e, f, et i montrent quatre images dégradées depuis l'image originale 4.31-a avec un nombre croissant de bruits générés. Dans la plupart des cas, le défaut généré est réaliste, même si, comme sur la figure 4.31-j, le nombre de dégradations demandées est important.



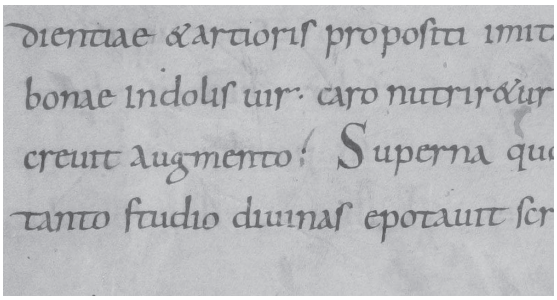
(a) nb = 1 (b) nb = 2 (c) nb = 3 (d) nb = 4 (e) nb = 5 (f) nb = 6

FIGURE 4.30: Test 2 : le nombre de points de dégradation (nb) augmente tandis que les pourcentage des trois types sont fixés. Les images dégradées sont visuellement réalistes jusqu'au nombre nb = 3 par caractère.

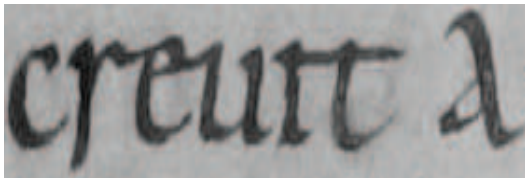
Les deux tests montrent que le modèle proposé peut générer des régions de bruit de façon réaliste. L'impact visuel de ce bruit augmente quand la quantité à générer augmente. Cette augmentation conduit à la génération croissante du nombre de zones de dégradation par mot/caractère. Par exemple, la figure 4.30 montre que les images dégradées sont visuellement réalistes jusqu'à un certain nombre de points de dégradation (dans nos expérimentations, nous avons pu observer que ce nombre de points est de l'ordre de 3). Dans la figure 4.29-d, le mot "tam" contient 9 taches de type déconnexion dont le nombre de points de dégradation est égal à 3 points par caractère, il est donc trop synthétique.



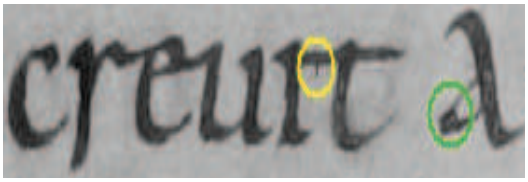
(a) Image originale



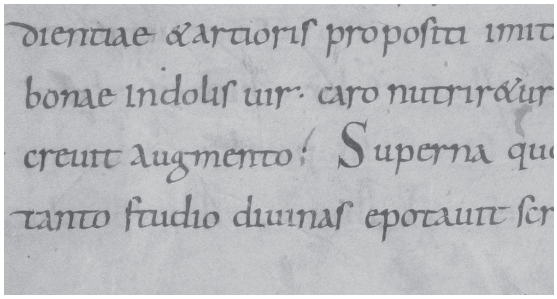
(b) Image dégradée $n_{user} = 25$



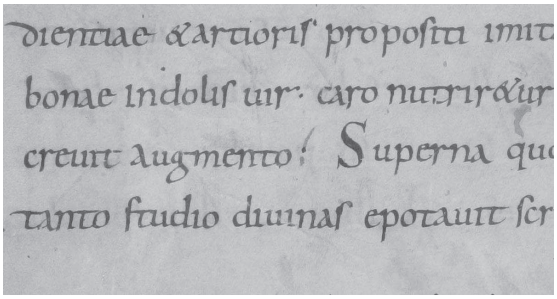
(c) Région zoomée de (a)



(d) Région zoomée de (b) $n_{user} = 25$

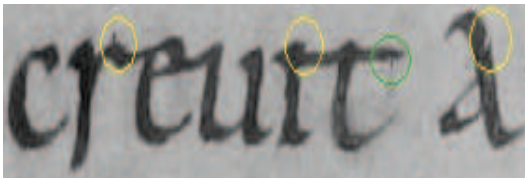


(e) Image dégradée $n_{user} = 37$

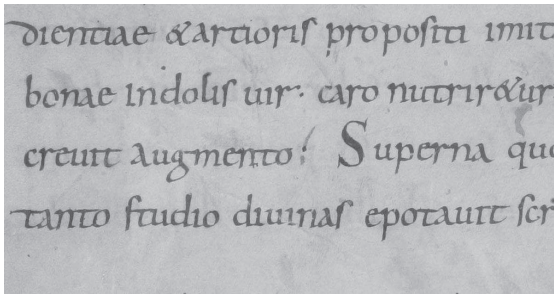


(f) Image dégradée $n_{user} = 56$

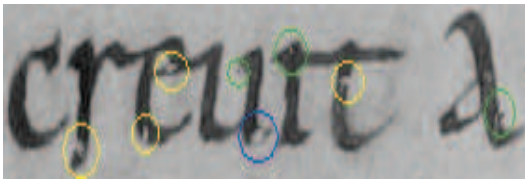
(g) Région zoomée de (e) $n_{user} = 37$



(h) Région zoomée de (f) $n_{user} = 56$



(i) Image dégradée $n_{user} = 75$

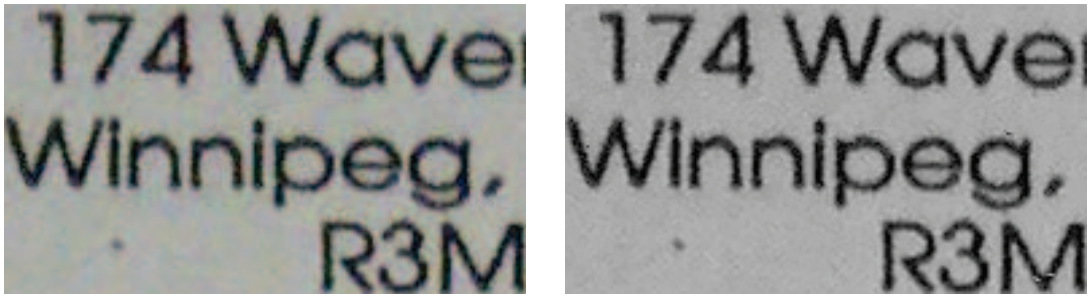


(j) Région zoomée de (i) $n_{user} = 75$

FIGURE 4.31: Test 2 : chaque image dégradée contient les trois types de bruit mais la quantité de bruit augmente progressivement.

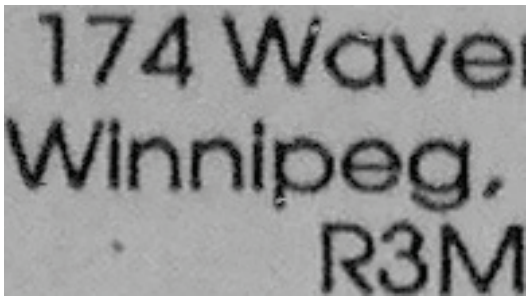
4.2.2.2 Validation automatique : l'impact de trois types de bruit sur les performances d'un logiciel d'OCR

Des caractères fins ou un caractère dont certaines parties sont effacées ou connectées à d'autres caractères induisent des baisses de performances des algorithmes d'analyse de documents (segmentation de lignes de texte, OCR, etc.). L'étude menée dans [Blando *et al.*, 1995] conforte cette hypothèse. Dans cette sous-section, nous montrons que le bruit synthétique que nous générons, de la même manière que des dégradations réelles, impacte les résultats d'un OCR. De plus, nous détaillerons l'impact de chaque type de dégradation sur les résultats de l'OCR.

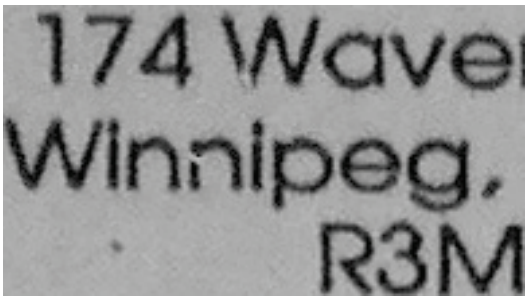


(a) Image originale

(b) Seulement des taches isolées



(c) Seulement des taches connectées



(d) Seulement des taches déconnexion

FIGURE 4.32: Exemple de trois images dégradées : l'image (b) contient seulement des taches isolées, l'image (c) contient des taches connectées, et l'image (d) ne contient que des taches de type déconnexion.

Nous utilisons une base de données nommée DIQA et qui a été publiée à ICDAR 2013 [Jayant Kumar *et al.*, 2013]. Cette base nous permet de tester deux logiciels d'OCR (OCRopus et finereader). Cette base contient 175 images numérisées à partir de 25 documents réels. Cette base est mise à disposition avec la vérité-terrain des caractères. Les images présentent certaines zones de flou de niveaux différents. Nous avons choisis 51 images com-

portant peu (ou pas) de flou afin de tester les OCR dans des conditions idéales. Les 51 images originales sont divisées en trois ensembles. Dans le premier ensemble, chaque image est dégradée avec des taches isolées selon 12 niveaux de dégradation croissants (le niveau 1 contient $1/6 \times$ le nombre de composantes connexes (nCCs) jusqu'au niveau 12 qui contient $12 \times 1/6 \times$ nCCs) (*e.g* figure 4.32-b). De la même façon, chaque image des deuxième et troisième ensembles est dégradée respectivement avec des taches connectées (cf. figure 4.32-c) et avec des taches de type déconnexion (cf. figure 4.32-d) selon 12 niveaux de dégradation croissants. La base synthétique contient donc $17 \times 3 \times 12 = 612$ images dégradées.

Le protocole de test est le suivant. Trois quart des images de la base synthétique sont utilisées pour entrainer les logiciels testés. Le reste de la base est utilisé pour les tests. La distance Levenshtein entre un mot et sa vérité-terrain est utilisée afin d'évaluer le taux d'erreur de chaque moteur.

Les figures 4.33 et 4.34 montrent le taux d'erreur moyen au niveau du mot obtenu par les deux logiciels de reconnaissance de caractères OCRopus et ABBYY finereader. Dans les deux cas, le taux d'erreur augmente avec l'état de dégradation de l'image. Le taux d'erreur d'OCRopus augmente en moyenne de 4%, et celui de Finereader augmente en moyenne de 6,4% lors que l'on passe de 0 à 2 taches par composante connexe en moyenne. Les résultats obtenus sur l'ensemble des images dégradées uniquement avec des taches de type déconnexion, montrent que c'est le cas de figure le plus complexe à gérer pour ces logiciels. Les caractères "coupés" en plusieurs composantes sont donc les plus difficiles à identifier. A l'opposé, ces deux logiciels semblent être robustes à la présence de petites taches isolées.

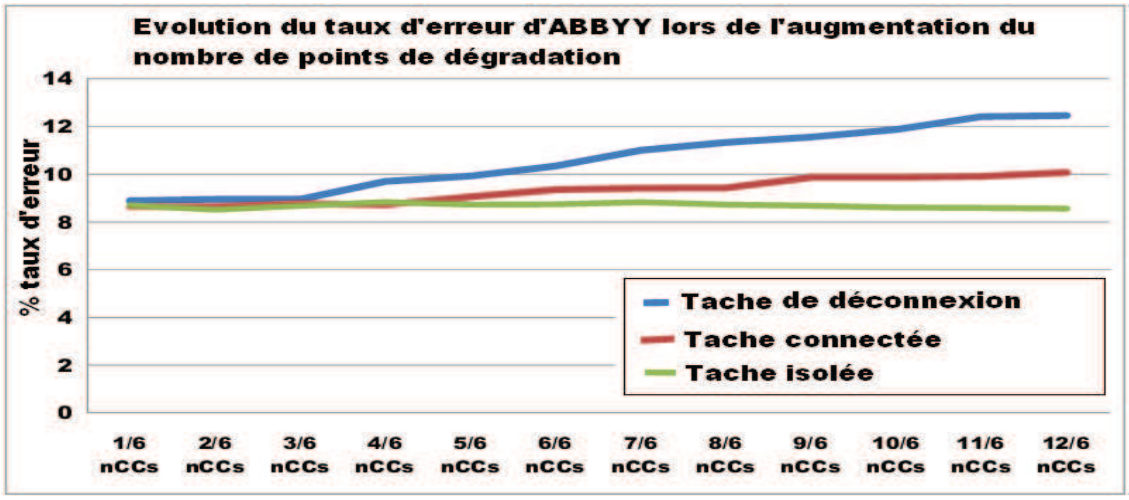


FIGURE 4.33: Le taux d'erreur réalisé par FineReader d'ABBYY selon la quantité et le type de bruit présent dans l'image de document. La quantité de points de dégradation augmente à partir de $1/6 \times$ le nombre de caractères total (nCCs) à $12/6 \times$ nCCs.

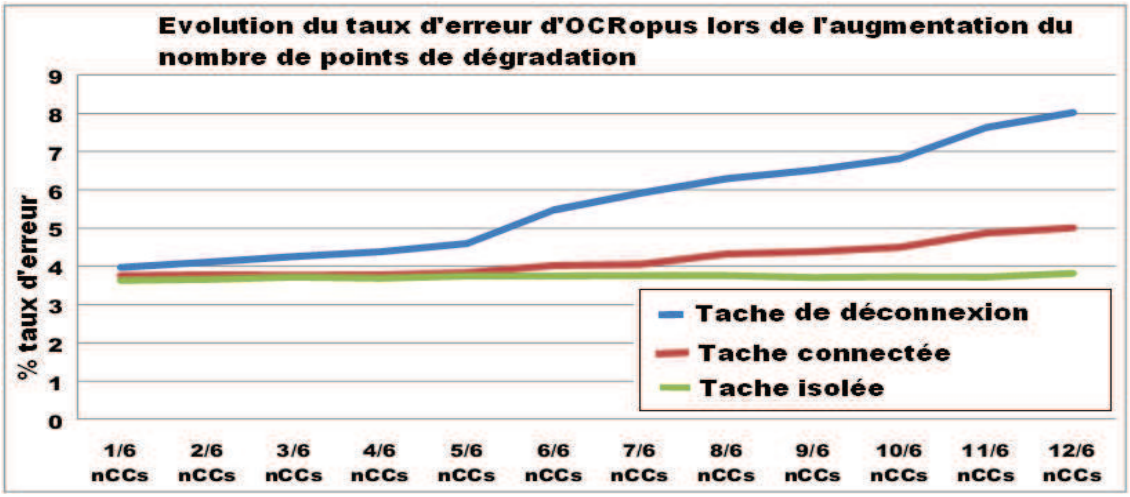


FIGURE 4.34: Le taux d'erreur réalisé par OCROpus selon la quantité et le type de bruit présent dans l'image de document. La quantité de points de dégradation augmente à partir de $1/6 \times$ le nombre de caractères total (nCCs) jusqu'à $12/6 \times$ nCCs.

4.3 Conclusion

Ce chapitre a permis de mettre en évidence que les modèles de dégradation existants dans la littérature sont performants pour imiter des bruits apparaissant dans des images

binaires et/ou modernes. Ils ne sont pas adaptés au contexte du document ancien en niveaux de gris présentant des distorsions complexes du papier comme des plis, des trous, etc. Nous avons donc proposé deux nouveaux modèles de dégradation : un permettant de reproduire, en 3D, des distorsions du papier et un second permettant de générer les dégradations en niveaux de gris liées à l'encre.

Le premier modèle permet de reproduire des distorsions globales et locales dans les documents anciens. D'abord, un document est numérisé en 3D pour obtenir un maillage. Puis, ce maillage est déplié sur un plan 2D pour calculer les coordonnées de texture des sommets du maillage. Finalement, une image 2D peut être plaquée sur le maillage pour générer l'image déformée. Ce modèle peut s'appliquer sur les images de documents anciens pour reproduire des distorsions réalistes.

Le deuxième modèle permet de générer des taches noires et blanches (à l'intérieur ou à proximité d'un caractère) et pouvant même couper la connectivité de certains d'entre eux. Ces taches apparaissent autour de la bordure des caractères. Pour les générer, un nombre souhaité de points de dégradation (les centres des taches) sont aléatoirement choisis. Puis, ces points sont classés en trois types de dégradations possibles en fonction de la distance minimale d'un point à la bordure des caractères. Finalement, à chaque point de dégradation, une région elliptique est définie.

Nous avons également réalisé une évaluation qualitative et la validation automatique de deux modèles. Ces travaux montrent que nos dégradations sont visuellement réalistes dès lors que la variation des paramètres est connue. Ils mettent également en évidence leur intérêt de réaliser une évaluation systématique de la robustesse d'algorithmes d'analyse de documents vis-à-vis de la nature et/ou du niveau de dégradation présent dans le document.

Dans le chapitre suivant, des applications réalisées en collaboration avec d'autres chercheurs permettent de confirmer la pertinence de nos modèles, et plus généralement de l'utilisation de données synthétiques, à la fois pour évaluer les performances des algorithmes de traitement et d'analyse d'images, et pour améliorer, le cas échéant, l'apprentissage de ces algorithmes.

Chapitre 5

Utilisation d'images de documents semi-synthétiques pour l'évaluation de performances ou le ré-apprentissage

Le chapitre précédent a permis de présenter notre méthodologie relative à la génération d'images de documents semi-synthétiques. Nous avons ensuite présenté dans quelles mesures la variété des défauts synthétiques joue sur les performances de méthodes d'analyse d'images de documents. Enfin nous avons présenté deux nouveaux modèles de dégradation ; l'un reproduit certaines dégradations de l'encre et l'autre simule les distorsions 3D du papier. Dans ce chapitre, nous intégrons ces deux modèles de dégradation et d'autres modèles existants dans notre générateur pour créer des bases d'images de documents semi-synthétiques. Ces bases peuvent être utilisées pour deux tâches : l'évaluation de performances et l'enrichissement de bases d'apprentissage. L'ambition de chapitre est de prouver l'efficacité de nos deux modèles de dégradations. Dans la section 5.1, nous présentons des travaux de l'état de l'art montrant la pertinence de la création d'images de documents semi-synthétiques. Ensuite, dans la section 5.2, nous détaillons plusieurs tests que nous avons réalisés avec d'autres chercheurs. Ces tests visent à évaluer finement les performances des algorithmes de traitement ou d'analyse d'images de documents de nos partenaires. Finalement nous présentons en section 5.3, deux campagnes de tests dans lesquelles nos images semi-synthétiques ont permis de mettre en évidence l'intérêt de nos modèles dans le cadre d'une méthode avec apprentissage.

5.1 Cas d'usage d'images de documents semi-synthétiques

L'analyse réalisée dans la section 3.1.2 a montré que la présence de dégradations est un facteur responsable de la baisse de performances d'algorithmes d'analyse de documents. Dans la section 2.3, nous avons montré que la génération d'images de documents semi-synthétiques est une solution possible pour compléter (voire remplacer) les actions réalisables sur des bases d'images de documents réels dans les deux tâches (évaluation de performances et ré-apprentissage). Dans cette section nous montrons, au travers de plusieurs expérimentations, que nos modèles peuvent être utiles à ces fins. Ces expérimentations permettent également de répondre à plusieurs questions relatives à l'utilisation d'images de documents semi-synthétiques :

- Évaluation de performances : l'influence d'une dégradation sur la performance est montrée dans la section 3.1.2, mais comment mesurer concrètement que ces dégradations influent sur les performances d'un algorithme ? Faut-il uniquement générer des images visuellement réalistes (des images avec de faibles dégradations) ou peut-on utiliser des images non-réalistes (des images avec de fortes

dégradations) ?

– Ré-apprentissage : les images de documents semi-synthétiques ajoutées dans la base d'apprentissage peuvent-elles améliorer les performances ? Tout comme pour l'évaluation de performances, à quel point les images peuvent elles être dégradées sans pour autant dégradées les performances de l'apprentissage.

Nous allons dans un premier temps replacer notre travail par rapport au contexte de l'état de l'art.

5.1.1 Utilisation d'images de documents semi-synthétiques pour l'évaluation de performances

L'idée d'intégrer des dégradations dans une base d'images de documents a été proposée à partir des années 90s pour évaluer les performances d'algorithmes d'analyse de documents. Les premiers modèles de dégradation proposés sont liés à l'étape de numérisation. Par exemple, des modèles de flou ou de bruits divers s'appliquant sur les images binaires ou en niveaux de gris ont été proposés dans [Jenkins et Kanai, 1994, Ho et Baird, 1995, Zhai *et al.*, 2003, Smith et Andersen, 2005, Liang *et al.*, 2008, Fornés *et al.*, 2011].

Une série d'études réalisées par Baird *et al.* concernant les dégradations induites par la numérisation est présentée dans [Baird, 1990, Baird, 1993, Ho et Baird, 1995, Baird, 2000]. Elle a pour but d'étudier l'impact des dégradations sur la performance de logiciels OCR. En 1990, Baird a proposé un modèle de dégradation globale basé sur le changement de 11 paramètres physiques d'un scanner détaillé dans la sous-section 3.2.1.1. [Ho et Baird, 1995] ont réalisé des tests d'un système OCR sur une base d'images synthétiques de documents imprimés en appliquant leur modèle de dégradation. Leurs études se focalisent sur trois types de dégradations : le flou, le bruit dû à l'impression, et le bruit de binarisation. Une base d'images contenant 6250 caractères semi-synthétiques est générée pour évaluer l'OCR. La figure 5.1 montre plusieurs exemples du caractère "e" de cette base. L'OCR est entrainé en utilisant la moitié de la base et testé avec le reste. Les résultats montrent que les performances du classifieur varient entre 83,74% pour un haut niveau de dégradation et 98,17% pour de bas niveaux de dégradation. Les auteurs concluent sur le fait que la quantité de flou est un défaut source de nombreuses erreurs d'OCR. Plus généralement, les études de Baird *et al.* ont montré tout l'intérêt de la génération d'images semi-synthétiques pour évaluer la robustesse d'algorithmes d'analyse de documents.

Elisa B. Smith analyse l'impact des dégradations dues au processus de numérisation dans [Barney Smith, 1998, Barney Smith, 2000, Smith, 2001, Hale et Barney Smith, 2007,

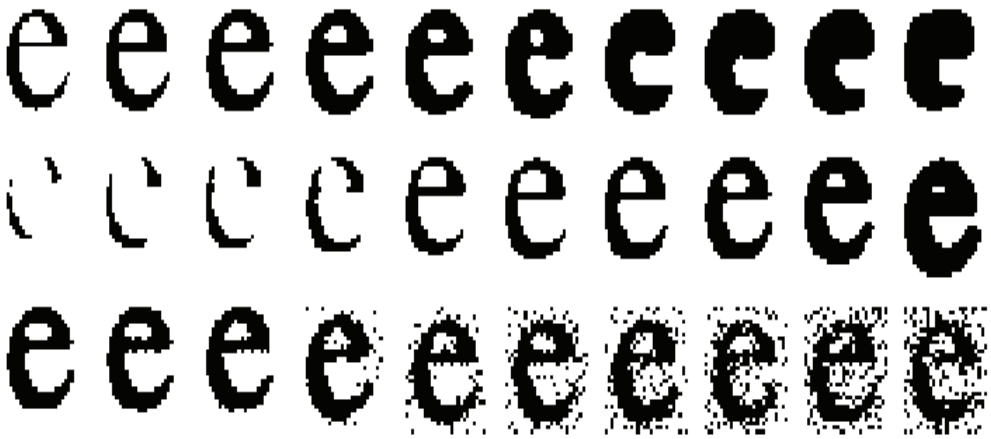


FIGURE 5.1: Exemples de trois types de dégradation issues de [Ho et Baird, 1995] : (a) caractères "e" dégradés par le flou (b) caractères "e" dégradés par le bruit d'impression, (c) caractères "e" dégradés par le bruit de binarisation

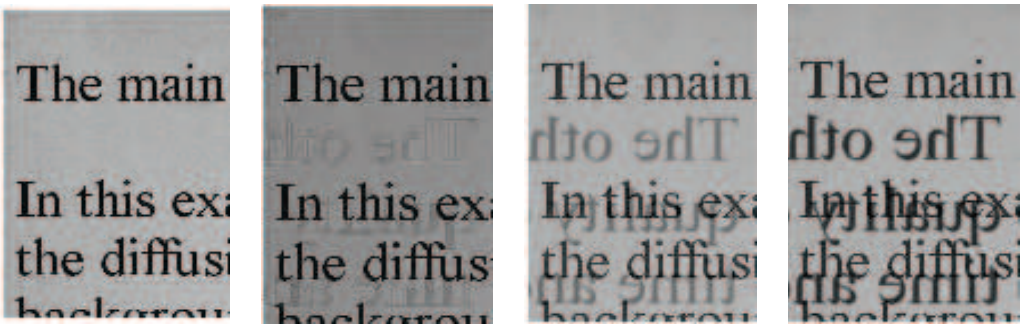
Smith, 2008, Barney Smith, 2009]. Les dégradations étudiées sont le bruit local proche du contour des objets (sorte de flou) et le bruit de binarisation. Elles sont générées par un modèle de bruit publié dans [Barney Smith, 1998]. Le seuil de binarisation et le flou de la fonction PSF (Point Spread Function) du modèle sont estimés dans l'article [Barney Smith, 2000, Hale et Barney Smith, 2007]. Une validation visuelle (comparaison entre des caractères scannés par deux matériels différents et des caractères synthétiques) et une validation qualitative (test de Turing avec 93 participants) sont présentées dans ces articles. Une étude statistique comparant la distribution des niveaux de gris entre dégradations réelles et synthétiques est présentée dans [Smith et Qiu, 2003]. Cela permet aux auteurs de trier les images de la base de test pour créer plusieurs sous-bases dans lesquelles les images synthétiques sont plus ou moins similaires à celles de la base d'apprentissage. L'objectif étant d'utiliser les meilleures images synthétiques pour améliorer les performances de l'étape de classification. Plus précisément, les auteurs de [Smith et Andersen, 2005] suivent le protocole suivant : quatre classifieurs sont entrainés sur une des quatre bases d'apprentissage. Chaque classifieur est testé sur une base dont les éléments sont choisis aléatoirement. Les performances obtenues sont de 96,2% en moyenne. Puis, chaque classifieur est testé sur une base qui est triée en quatre sous-bases selon leur niveau de similarité. Les performances passent, en moyenne, à 97,6%. Ces résultats montrent qu'un classifeur peut s'adapter mieux à une base ciblée s'il est guidé par l'information utile extraite d'images semi-synthétiques.

Le travail de [Zhai *et al.*, 2003] tente d'évaluer les performances de trois algorithmes

de détection de symboles [Wenyin et Dori, 1994, Elliman, 2002, Libenzi,] relativement à la présence de 4 bruits : le bruit Gaussien (poivre et sel), le bruit dit "de haute fréquence" (similaire au bruit de [Kanungo *et al.*, 1993]), le flou dû au mouvement des documents lors de la numérisation, et le bruit "hard pencil noise" présenté dans la sous-section 3.2.2.3. La base originale de 18 images référencée dans [Wenyin *et al.*, 2002] est dégradée en dix niveaux (0 : sans dégradation, 10 : très dégradée) pour obtenir une base de test contenant 18×10 niveaux $= 180$ images semi-synthétiques. Les performances des trois algorithmes chutent quand le niveau de dégradation augmente. Les performances ont chuté de 70% en moyenne quand le niveau de dégradation passe de 0 à 10. C'est une évidence que plus le niveau de bruit augmente, plus la performance chute. Néanmoins, l'intérêt de cette étude est qu'elle montre la pertinence de l'usage d'images semi-synthétiques pour l'évaluation de performances et même plus précisément de l'intérêt qu'il y a à moduler le niveau de bruit des images générées.

Deux études de [Liang *et al.*, 2008] et [Fornés *et al.*, 2011] montrent que les distorsions ont un impact sur la performance d'algorithmes de détection et de restauration de documents et sur les performances d'un logiciel d'OCR. Dans leur première étude, [Liang *et al.*, 2008] ont utilisé leur modèle de distorsion résumé dans la sous-section 3.2.1.3 afin de tester leur méthode de restauration de distorsion. Une base de 120 images semi-synthétiques est générée à partir de 5 images de documents réelles. Les tests réalisés avec l'OCR OmiPage montrent que les distorsions diminuent les performances et que grâce à leur méthode de restauration la performance de l'OCR est améliorée. Dans la seconde étude, [Fornés *et al.*, 2011] génèrent 6000 images semi-synthétiques d'apprentissage et 6000 images semi-synthétiques de test à partir de 1000 images de documents musicaux. Ces images contiennent des distorsions en 2D (rotation, translation) et des bruits (les taches blanches et le bruit de [Kanungo *et al.*, 1993]) pour évaluer les méthodes de suppression de lignes (compétition ICDAR 2011). Ces études mettent en évidence l'impact des dégradations. Ce qu'il faut principalement retenir de ces expérimentations, c'est le protocole de génération d'images utilisé et les choix faits en termes de quantité d'images générées.

Deux méthodes de restauration de transparence ([Moghaddam et Cheriet, 2009] et [Oja et Yuan, 2006]) sont évaluées et comparées en utilisant les images semi-synthétiques générées par le modèle de transparence de Moghaddam [Moghaddam et Cheriet, 2009]. Ces deux modèles sont analysés en modifiant les paramètres gérant la diffusion de l'encre. Une base d'images est générée à partir de deux images réelles ne contenant pas de transparence. La quantité d'encre diffusée augmente à chaque itération. La figure 5.2.b-d montre trois



(a) (b) (c) (d)

FIGURE 5.2: Exemples de la base de test utilisée par [Moghaddam et Cheriet, 2009] : (a) image originale, (b) image peu dégradée, (c) image moyennement dégradée, (d) image fortement dégradée

images dégradées à partir dune image originale 5.2.a. Les résultats présentés dans [Moghaddam et Cheriet, 2009], permettent de montrer que les deux méthodes testées réalisent une restauration visuellement acceptable pour des dégradations synthétiques faibles. La méthode de restauration [Moghaddam et Cheriet, 2009] donne de meilleurs résultats quand les dégradations sont plus fortes. Ces expérimentations mettent donc en évidence que la génération d'images semi-synthétiques binaires peut avoir son utilité.

Les tests présentés préalablement montrent tous, évidemment, que plus le niveau de dégradation synthétique augmente, plus les performances diminuent. Ce qui est plus intéressant, c'est de faire le lien entre le niveau de dégradations et la performance comme dans [Baird, 1990, Barney Smith, 1998, Zhai *et al.*, 2003, Liang *et al.*, 2008, Fornés *et al.*, 2011], et que ceci permet de connaitre l'influence de chaque défaut sur les performances. Il est à noter que seule l'étude réalisée par [Moghaddam et Cheriet, 2009] porte sur des documents anciens en niveaux de gris.

5.1.2 Utilisation d'images de documents semi-synthétiques pour le ré-apprentissage

Dans cette section, nous étudions des expérimentations dans lesquelles des images de document semi-synthétiques ont été utilisées pour enrichir une base d'apprentissage.

Au travers de ces études nous allons essayer de savoir comment les auteurs ont appréhendé les questions suivantes : (1) l'utilisation de données semi-synthétiques en complément de données réelles peut-elle modifier la qualité de l'apprentissage ? (2) l'introduction

de bruit dans les images semi-synthétiques améliore-t-elle systématiquement la qualité de l'apprentissage ou peut-elle la faire chuter ? (3) Comment fixer le bon niveau de dégradation ? (4) Est-ce que la combinaison de dégradations (combiner les modèles) peut améliorer la reconnaissance ? (5) Combien faut-il intégrer d'images de documents semi-synthétiques dans la base d'apprentissage ? (6) A quel point les images synthétiques doivent elles être similaires aux images réelles ?

Les études menées par [Baird, 2000] permettent de répondre partiellement aux questions précédemment énoncées. Une conclusion intéressante est que les tests qu'ils ont réalisés avec des documents synthétiques volontairement trop dégradés ont détérioré les résultats. Ils concluent également leurs tests sur le fait que les images synthétiques doivent obligatoirement être issues de la base d'apprentissage (réelle) originale. Si les images synthétiques sont générées à partir de documents (même relativement similaires) n'appartenant pas à la base d'origine, des informations incohérentes sont artificiellement intégrées lors de l'apprentissage. Les résultats peuvent alors devenir incohérents. Cette conclusion semble être confirmée par les tests réalisés dans [Mori *et al.*, 2000, Varga et Bunke, 2003a].

[Mori *et al.*, 2000] alimentent leur base d'apprentissage avec uniquement des images semi-synthétiques afin d'entraîner un moteur de reconnaissance de chiffres manuscrits. Un générateur basé sur la forme normalisée des chiffres permet de générer des images de chiffres à partir d'une base originale. Le modèle d'un chiffre est généré à partir de plusieurs exemples d'un même chiffre (cf. figure 5.3-a-b). Les pixels du squelette du modèle sont mis en correspondance avec ceux d'un chiffre (cf. la figure 5.3-c). Des chiffres semi-synthétiques sont créés en faisant varier les liens entre deux pixels (cf figure 5.3-c). Un paramètre permet de contrôler la variation des liens. La figure 5.3-d montre plusieurs exemples d'images semi-synthétiques générées selon des valeurs différentes de paramètres. Trois bases de 200, 500, et 1000 images sont créées (50% d'images semi-synthétiques et 50% images réelles).

La figure 5.4 présente le taux de reconnaissance global en fonction du niveau de dégradation des images de la base. Le paramètre de distorsion fixé entre $-0,2$ et $0,5$ correspond à des images contenant des caractères faiblement dégradées. Ces images synthétiques, utilisées lors de l'apprentissage, permettent clairement d'améliorer le taux de reconnaissance de 0,3% en moyenne. Au contraire, des images fortement dégradées (paramètre de distorsion $<$-0,2 ou $>$0,5), sont responsables d'un taux de reconnaissance plus faible. Ces tests permettent de répondre aux trois premières de nos interrogations mentionnées dans les section 5.1.2 : l'utilisation de données semi-synthétiques modifie les résultats d'apprentissage sans avoir à enlever des images composant la base réelle. Le niveau de dégradation d'images semi-

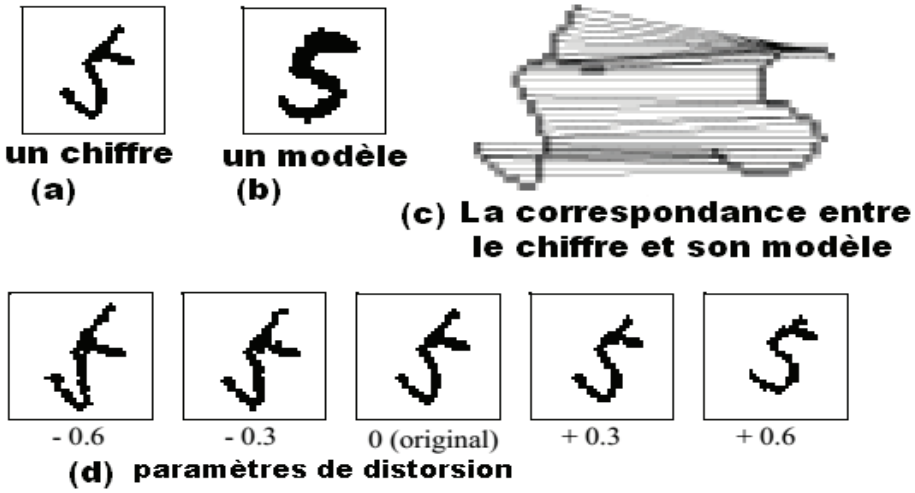


FIGURE 5.3: Le processus de génération d'images de chiffres utilisant des points de correspondance entre un chiffre et un modèle [Mori *et al.*, 2000]

synthétiques a un impact important sur la qualité de la base d'apprentissage. Concrètement, un faible niveau de dégradations permet d'améliorer le taux de reconnaissance tandis qu'un haut niveau le diminue.

La figure 5.5 permet, quant à elle, de montrer que l'ajout d'images semi-synthétiques améliore fortement l'apprentissage, et donc les résultats de l'étape de reconnaissance. Pour cela, les auteurs comparent le taux de reconnaissance obtenu avec une base de plus en plus conséquente mêlant images réelles et synthétiques avec les résultats obtenus en utilisant une base comportant uniquement des images réelles. L'augmentation de taille de la base d'apprentissage permet d'améliorer la performance du moteur de reconnaissance dans les deux cas (voir la figure 5.5). Ce qui est intéressant ici c'est que pour un même nombre d'images dans la base d'apprentissage, les résultats sont meilleurs en mélangeant images semi-synthétiques et réelles plutôt qu'en utilisant uniquement des images réelles. La base "mixte" est pertinente jusqu'à un nombre d'images proche de 700. Par conséquent, on peut imaginer que sur une "petite" quantité d'images, il est plus simple d'avoir une base variée en la générant plutôt qu'en choisissant des images réelles.

Le modèle de distorsion 2D présenté dans [Varga et Bunke, 2003b] permet de déformer des lignes de texte. Il est appliqué pour générer des images de lignes de texte semi-synthétiques qui sont ensuite intégrées dans des bases d'apprentissage [Varga et Bunke, 2003a]. La base originale est présentée dans [Marti et Bunke, 2002]. Les images semi-synthétiques sont dégradées selon quatre niveaux : très bas niveau, bas niveau, moyen ni-

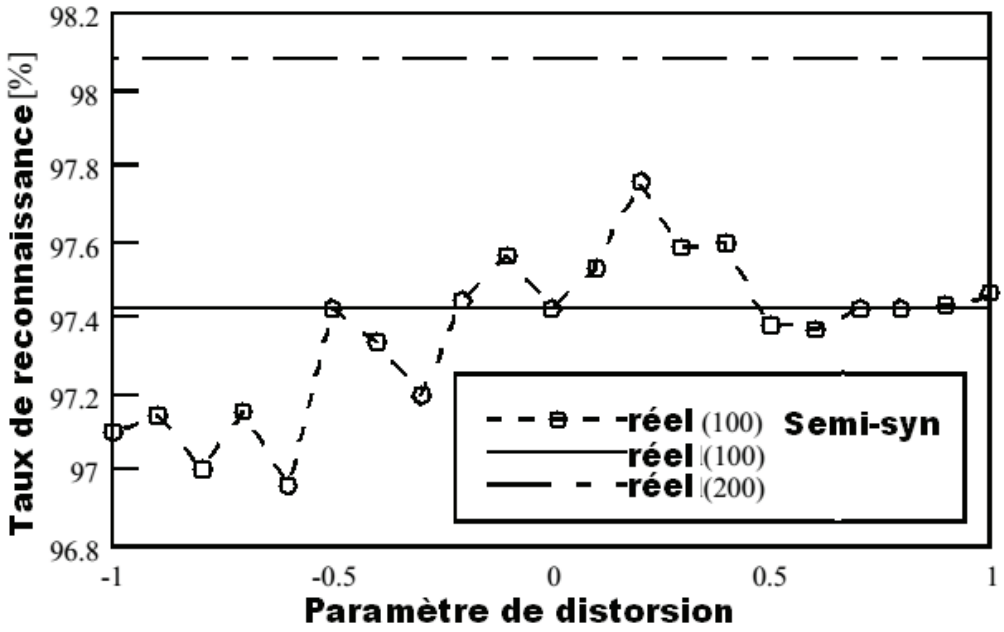


FIGURE 5.4: Taux de reconnaissance de chiffres manuscrits par rapport à l'évolution du niveau de distorsion (dans le cas où la base d'apprentissage contient 100 images) [Mori *et al.*, 2000]

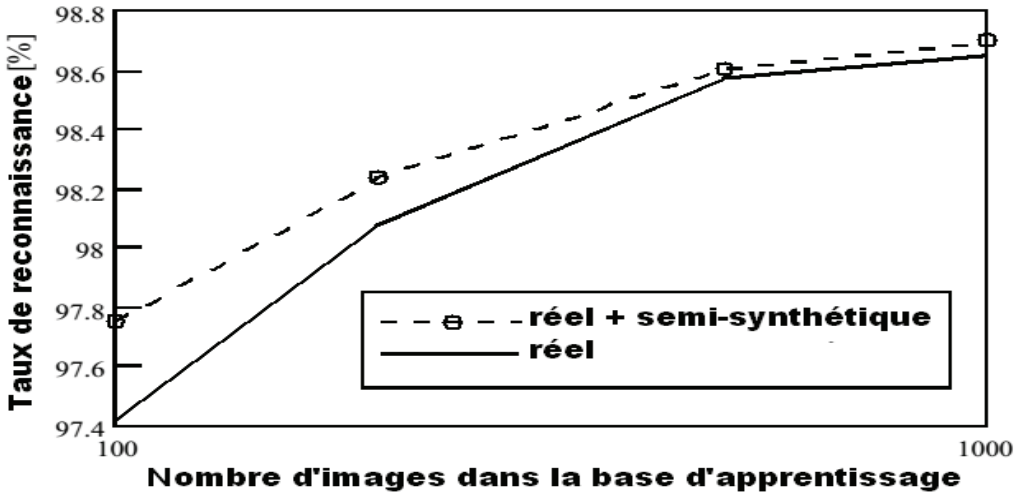


FIGURE 5.5: Taux de reconnaissance de chiffres en fonction de l'évolution de la taille de la base d'apprentissage [Mori *et al.*, 2000]

veau, haut niveau. Quatre bases d'apprentissage de 81, 162, 243, et 324 images sont générées. Dans chaque base d'apprentissage, il y a 5 images semi-synthétiques générées pour une image réelle. La base de test contient 47 images réelles de lignes de texte.

TABLE 5.1: Taux de reconnaissance en fonction de la taille de la base et du niveau de dégradation des images générées. [Varga et Bunke, 2003a]

	Origine	Très bas	Bas	Moyen	Haut
taille=81	54,20	62,81	**65,76**	63,72	62,81
taille=162	63,04	70,29	**72,79**	70,29	67,80
taille=243	70,07	71,20	**72,79**	71,88	67,80
taille=324	71,20	73,47	**74,60**	70,52	70,07

Les résultats présentés dans la table 5.1 permettent de conclure que plus la taille de la base d'apprentissage devient grande, plus la performance du moteur augmente. Ces tests permettent également de confirmer ceux réalisés par [Mori *et al.*, 2000] concluant que des dégradations trop fortes ont tendance à faire baisser les performances. Il faut donc générer des images représentatives de la variété des dégradations que l'on peut observer sur des images réelles, sans trop accentuer ces dégradations.

5.2 Contribution 3 : utilisation d'images de documents semi-synthétiques pour l'évaluation de performances

Afin de confirmer les conclusions issues des expérimentations de la section précédente, mais également pour répondre aux interrogations (4) et (5) (Est-ce que la combinaison de dégradations dans des images semi-synthétiques peut améliorer la reconnaissance ? Combien faut-il intégrer d'images de documents semi-synthétiques dans la base d'apprentissage ?), nous avons réalisé un ensembles de tests. L'ensemble des tests présentés par la suite ont été réalisés en collaboration avec des chercheurs issus de différents laboratoires dont les activités portent sur l'analyse d'images de documents. Concrètement cela a consisté pour nous, à nous approprier les bases d'images réelles des chercheurs avec lesquels nous avons collaboré et de réfléchir avec eux sur les points suivants : est-ce que les dégradations proposées sont pertinentes pour évaluer les performances des algorithmes d'analyse d'images de documents anciens ? Comment génère-t-on la base de test avec les modèles de dégradation que nous avons proposés (et ceux existants) afin de s'adapter à différents contextes d'évalua-

tion ? Nous considérons que cette section est un apport important de ces travaux de thèse puisque ces tests permettent non seulement de prouver la qualité et l'utilité de nos modèles de dégradation, mais également de répondre aux questions relatives à l'utilisation de telles données. Nous allons tout d'abord détailler deux expérimentations réalisées respectivement avec le laboratoire CVC de l'Université Autonome de Barcelone [1] et le laboratoire L3I de l'Université de la Rochelle.

5.2.1 Génération d'images de documents semi-synthétiques pour la compétition ICDAR 2013

Les documents musicaux sont une catégorie d'images pour lesquelles des nombreuses méthodes d'analyse ont été proposées [Blostein et Baird, 1992,Rebelo *et al.*, 2012,Dalitz *et al.*, 2008,dos Santos Cardoso *et al.*, 2009]. Généralement, un document musical contient des lignes, des symboles musicaux, des textes et des illustrations. La figure 5.6 montre un exemple d'un document musical. Dans les mécanismes d'analyse, supprimer les lignes de portée est une étape fréquemment utilisée avant de pouvoir tenter la reconnaissance des symboles musicaux. Comme pour tout type d'image de document, la présence de dégradations et la variabilité du contenu rendent cette analyse complexe. Par ailleurs, il existe peu d'images disposant de vérité-terrain, permettant de qualifier ces algorithmes [Rebelo *et al.*, 2012].

[Fornés *et al.*, 2011] ont organisé un concours pour évaluer des algorithmes de suppression de lignes de portée de documents musicaux. Ce concours utilise la base CVC-MUSICMA [Fornés *et al.*, 2012] qui contient 1000 images de documents musicaux produits par 50 scripteurs. Une partie des images sont dégradées en utilisant des modèles de bruit : bruit de [Kanungo *et al.*, 1993], taches blanches, déconnection et changement d'épaisseur de lignes. Ces dégradations sont générées selon différents niveaux (bas, moyen et haut). Des modèles de distorsions 2D (rotation, courbure, translation de blocs, translation plus ou moins forte selon l'axe Oy) sont également utilisés pour générer des images semi-synthétiques (cf. la figure 5.7). Une base de test composée de 500 $\times$ 11 dégradations + 500 images originales donne finalement 6000 images semi-synthétiques. Ces images ont été utilisées pour évaluer 7 algorithmes de suppression de lignes de portée. Ces algorithmes sont évalués au travers de deux tests. Dans le premier test, chaque image semi-synthétique contient une seule dégradation. Les algorithmes produisent de bons résultats (taux d'erreur entre 1,9% et 2,8%). Dans le deuxième test, les images contiennent des combinaisons de dégradations.

1. http://www.cvc.uab.es/

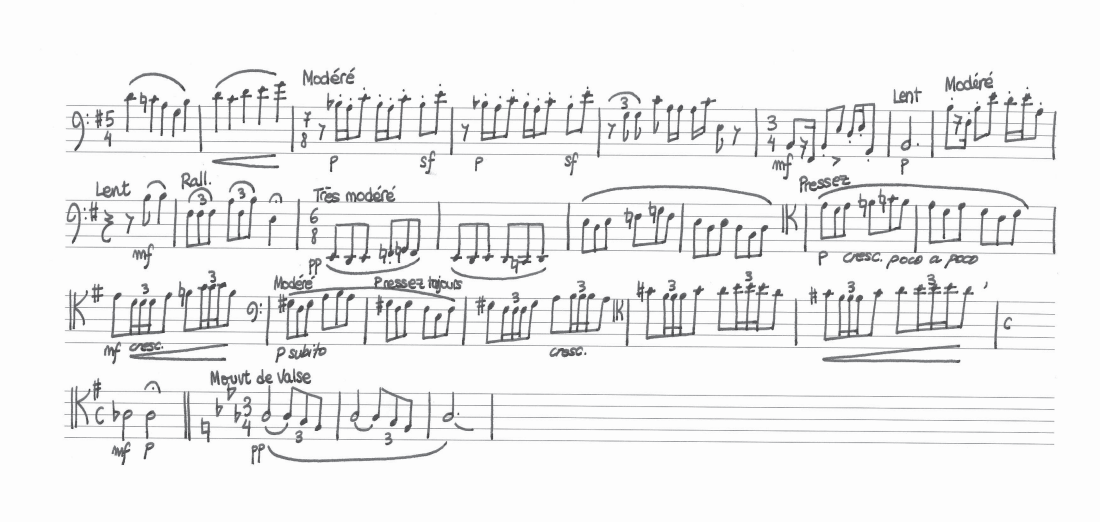


FIGURE 5.6: Exemple d'une image de document musical issue de la base CVC-MUSICMA [Fornés *et al.*, 2012]

Les performances sont moins bonnes que dans le premier test.

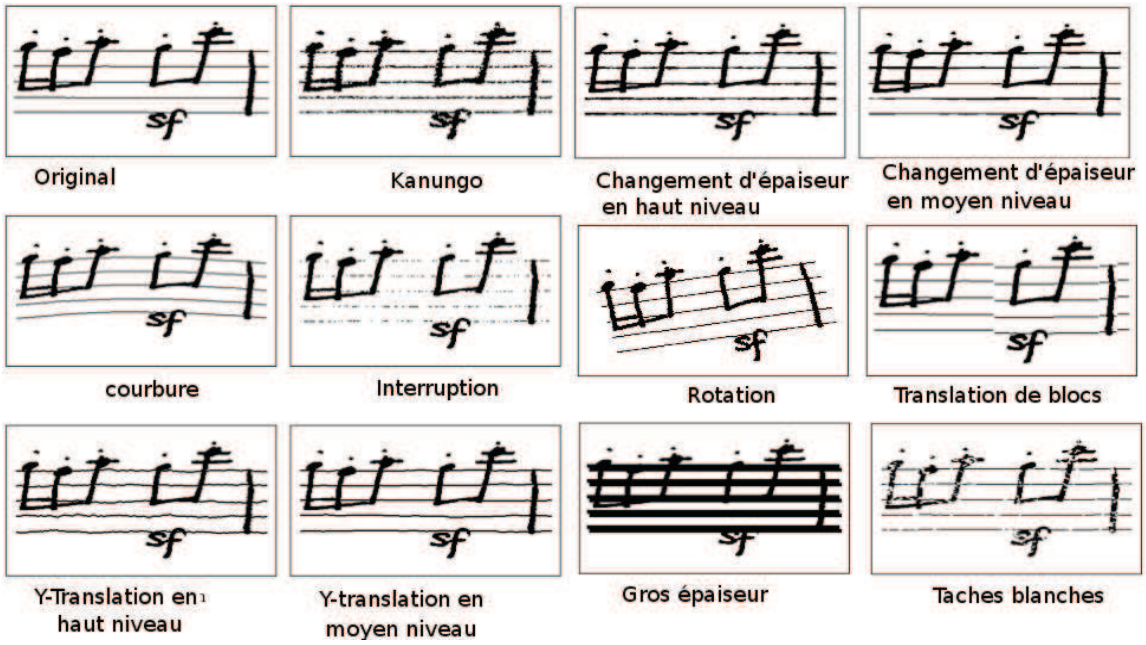


FIGURE 5.7: Défauts apparaissant dans des documents musicaux.

Néanmoins, on constate que dans le cadre de la génération d'images semi-synthétiques, les images ne sont pas très réalistes. En effet, la version 2011 de cette base contient des images

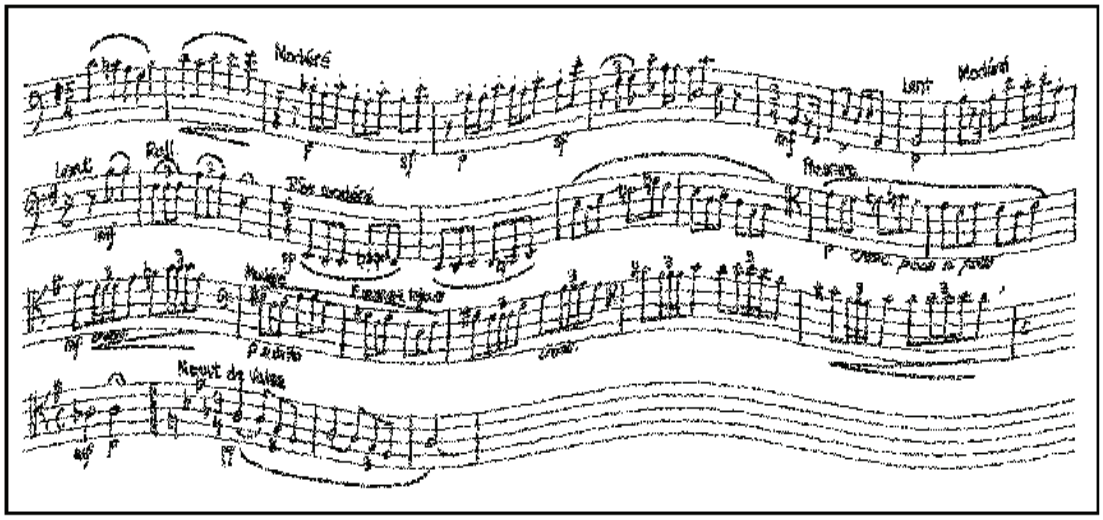


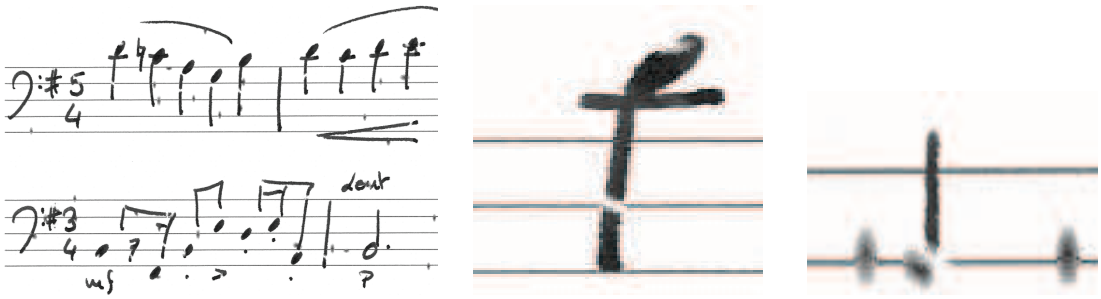
FIGURE 5.8: Exemple d'une image de document musical semi-synthétique contenant plusieurs type de dégradations pour la compétition [Fornés *et al.*, 2011]

dont les dégradations semi-synthétiques ne sont pas assez réalistes ni suffisamment variées. Les distorsions 2D sont toutes de la même forme et le bruit ajouté sur les lignes de portée et les notes sont très similaires les uns par rapport aux autres. Elle sont similaires à un simple bruit poivre et sel (cf. figure 5.8). Enfin, ces dégradations sont des dégradations binaires peu représentatives de celles observées sur des documents musicaux anciens réels en niveaux de gris. En améliorant l'étape de génération des images semi-synthétiques, cette base (générée en 2011) peut potentiellement être rendue plus intéressante.

Notre collaboration avec [Fornés *et al.*, 2011] a donc dans un premier temps consisté à générer une nouvelle base, plus réaliste et disposant d'images comportant des bruits plus variés. Cette base est maintenant plus représentative de bases réelles de documents musicaux anciens. Nous avons pour cela utilisé nos deux modèles (bruit local et distorsion en 3D) sur des images originales du concours de 2011 pour générer des images semi-synthétiques. Nous avons ainsi co-organisé la deuxième édition de cette compétition. Elle s'est tenue lors de la conférence ICDAR 2013. Notre base a permis en particulier d'ouvrir ce concours aux chercheurs disposant de méthodes utilisant des images en niveaux de gris.

Les images de cette base sont des images musicales anciennes en niveaux de gris. Nos modèles sont donc parfaitement adaptés au contexte du concours. Les taches de type déconnexion permettent de couper les lignes dans les documents musicaux (cf. la figure 5.9-b). Notre modèle permet également de créer des taches sombres pouvant amener les

algorithmes à les confondre avec des symboles musicaux comme dans la figure 5.9-c. Les distorsions 3D ont tendance à plier les lignes des images de façon beaucoup diversifiée que dans version 2011 de la base, comme montré figure 5.10.



(a) image semi-synthétique (b) déconnection d'une ligne (c) bruit pouvant être assimilé à une note

FIGURE 5.9: Une image de document musical semi-synthétique contenant des bruits locaux

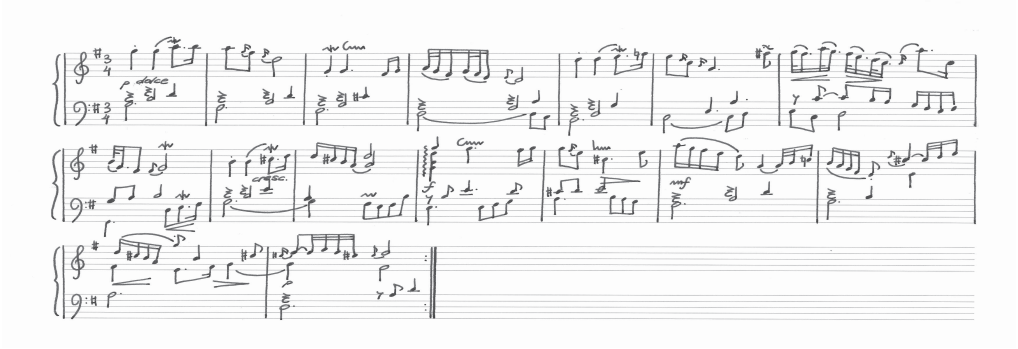
Avec cette compétition, nous souhaitons pouvoir obtenir des informations sur l'influence de chaque type et degré de défaut sur les performances des algorithmes testés.

5.2.1.1 Base d'images de documents musicaux semi-synthétiques

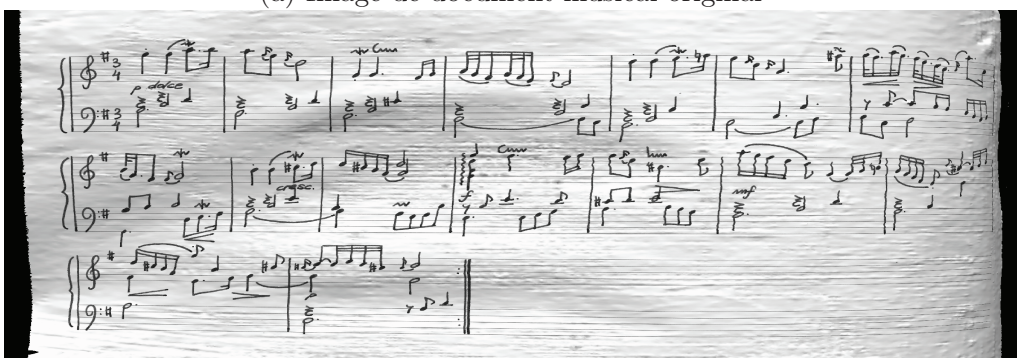
Nous sommes donc partis de la version 2011 de la base présentée dans [Fornés *et al.*, 2012]. Nous avons utilisé les 1000 images réelles de cette base et les avons divisées en deux groupes : le premier contient 667 images qui nous serviront à générer une base sur laquelle les participants du concours pourront entraîner leur algorithmes avant la compétition, le second groupe contient 333 images qui seront dégradées pour générer la base évaluant les algorithmes des participants lors du concours. Ces 1000 images réelles seront donc dégradées puis ne seront plus jamais utilisées dans la suite du processus. Le concours s'opère uniquement sur des images semi-synthétiques. Ainsi, le premier groupe d'images est dégradé par nos deux modèles afin de générer une base d'apprentissage de 4000 images semi-synthétiques. Le second groupe est dégradé de la même manière, mais en changeant légèrement le jeu de paramètres utilisés pour générer le premier groupe. Nous obtenons donc une base de test de 2000 images semi-synthétiques. Les paragraphes suivants expliquent en détail le processus de génération de ces deux bases.

Le premier groupe d'images (bases d'apprentissage) contient donc 4000 images semi-synthétiques générées à partir de 667 images originales. Ces images semi-synthétiques sont de nouveau séparées en trois sous-ensembles suivant le modèle de dégradation choisi :

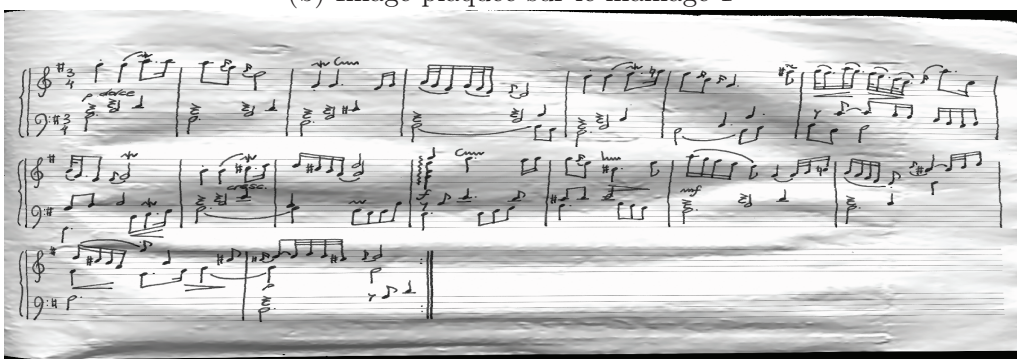
– Ensemble 1 : composé de 1000 images semi-synthétiques qui sont dégradées par notre modèle de distorsion en 3D utilisant deux maillages. Le premier maillage se caractérise par la présence de petites déformations locales tandis que le deuxième maillage contient un nombre plus important de petits plis. Nous plaquons 667 images originales sur deux maillages pour produire 667×2 = 1334 images semi-synthétiques. 1000 images semi-synthétiques sont sélectionnées au hasard parmi ces 1334 images. La figure 5.10-b est une image générée avec le premier maillage et la figure 5.10-c est une image générée à partir du second maillage.



(a) Image de document musical original



(b) Image plaquée sur le maillage 1

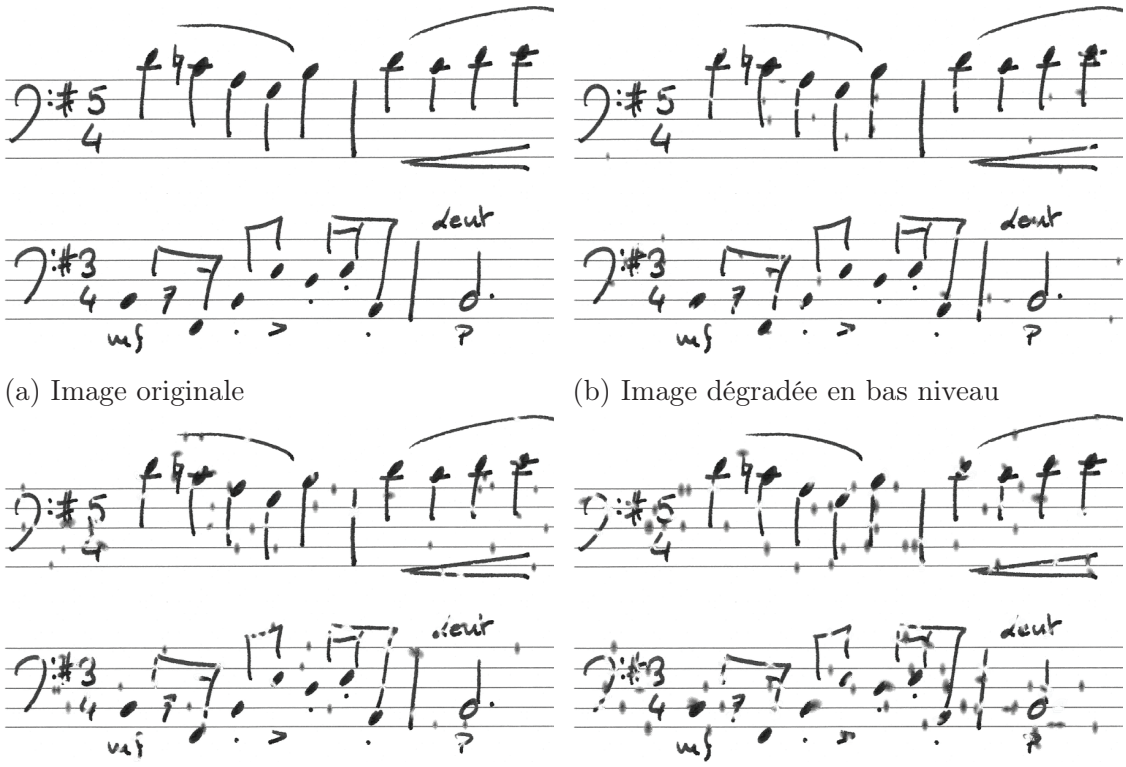


(b) Image plaquée sur le maillage 2

FIGURE 5.10: Images semi-synthétiques issues du groupe 1 : (a) image originale, (b) image plaquée sur le maillage 1, (c) image plaquée sur le maillage 2

– Ensemble 2 : composé de 1000 images semi-synthétiques générées par notre modèle de bruit local selon trois niveaux : bas, moyen, haut. Le facteur d'aplatissement g est fixé à $0,6$ tandis que le nombre de points de dégradation et la taille de régions de bruit a_0 augmentent pour chaque niveau. Ces images sont divisées en trois sous-ensembles correspondants aux trois niveaux de bruit :

- Bas niveau : 333 images semi-synthétiques. Chacune contient 500 points de dégradation avec une taille de régions de bruit fixée $a_0 = 7$. Un exemple est présenté dans la figure 5.11-b ;
- Moyen niveau : 334 images semi-synthétiques. Chacune contient 1000 points de dégradation avec la taille de régions de bruit fixée à $a_0 = 8.5$. Un exemple est présenté dans la figure 5.11-c ;
- Haut niveau : 333 images semi-synthétiques. Chacune contient 1300 points de dégradation avec une taille de régions de bruit fixée à $a_0 = 10$. Un exemple est présenté dans la figure 5.11-d ;

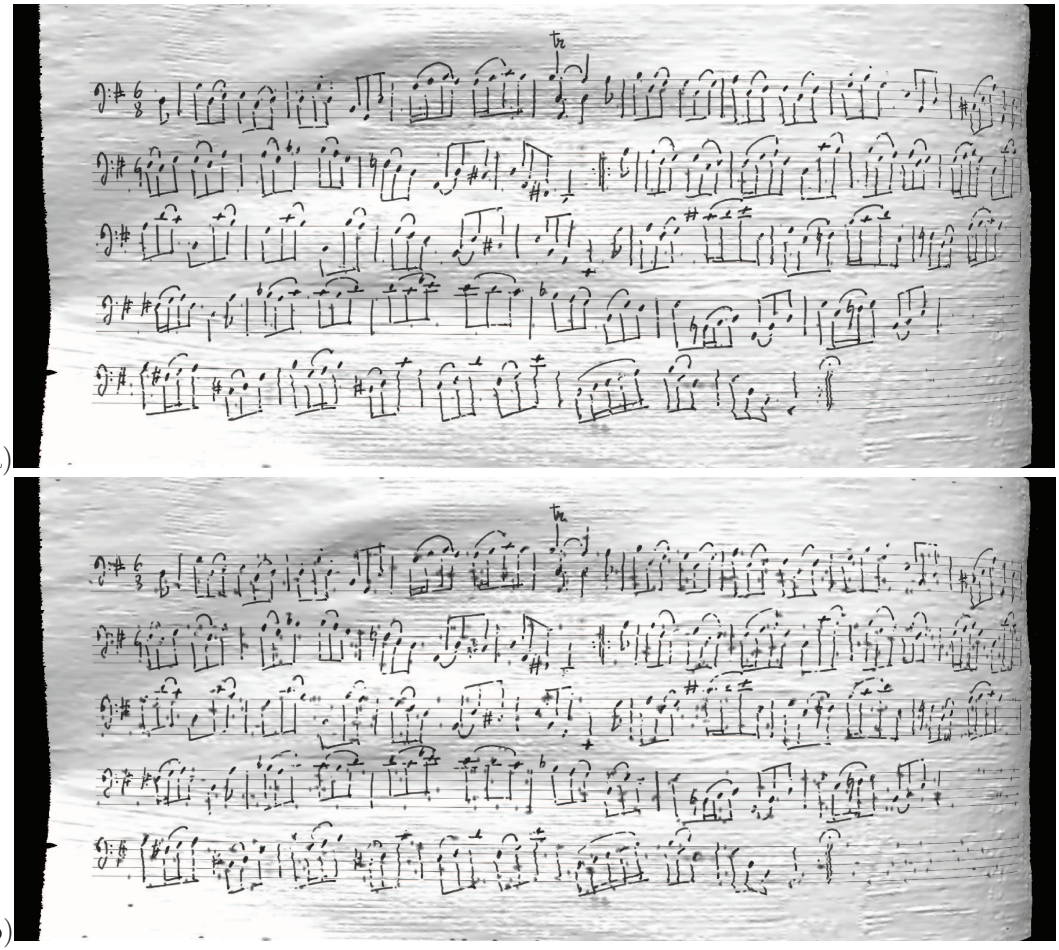


(a) Image originale (b) Image dégradée en bas niveau

(c)Image dégradée en niveau moyen (d) Image dégradée en haut niveau

FIGURE 5.11: Images semi-synthétiques issues de l'ensemble 2 : (a) image originale, (b) image peu dégradée, (c) image moyennement dégradée, (d) image fortement dégradée

- Ensemble 3 : composé de 2000 images générées en combinant nos deux modèles de dégradation. On obtient donc (1000 images bruitées selon trois niveaux de dégradation × 2 maillages) = 2000 images semi-synthétiques. Les figures 5.12 et 5.13 présentent 4 exemples issus de cet ensemble.

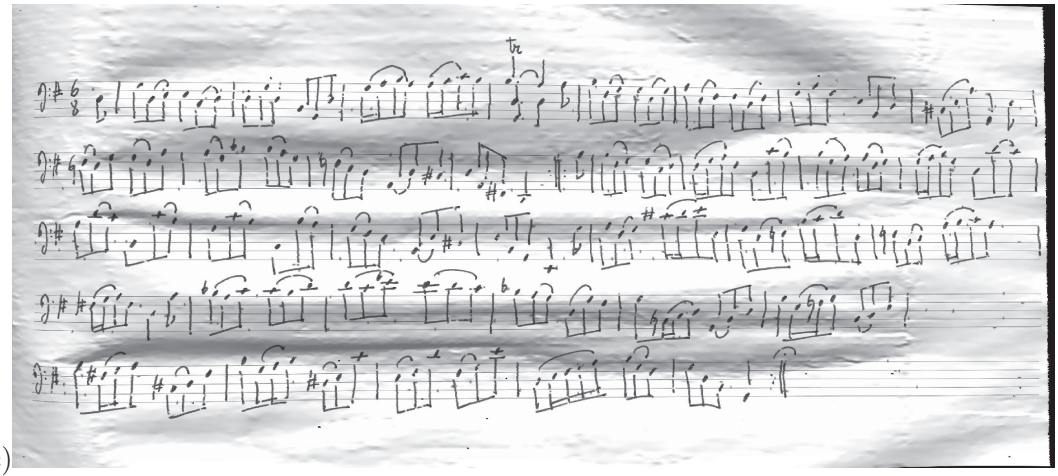


(a)

(b)

FIGURE 5.12: Images semi-synthétiques issues de l'ensemble 3 : (a) maillage 1 + faible niveau de bruit, (b) maillage 1 + fort niveau de bruit

(c)



(d)

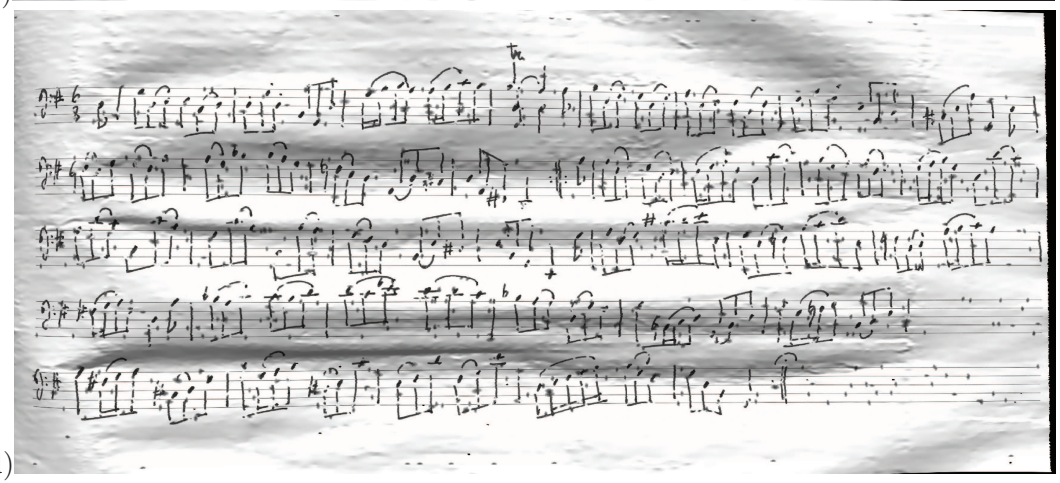


FIGURE 5.13: Images semi-synthétiques issues de l'ensemble 3 : (c) maillage 2 + faible niveau de bruit, (d) maillage 2 + fort niveau de bruit

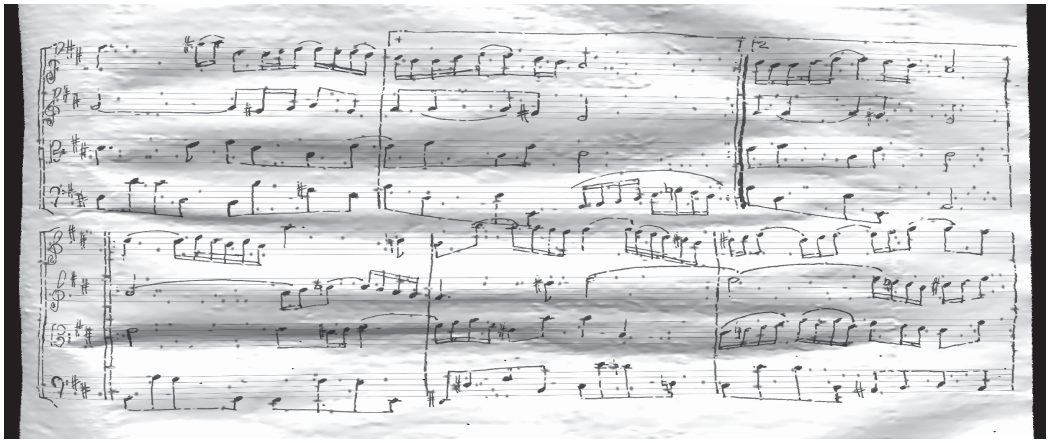
La base de test contient 2000 images semi-synthétiques générées à partir des 333 images originales restantes des 1000 de la base d'origine. Cette base est également séparée en trois ensembles :

- Ensemble 1 : composé de 500 images dégradées par notre modèle de distorsion en 3D. Les deux maillages utilisés sont différents de ceux de la génération de la base précédente. Par conséquent, on obtient 333 $\times$ 2 maillages = 666 images semi-synthétiques mais pour chaque maillage on sélectionne aléatoirement 250 images semi-synthétiques.
- Ensemble 2 : composé de 500 images dégradées par le modèle de bruit local selon trois niveaux de bruit. Les paramètres du modèle de bruit sont les mêmes que dans la base précédente.

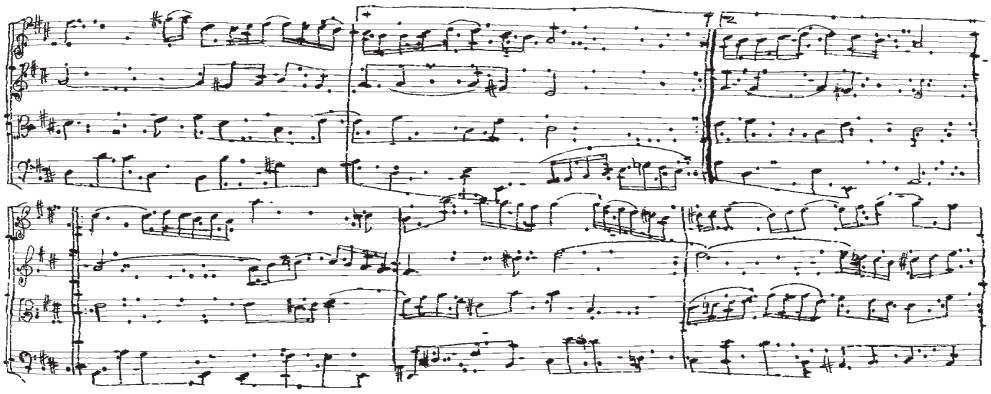
– Ensemble 3 : composé de 1000 images dégradées en utilisant conjointement nos deux modèles de dégradation. On a donc 6 niveaux de dégradation (2 maillages $\times$ 3 niveaux de bruit). Par conséquent, cet ensemble contient 1000 images semi-synthétiques = 500 images semi-synthétiques de l'ensemble précédent $\times$ 2 maillages. Les paramètres du modèle de bruit et les maillages sont les mêmes que pour les 2 ensembles précédents.

Pour ne pas exclure les compétiteurs proposant uniquement des méthodes utilisant en entrée des images binaires, nous avons généré une version binaire des 6000 images synthétiques. Nous proposons donc de participer au concours sur des images semi-synthétiques en niveaux de gris ou binaires. La vérité-terrain utilisée pour l'évaluation finale des algorithmes est générée en binarisant manuellement l'image originale sans les lignes de portée. La figure 5.14 montre un exemple de vérité terrain utilisée pour évaluer les algorithmes utilisés sur les deux premières images.

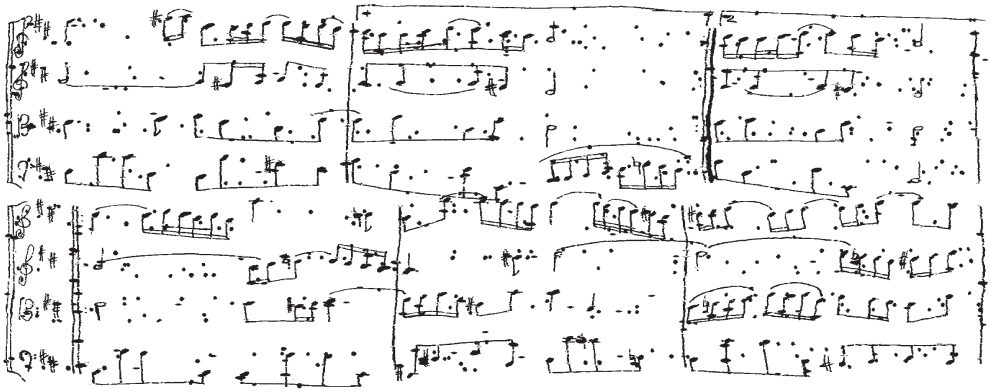
La base d'apprentissage (4000 images) et sa vérité terrain associée ont été données aux participants 46 jours avant la compétition. Le jour de la compétition les 2000 images de la seconde base ont été données sans vérité terrain, puis les concurrents ont fourni les résultats de leurs algorithmes sous forme d'images binaires. Nous avons utilisé ces images résultat pour calculer des mesures de performance à partir de la vérité-terrain.



(a) image semi-synthétique en niveaux de gris



(b) image semi-synthétique binaire



(c) la vérité-terrain des images (a) et (b)

FIGURE 5.14: Exemples de deux versions : (a) image semi-synthétique en niveaux de gris, (b) image semi-synthétique binaire, (c) la vérité-terrain des images (a) et (b)

5.2.1.2 Participants

La compétition a réuni 5 participants provenant de : France, Israël, Singapour, Allemagne et Portugal. Ils ont au total soumis 8 méthodes de suppression de lignes de portée. Six d'entre elles étaient conditionnées à l'utilisation d'images binaires. Les deux dernières méthodes prenaient en entrée des images en niveaux de gris. Nous décrivons ci-dessous ces 8 méthodes.

Le participant 1 : il a soumis une méthode nommée TAU-bin (basée sur celle de [Fujinaga et Adviser-Pennycook, 1997]) qui tout d'abord de calculer la largeur de lignes (*staffline_height*) et l'espace inter-lignes (*staffspace_height*). Puis, tous les symboles musicaux et le texte sont enlevés. Il ne reste normalement plus que les lignes de portée. La rotation est corrigée pour que toutes les lignes soient perpendiculaires avec l'axe de l'image.

Le participant 2 : il a soumis la méthode référencée sous le nom de NUS-bin et décrite dans [Su *et al.*, 2012b]. Les *staffline_height* et *staffspace_height* sont calculées de la même manière que dans la première méthode. Puis, une fonction prédictive est utilisée pour tenter de déterminer la direction des lignes et ensuite les détecter.

Le participant 3 : il a proposé deux méthodes :

- NUASi-bin-lin : La méthode est détaillée dans [Dalitz *et al.*, 2008] et s'applique sur des images binaires. Les *staffline_height* et *staffspace_height* sont détectées en utilisant l'algorithme "vertical scans" [Fujinaga et Adviser-Pennycook, 1997]. Les symboles et le bruit sont enlevés pour localiser uniquement les lignes. Néanmoins, il reste à identifier les régions d'intersection entre les symboles et les lignes (si on enlève les lignes, on enlève également une partie des symboles). Pour résoudre ce problème, les squelettes de lignes sont extraits. Les squelettes sont les entrées d'une fonction *chordlength*(φ) permettant de filtrer les pixels de lignes qui n'appartiennent pas aux symboles (φ est l'angle entre le squelette et le symbole). La méthode est disponible en ligne [2].
- NUASi-bin-skel : La méthode est également détaillée dans l'article [Dalitz *et al.*, 2008] et s'utilise sur des images binaires. Elle détecte les squelettes de lignes, puis les coupe en plusieurs segments au niveau des intersection. Une heuristique permet d'éliminer des segments supposés ne pas appartenir aux lignes. La méthode est disponible en ligne à le même adresse que la précédente.

Le participant 4 : il a également soumis deux méthodes, l'une pour les images binaires (LRDE-bin) et l'autre pour les images en niveaux de gris (LRDE-gray). Les deux

2. http://music-staves.sourceforge.net/

méthodes sont détaillées sur le site[3] :

- LRDE-bin : Cette méthode s'appuie sur des opérateurs mathématiques morphologiques binaires. Tout d'abord est appliqué un filtre dont l'élément structurant est horizontal. Cela permet d'éliminer les symboles et le texte. Puis, un filtre médian est appliqué pour enlever le bruit. Ensuite, une opération de dilatation est réalisée afin de reconstruire les lignes horizontales. Finalement, les lignes horizontales sont filtrées par fermeture morphologique.
- LRDE-gray : L'auteur a amélioré la méthode LRDE-bin afin qu'elle s'adapte aux images en niveaux de gris.

Le participant 5 : il a soumis deux méthodes détaillées dans [dos Santos Cardoso *et al.*, 2009]. Ces deux méthodes utilisent le graphe de "Strong Staff Pixels" – SSPs (un SSP est un pixel ayant une forte probabilité de devenir un pixel de ligne).

- INESC-bin : Les *staffline_height* et *staffspace_height* sont détectées en utilisant l'algorithme proposé dans [Cardoso et Rebelo, 2010]. Tous les pixels de premier-plan détectés lors de l'étape "vertical-run" sont considérés comme des SSPs. Les nœuds du graphe sont les SSPs, les arcs du graphe sont les connexions entre pixels voisins. Les auteurs définissent des heuristiques pour classer les SSPs en deux catégories : les pixels de lignes et les pixels appartenant à des symboles.
- INESC-gray : Pour construire le graphe de SSPs, une méthode de binarisation récente est appliquée. Les pixels de l'arrière-plan ont les poids les plus petits. Inversement, les pixels du première-plan ont les poids plus grands dans le graphe. Une fois que le graphe est construit, la méthode INESC-bin est appliquée [Rebelo et Cardoso, 2013].

La méthode de référence : afin de comparer ces résultats à ceux d'une approche de référence, nous avons décidé d'implanter une méthode nommée baseline et proposée dans [Dutta *et al.*, 2010]. Les *staffline_height* et *staffspace_height* sont détectées pour obtenir des segments de lignes et enlever les symboles. Les auteurs supposent qu'une ligne est caractérisée par une succession de segments générant une ligne de texte de hauteur constante.

5.2.1.3 Mesure d'évaluation

Nous avons utilisé 5 mesures pour comparer les résultats des participants. Les mesures sont calculées pour chaque niveau de dégradation des trois bases sous ensembles de la base de test. Les mesures choisies sont : taux de classification (A), précision (P), rappel

3. http://www.lrde.epita.fr/cgi-bin/twiki/view/Olena/Icdar2013Score

(R), f-mesure (F-M), et taux de spécificité (S). Ces mesures sont respectivement données dans les équations (5.1), (5.2), (5.3), (5.4), et (5.5) où TN est le nombre de "vrais négatifs" (les pixels sont correctement classés comme étant des pixels n'appartenant pas à une ligne de portée), TP est le taux de "vrais positifs" (les pixels sont correctement identifiés comme étant des pixels de ligne de portée), FP est le nombre de "faux positifs" (les pixels sont classés comme des pixels de ligne de portées alors qu'ils ne le sont pas), et FN est le nombre de "faux négatif" (les pixels sont classés comme des pixels n'appartenant pas à une ligne de portée alors qu'ils le sont).

$$taux\ de\ classification = A = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

$$\text{précision} = P = \frac{TP}{TP + FP} \tag{5.2}$$

$$rappel = R = \frac{TP}{TP + FN} \tag{5.3}$$

$$F - Measure = F - M = 2 \times \frac{P \times R}{P + R} \tag{5.4}$$

$$\text{specificité} = S = \frac{TN}{TN + FP} \tag{5.5}$$

5.2.1.4 Analyse des résultats

La majorité des méthodes testées ont une forte valeur de précision mais des valeurs de rappel relativement bas. La figure 5.15 présente les F-mesures des 8 méthodes sous la forme de pourcentages. Nous remarquons que les distorsions influencent les performances de manière plus importante que le bruit local (la F-mesure moyenne sur les distorsions est de 84,91%, celle sur le bruit est de 87,74%). En cas de combinaison des deux dégradations, la F-mesure correspondant aux méthodes utilisées sur des images binaires sont généralement robustes (la F-mesure est égale à 87,4% en moyenne). Dans le cas des images en niveaux de gris, les F-mesures chutent à 61,74% en moyenne (81,86% pour LRDE-gray, 41,63% pour INESC-gray). Ces résultats peuvent s'expliquer par le fait que ces deux méthodes utilisent une méthode de binarisation qui produit donc une perte d'information.

Les F-mesure de la figure 5.15 montrent la tendance globale des résultats obtenus avec toutes les méthodes. Trois méthodes LRDE-bin, NUASi-bin-lin, et NUASi-bin-skel ont des résultats stables et semblent robustes aux deux types de dégradations (F-mesures toujours supérieures à 94%). Les méthodes NUS-bin et la baseline sont robustes au bruit local (F-mesures moyenne $\geq$ 95,33%). En revanche, la présence de distorsions dégrade les

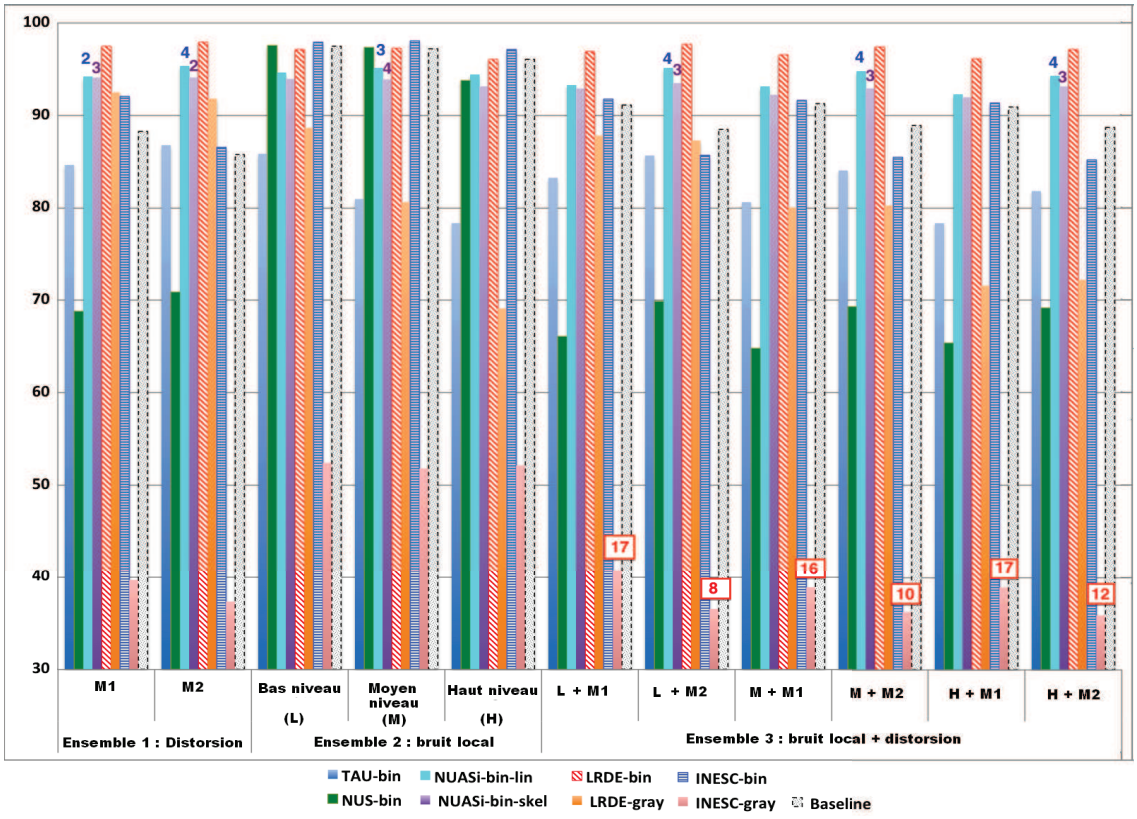


FIGURE 5.15: F-mesure (en %) de 8 méthodes testées lors de la compétition de suppression de lignes de portée à ICDAR 2013. Les chiffres montrés sur les bâtons correspondent au nombre d'images où aucune ligne de portée n'a été détectée.

résultats de ces deux méthodes (la F-mesures moyenne de NUS-bin sur les ensembles de test 1 et 3 est inférieure à 70% et celle de baseline est inférieure à 86%). Pour les autres méthodes, leurs F-mesures ne permettent pas de conclure clairement sur leur robustesse face à nos dégrdations. Les tables 5.2, 5.3, et 5.4 présentent les résultats complets des performances selon trois types de dégradations : les distorsions 3D (l'ensemble de test 1), le bruit local (l'ensemble de test 2), et la combinaison (l'ensemble de test 3).

TABLE 5.2: Résultats de la compétition ICDAR 2013 : images avec distorsion 3D. "#" correspond au nombre d'images rejetées. Les mesures (M.) en % sont : P=Précision, R=Rappel, F-M=F-Mesure, S=Spécificité, A=Exactitude. "M1" et "M2" correspondent aux maillages 1 et 2. Les valeurs entre parenthèses sont les valeurs des mesures calculées avec les images rejetées, c'est-à-dire où aucune ligne de portée n'a été détectée).

Niveau	M.	TAU-bin	NUS-bin	NUASi-bin-lin	NUASi-bin-skel	LRDE bin	LRDE gray	INESC bin	INESC gray	Baseline
M1	P	75.51	98.75	99.05	98.58	98.89	87.26	**99.76**	32.50	98.62
	R	96.32	52.80	#2 $89.90_{(89.77)}$	#3 $90.26_{(90.03)}$	96.19	**98.41**	85.41	50.91	79.86
	F-M	84.65	68.81	$94.25_{(94.18)}$	$94.24_{(94.11)}$	**97.52**	92.50	92.03	39.67	88.26
	S	98.81	99.97	$99.96_{(99.96)}$	$99.95_{(99.95)}$	97.52	99.45	**99.99**	95.97	99.95
	A	98.721	98.25	$99.60_{(99.60)}$	$99.60_{(99.60)}$	**99.82**	99.42	99.46	94.32	99.22
M2	P	82.22	99.50	99.70	99.39	99.52	86.59	**99.90**	34.36	99.29
	R	91.90	55.05	#4 $92.07_{(91.38)}$	#2 $89.63_{(89.36)}$	96.39	**97.76**	76.33	40.85	75.47
	F-M	86.79	70.88	$95.73_{(95.36)}$	$94.26_{(94.11)}$	**97.93**	91.83	86.54	37.33	85.76
	S	99.26	99.99	$99.98_{(99.99)}$	$99.97_{(99.97)}$	99.98	99.44	**99.99**	97.12	99.98
	A	99.01	98.39	$99.71_{(99.68)}$	$99.61_{(99.60)}$	**99.86**	99.38	99.16	95.12	99.10

La table 5.2 détaille les résultats obtenus pour chaque méthode en fonction des distorsions. S'il est difficile de faire émerger un gagnant, il nous semble que la méthode LRDE-bin semble être très bien placée avec des mesures supérieures à 96,19%. La méthode LRDE-gray donne également de bons résultats avec des résultats supérieurs à 87,26%. Les déformations dégradent fortement les résultats des méthodes TAU-bin (P est faible) et NUS-bin (R est faible). Les deux méthodes sont robustes aux rotations, mais ne semblent pas pouvoir gérer des images contenant des courbures du support papier. Ces résultats font bien ressortir le fait que les images dégradées en utilisant le maillage 1 sont plus complexes à analyser que celles utilisant le maillage 2. Autrement dit, les distorsions globales/homogènes, pour les méthodes testées, semblent être plus complexes à gérer que des distorsions locales/hétérogènes. Cela peut s'expliquer par les heuristiques utilisées par la plupart de ces approches souvent basées sur une analyse globale de l'image.

La table 5.3 montre l'impact du bruit local sur chaque méthode. Les performances des méthodes chutent environ de 13 points quand le niveau de bruit augmente. Dans le meilleur des cas, la performance a chuté de seulement 1 point (la méthode INESC-bin). Dans le pire des cas, la performance a chuté de 20 points (LRDE-gray). Les trois méthodes INESC-bin, LRDE-bin, NUS-bin, et la méthode "baseline" sont sensées détecter la direction des lignes, puis de restaurer les lignes contenant des déconnections. Nos tests confirment que ces méthodes sont bel et bien résistantes à un bruit de type local.

TABLE 5.3: Résultats de la compétition ICDAR 2013 : le cas du bruit local. "#" correspond au nombre d'images rejetées. Les mesures (M.) en % sont : P=Précision, R=Rappel, F-M=F-Mesure, S=Spécificité, A=Exactitude. Les valeurs dans les parenthèses sont les valeurs des mesures calculées avec les images rejetées

Niveau	M.	TAU-bin	NUS-bin	NUASi-bin-lin	NUASi-bin-skel	LRDE bin	LRDE gray	INESC bin	INESC gray	Base-line
Haut (H)	P	65.71	95.37	**98.41**	97.28	95.54	53.22	97.63	38.81	95.65
	R	97.01	92.27	90.81	89.35	96.65	**98.58**	96.62	79.35	96.53
	F-M	78.35	93.79	94.46	93.15	96.09	69.12	**97.13**	52.13	96.09
	S	98.59	99.87	**99.95**	99.93	99.87	97.58	99.93	96.51	99.87
	A	98.55	99.67	99.71	99.64	99.79	97.61	**99.85**	96.05	99.78
Moyen (M)	P	69.30	97.82	**99.24**	98.38	97.50	68.10	98.95	39.61	97.26
	R	97.34	96.97	#3 $91.94_{(91.41)}$	#4 $90.56_{(89.80)}$	97.13	**98.77**	97.19	74.83	97.10
	F-M	80.96	97.39	$95.45_{(95.16)}$	$94.31_{(93.90)}$	97.32	80.62	**98.07**	51.81	97.18
	S	98.71	99.93	$99.97_{(99.97)}$	$99.95_{(99.95)}$	99.92	98.61	**99.96**	96.58	99.91
	A	98.67	99.85	$99.75_{(99.73)}$	$99.68_{(99.66)}$	99.84	98.62	**99.89**	95.96	99.83
Bas (L)	P	77.07	98.56	99.25	98.07	97.89	80.65	**99.42**	40.13	98.52
	R	96.88	96.58	90.48	90.17	96.47	**98.47**	96.52	75.48	96.45
	F-M	85.85	97.56	94.66	93.95	97.17	88.67	**97.95**	52.40	97.47
	S	99.12	99.95	**99.97**	99.94	99.93	99.28	99.98	96.59	99.95
	A	99.06	99.86	99.70	99.66	99.84	99.26	**99.88**	95.98	99.85

L'impact de la combinaison de deux dégradations est présenté dans la table 5.4. Les résultats précédents montrent que le maillage M1 est plus complexe à analyser que le maillage M2. Par conséquent, la combinaison de (M1 + H) génèrera des images complexes à analyser alors que celles issues de la combinaison (M2 + L) le seront moins. Les F-mesures présentées dans la table 5.4 confirment ce postulat. La F-mesure moyenne sur l'ensemble (M1+H) est égale à 79,79% tandis que celle sur l'ensemble (M2+L) est égale à 82,40%. Les méthodes NUASi-bin-lin, LRDE-bin, et NUASi-bin-skel sont encore une fois les plus robustes aux différentes dégradations.

TABLE 5.4: Résultats de la compétition ICDAR 2013 : combinaison de dégradations. "#" correspond au nombre d'images rejetées. Les mesures (M.) en % sont : P=Précision, R=Rappel, F-M=F-Mesure, S=Spécificité, A=Exactitude. "M1" et "M2" correspondent aux maillages 1 et 2. Les valeurs entre parenthèses sont les valeurs des mesures calculées avec les images rejetées.

Niveau	M.	TAU-bin	NUS-bin	NUASi-bin-lin	NUASi-bin-skel	LRDE bin	LRDE gray	INESC bin	INESC gray	Base-line
H+M1	P	66.01	94.31	96.88	96.37	96.14	56.19	**97.63**	31.70	96.41
	R	96.35	50.00	88.03	87.93	96.13	**98.59**	85.79	#17 $55.21_{(50.48)}$	85.98
	F-M	78.34	65.35	92.25	91.96	**96.14**	71.58	91.33	$40.27_{(38.94)}$	90.90
	S	98.30	99.89	99.90	99.88	99.86	97.37	**99.92**	$95.93_{(96.27)}$	99.89
	A	98.24	98.25	99.51	99.49	**99.74**	97.41	99.46	$94.58_{(94.76)}$	99.43
H+M2	P	73.40	97.50	**98.55**	98.07	97.61	57.18	98.35	33.11	97.62
	R	92.42	53.56	#4 $90.99_{(90.32)}$	#3 $89.15_{(88.68)}$	96.66	**98.00**	75.17	#12 $42.15_{(39.19)}$	81.26
	F-M	81.82	69.14	$94.62_{(94.25)}$	$93.40_{(93.14)}$	**97.13**	72.22	85.22	$37.09_{(35.90)}$	88.69
	S	98.86	**99.95**	$99.95_{(99.95)}$	$99.94_{(99.94)}$	99.92	97.51	99.95	$97.11_{(97.31)}$	99.93
	A	98.65	98.43	$99.66_{(99.64)}$	$99.59_{(99.57)}$	**99.81**	97.53	99.14	$95.31_{(95.41)}$	99.32
M+M1	P	69.26	95.45	97.52	96.93	97.11	67.44	**98.51**	32.34	97.29
	R	96.44	49.07	89.15	87.98	95.98	**98.46**	85.63	#16 $53.52_{(48.76)}$	85.96
	F-M	80.62	64.81	93.15	92.24	**96.54**	80.05	91.62	$40.31_{(38.88)}$	91.27
	S	98.47	99.91	99.91	99.90	99.89	98.30	**99.95**	$96.01_{(96.36)}$	99.91
	A	98.406	98.168	99.549	99.491	**99.763**	98.312	99.461	$94.556_{(94.730)}$	99.43
M+M2	P	77.50	98.39	99.02	98.53	98.42	68.09	**99.06**	33.76	98.35
	R	91.83	53.47	#4 $91.57_{(90.85)}$	#3 $88.43_{(87.94)}$	96.52	**97.92**	75.21	#10 $41.64_{(39.13)}$	81.08
	F-M	84.05	69.29	$95.15_{(94.76)}$	$93.20_{(92.93)}$	**97.46**	80.27	85.50	$37.29_{(36.25)}$	88.88
	S	99.06	99.96	$99.96_{(99.96)}$	$99.95_{(99.95)}$	99.94	98.38	**99.97**	$97.12_{(97.30)}$	99.95
	A	98.87	98.39	$99.68_{(99.66)}$	$99.56_{(99.54)}$	**99.83**	98.37	99.13	$95.24_{(95.32)}$	99.31
L+M1	P	73.28	96.75	98.06	97.50	97.92	79.32	**99.14**	32.77	97.96
	R	96.38	50.22	88.96	88.74	95.92	**98.38**	85.48	#17 $53.83_{(48.83)}$	85.23
	F-M	83.26	66.12	93.29	92.92	**96.91**	87.83	91.80	$40.74_{(39.22)}$	91.15
	S	98.70	99.93	99.93	99.91	99.92	99.05	**99.97**	$95.93_{(96.30)}$	99.93
	A	98.62	98.17	99.55	99.52	**99.78**	99.03	99.46	$94.44_{(94.62)}$	99.41
L+M2	P	80.17	99.00	99.39	98.94	99.02	78.81	**99.53**	34.31	98.84
	R	91.98	54.01	#4 $91.97_{(91.22)}$	#3 $89.14_{(88.63)}$	96.46	**97.85**	75.18	#8 $41.34_{(39.08)}$	80.14
	F-M	85.67	69.89	$95.54_{(95.13)}$	$93.78_{(93.50)}$	**97.72**	87.30	85.66	$37.50_{(36.54)}$	88.52
	S	99.17	**99.98**	$99.97_{(99.98)}$	$99.96_{(99.96)}$	99.96	99.04	**99.98**	$97.13_{(97.28)}$	99.96
	A	98.92	98.37	$99.70_{(99.67)}$	$99.59_{(99.57)}$	**99.84**	99.01	99.12	$95.18_{(95.25)}$	99.27
∑images rejetées		#0	#0	#21	#18	#0	#0	#0	#80	#0

Les images de documents musicaux semi-synthétiques sont visuellement plus réalistes que celles générées en 2011 [Fornés *et al.*, 2011]. Ces images sont dégradées à l'aide de nos deux modèles de dégradation et permettent de faire varier facilement le niveau de dégradation (dans ce concours, le niveau de bruit local est réparti selon trois niveaux : bas niveau, moyen niveau, et haut niveau), et le type de dégradation (tache isolée, tache connectée, tache de déconnexion, distorsion globale, distorsion locale). Les résultats obtenus permettent également de supposer que la variabilité des dégradations que nous générons est réaliste vis

à vis de celles rencontrées dans la réalité. Par exemple, les meilleures méthodes (LRDE et INESC) sont aussi performantes sur des données réelles que sur nos données synthétiques (plus les dégradations sont fortes, plus les résultats chutent). En conclusion, ce concours a montré que nos deux modèles de dégradation permettent de générer facilement des bases d'images de documents semi-synthétiques pour l'évaluation de performances d'algorithmes d'analyse de structure.

5.2.2 Utilisation d'images de documents semi-synthétiques pour évaluer la robustesse d'un système de segmentation

L'objectif de cette section est de tester la robustesse du système de segmentation d'images présenté dans [Mehri *et al.*, 2013]. En générant une grande variété d'images avec des caractères et zones de texte plus ou moins déformés, nous serons donc en mesure d'identifier les cas de figure pour lesquels la méthode [Mehri *et al.*, 2013] est performante ou non.

Cette méthode se situe dans le contexte de la segmentation d'images de documents en zones de texte ou de graphique suivant une critère d'homogénéité. L'hypothèse avancée par les auteurs est qu'il est possible de s'abstraire du manque d'information à disposition sur les images à analyser (mise en page, taille du texte, script, etc.) en utilisant une approche texture multi-résolution. Sans utiliser de connaissance *a priori*, les auteurs proposent d'appliquer un protocole en deux temps permettant de segmenter les pixels d'une image. Tout d'abord, en utilisant une fenêtre glissante de taille 4×4 pixels, un grand nombre d'indices de texture est calculé pour chaque zone de l'image. Les indices de texture calculés sont : l'autocorrélation, la matrice de co-occurrence et la réponse au filtre de Gabor. En répétant cette étape avec des fenêtres de tailles 4×4, 16×16, 32×32, et 64×64 il est possible de capturer une information multi-résolution sur les textures présentes dans une image de document. A la fin de cette première étape, chaque pixel est décrit par un vecteur de caractéristiques qui est normalisé.

La seconde étape consiste à utiliser les informations calculées précédemment pour regrouper chaque pixel de l'image. Pour cela, un algorithme de "consensus clustering" est utilisé afin d'estimer le nombre de clusters présents dans un groupe d'images. Une fois ce nombre trouvé, une classification ascendante hiérarchique est appliquée. Elle permet d'affecter un label à chaque pixel des images. La figure 5.16 permet de montrer que cette méthode à tendance à regrouper des zones homogènes de pixels, ce qui était attendu. Si le nombre de clusters est fixé à 3, alors on observe un regroupement des pixels en groupes texte/illustration/fond. Lorsque le nombre de clusters augmente, cette méthode à tendance

soit à séparer le cluster de texte en fonction des fontes utilisées, soit à séparer les illustrations selon leurs caractéristiques visuelles.

FIGURE 5.16: Résultats illustrant la qualité de la segmentation d'images

Sur la base de tests réalisés avec un peu plus de 200 images sélectionnées dans Gallica (*http ://gallica.bnf.fr/*), les auteurs de [Mehri *et al.*, 2013] ont montré que leur méthode était robuste à la variété des contenus et des mises en page des documents anciens. Dans le contexte de l'analyse d'images de documents anciens, les auteurs souhaitaient savoir également dans quelle mesure leur méthode était robuste aux bruits présents sur les zones de texte.

5.2.2.1 Protocole expérimental

A partir de la base d'images utilisée dans les expérimentations originales, nous avons généré 150 images semi-synthétiques en ajoutant divers niveaux de notre modèle de déformation de caractères. Nous avons décidé de diviser cet ensemble en 6 sous-ensembles. Les trois premiers (*Is*, *Os*, et *Ds*) correspondent à des sous-ensembles composés d'images dégradées selon un nombre fixe de "points de dégradation" et en générant seulement l'un des 3 bruits. Ce choix vise à évaluer, pour un nombre fixe de "points de dégradation", si l'algorithme de classification est sensible au type de défaut généré. Le choix du nombre de points à utiliser a été manuellement fixé pour générer des dégradations réalistes. Ainsi, chaque image d'*Is* a été dégradée par des taches sombres/claires non connectées à la bordure d'un caractère. Chaque image d'*Os* a été dégradée par des dégradations connectées à un caractère. Enfin, *Ds* contient uniquement des images avec des points de dégradation déconnectant un caractère. Les trois derniers sous-ensembles (*Ld*, *Md*, et *Hd*) tendent à évaluer la méthode,

non plus en fonction du type de dégradation, mais uniquement de la quantité de dégradation présente. Ainsi, chacune des images est composée à part égale des trois types de "points de dégradation". D'un sous-ensemble à l'autre, nous faisons augmenter le nombre de "points de dégradation" et le paramètre jouant sur la taille de l'ellipse générée. Les sous-ensembles *Ld* (bas niveau), *Md* (moyen niveau), et *Hd* (haut niveau) sont générés respectivement avec 1000, 1500 et 2500 "points de dégradation" et des ellipses de taille croissante. La vérité terrain utilisée pour évaluer les performances est celle qui a été saisie manuellement à l'aide de l'environnement GEDI [4].

5.2.2.2 Analyse de résultats

La figure 5.17 permet d'illustrer le type de résultats obtenus sur la base dégradée. Ces résultats ont été obtenus avec le descripteur de texture Gabor, mais représente bien également les résultats obtenus avec les deux autres descripteurs. L'algorithme de classification des pixels est globalement robuste à n'importe quel type de dégradation. On observe une légère diminution de la qualité de la segmentation quand des caractères sont coupés en plusieurs fragments. La texture des caractères ainsi "cassés" est parfois assimilée à la texture d'une illustration (voir figure 5.17.c). Les tests réalisés sur les sous-ensembles *Ld*, *Md*, et *Hd* montrent par contre clairement que la méthode peine à discerner des textures de type texte ou caractère lorsque les dégradations sont trop importantes (Figure 5.17.e-f). De manière générale cela intervient quand les dégradations sont telles qu'elles déforment un caractère de près de la moitié de sa surface (suppression ou ajout de pixels).

Is (a) Os (b) Ds (c) Ld (d) Md (e) Hd (f)

FIGURE 5.17: Résultats des classifications de pixels sur des bases semi-synthétiques dégradées différemment. Les images correspondent à des zones extraites des images qui sont issues de : *Is* (a), *Os* (b), *Ds* (c), *Ld* (d), *Md* (e), et *Hd* (f).

La figure 5.18 résume les résultats obtenus après une analyse quantitative réalisée sur l'ensemble des images de la base semi-synthétique. Nous avons décidé d'utiliser les mêmes

4. http://gedigroundtruth.sourceforge.net/

métriques que celles utilisées par les auteurs de [Mehri *et al.*, 2013] pour les tests réalisés sur des images réelles. Pour chacun des 6 sous-ensembles des images et les images réelles (non dégradées), nous comparons la vérité terrain et les résultats de la segmentation automatique sur la base de 5 métriques : Précision (P), rappel (R), Classification accuracy (CA), Silhouette Width (SW), et purity per block (PPB). Ces résultats confirment qu'en règle générale les trois descripteurs sont résistants à n'importe quel type de bruit local puisque les résultats sont quasiment au même niveau que ceux obtenus sur les images originales.

Les tests effectués sur les trois ensembles *Ld*, *Md*, et *Hd* confirment également que les résultats chutent dès lors que la dégradation augmente. En moyenne, les performances ont chuté de 1% entre la base originale et les images issues de *Ld* et *Md*. Les tests effectués sur les images les plus dégradées montrent que les performances chutent en moyenne de 4%.

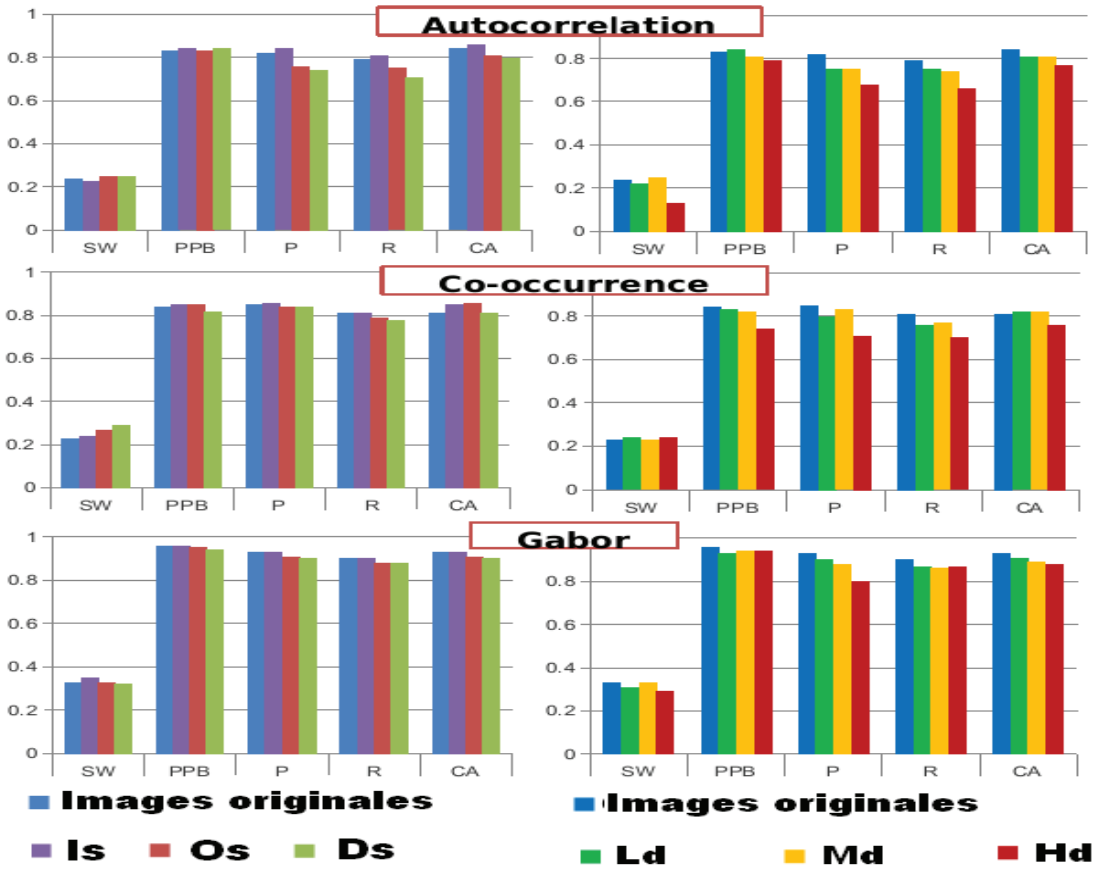


FIGURE 5.18: Les résultats statistiques de trois descripteurs obtenus avec : deux mesures d'évaluation de clustering non supervisées (Silhouette Width (SW), Purity Per Block (PPB)) et trois mesures supervisées (Précision (P), Rappel (R), et le taux de classification (CA)).

Dans cette expérience, nous avons tenté d'évaluer l'impact de notre bruit local sur l'étape d'extraction de trois descripteurs de texture. Grâce aux caractéristiques multi-résolutions textures, l'impact de notre bruit local (de trois types de bruit et de trois niveaux différent) est faible sur la qualité des trois descripteurs calculés. Cela explique pourquoi les résultats de segmentation d'images semi-synthétiques sont similaires à ceux calculés sur les images originales.

5.3 Contribution 4 : utilisation d'images de documents semi-synthétiques contenant les deux dégradations proposées pour l'enrichissement de bases d'apprentissage

Comme évoqué précédemment, l'étude de [Baird, 2000] tend à montrer que l'utilisation d'images semi-synthétiques peut améliorer une étape d'apprentissage. Les études menées dans [Mori *et al.*, 2000, Varga et Bunke, 2003a] ont montré que images semi-synthétiques permettent d'améliorer l'apprentissage dans certains cas précis où les dégradations ajoutées sont la déformation et la distorsion en 2D. De la même manière, nous tenons à réaliser plusieurs expérimentations permettant valider ou non l'utilité de nos modèles dans le cas précis du ré-apprentissage. Au travers de ces expérimentations, nous étudions tout particulièrement l'impact sur les résultats des choix effectués pour générer la base (quantité d'images, combinaison de dégradations ou pas, niveau de dégradation, ...) Dans cette section, nous détaillerons donc deux expérimentations de reconnaissance de caractères et de binarisation menées respectivement sur des images contenant des écritures manuscrites et sur des documents anciens. Ces expérimentations nous permettront de valider nos modèles et de proposer plusieurs conseils sur la manière d'utiliser des données synthétiques pour générer des images d'apprentissage.

5.3.1 Enrichissement d'une base d'apprentissage d'un moteur de reconnaissance d'écritures anciennes

Les bases d'images d'écritures anciennes annotées existantes et publiques sont généralement limitées à quelques dizaines, voire quelques centaines d'images de documents (voir la liste "Handwritten Documents " détaillée sur le site[5]). Dans le cadre des tests présentés ci-dessous, nous avons étudié comment nos modèles de dégradation pouvaient pallier

5. http://www.iapr-tc11.org/mediawiki/index.php/Datasets_List

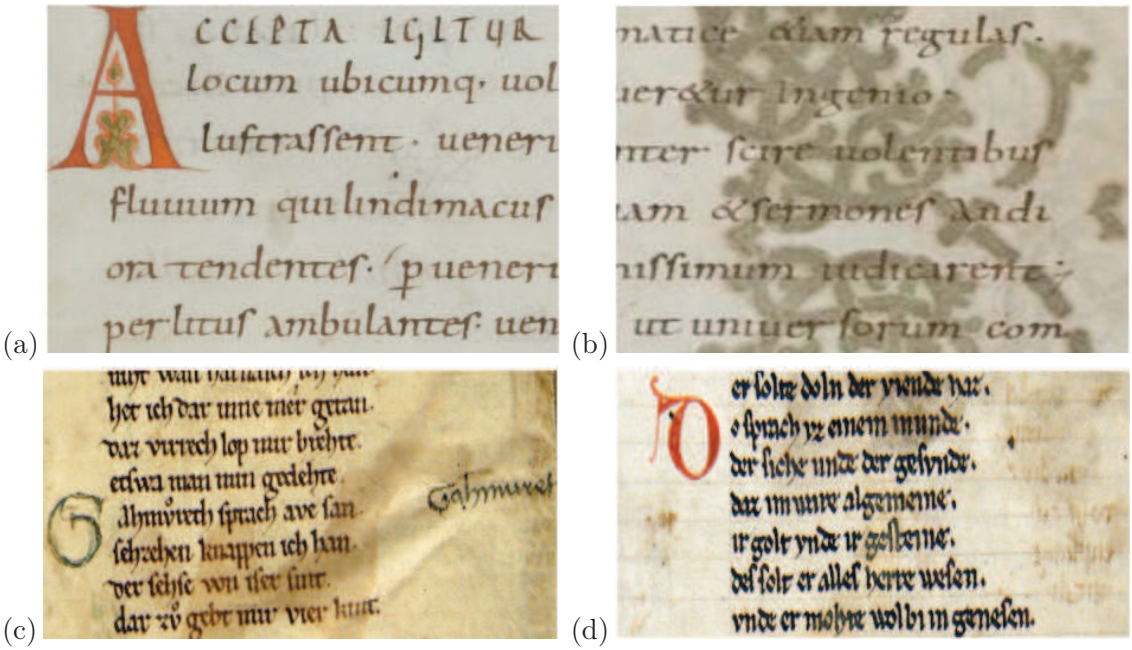
le problème du manque de données d'apprentissage. Nous essayons donc d'enrichir des bases d'images d'écritures anciennes avec des images semi-synthétiques. Ces tests seront l'occasion de répondre à la question suivante : la combinaison de dégradations dans une base d'apprentissage est-elle utile ?

En collaboration avec les auteurs de [Fischer *et al.*, 2012], nous avons travaillé à enrichir la base d'images IAM-HistDB [6]. Nous avons utilisé trois modèles de dégradation : le modèle de bruit local présenté dans la section 4.2, le modèle de bruit de Kanungo [Kanungo *et al.*, 1993], et le modèle de distorsion de [Liang *et al.*, 2008]. Ces modèles nous ont permis de dégrader deux sous-bases d'images de l'écriture manuscrite Parzival [Fischer *et al.*, 2012] et Saingall [Fischer *et al.*, 2011] issues la base IAM-HistDB.

5.3.1.1 Base d'images semi-synthétiques

Deux bases de documents manuscrits sont donc utilisées : les bases Parzival [Fischer *et al.*, 2012] et Saingall [Fischer *et al.*, 2011]. La base Saint Gall contient 60 manuscrits anciens écrits en latin par une seule personne et datant du 9ème siècle (cf. figure 5.19.a-b). Le pré-processus de traitement présenté dans [Fischer *et al.*, 2010] nous permet d'extraire 1410 lignes de texte de cette base. La deuxième base Parzival contient 47 documents rédigés en allemand par 3 scripteurs au 13ème siècle (cf. la figure 5.19.c-d). Pour cette base, nous avons extrait 4477 lignes de texte.

6. http://www.iam.unibe.ch/fki/databases/iam-historical-document-database



(a) (b) (c) (d)

FIGURE 5.19: Images issues de deux bases de documents manuscrits : (a) et (b) base Saint Gall, (c) et (d) base Parzival

Nous proposons de dégrader les deux bases pour créer des versions semi-synthétiques dans la perspective d'augmenter le nombre d'images de la base, ceci afin d'améliorer l'apprentissage. Par conséquent, nous avons convenu de créer quatre ensembles d'images semi-synthétiques :

- Ensemble 1 "Bruit-local" : dans notre article [Kieu *et al.*, 2012], nous avons montré que plus les valeurs de α et β diminuent, plus le niveau de dégradation augmente. De manière générale, plus la taille de chaque région de bruit est grande, plus le niveau de dégradation augmente. Par conséquent, nous avons fixé plusieurs jeux de paramètres pour générer des dégradations différentes.
  - Bas niveau : $\alpha = \beta = 8.5$, $a_0 = 5$, $g = 0.6$ (cf figure 5.20.e-f) ;
  - Niveau moyen : $\alpha = \beta = 7$, $a_0 \in [3, 7]$, $g = 0.6$ (cf figure 5.20.g-h) ;
  - Haut niveau : $\alpha = \beta = 6.2$, a_0 *in* $[3, 10]$, $g = 0.6$ (cf figure 5.20.i-j) ;

FIGURE 5.20: Images issues de l'ensemble "bruit-local" : (a) et (b) deux images originales, (c) et (d) images binarisées, (e) et (f) bas niveau de bruit, (g) et (h) niveau moyen, (i) et (j) haut niveau

– Ensemble 2 "Bruit-kanungo" : dans le modèle de Kanungo [Kanungo *et al.*, 1993] résumé dans la section 3.2.2.1, les deux paramètres α et β permettent de contrôler le nombre de pixels inversés. Plus la valeur des deux paramètres est petite, plus le nombre de pixels inversés augmente et plus le niveau de dégradation augmente. En conséquence, nous diminuons graduellement les valeurs de deux paramètres pour obtenir trois niveaux de bruit. La figure 5.21 présente des images générées comme suit :
  – Bas niveau : $\alpha = \beta = 7$;
  – Niveau moyen : $\alpha = \beta = 5.5$;
  – Haut niveau : $\alpha = \beta = 4.5$;

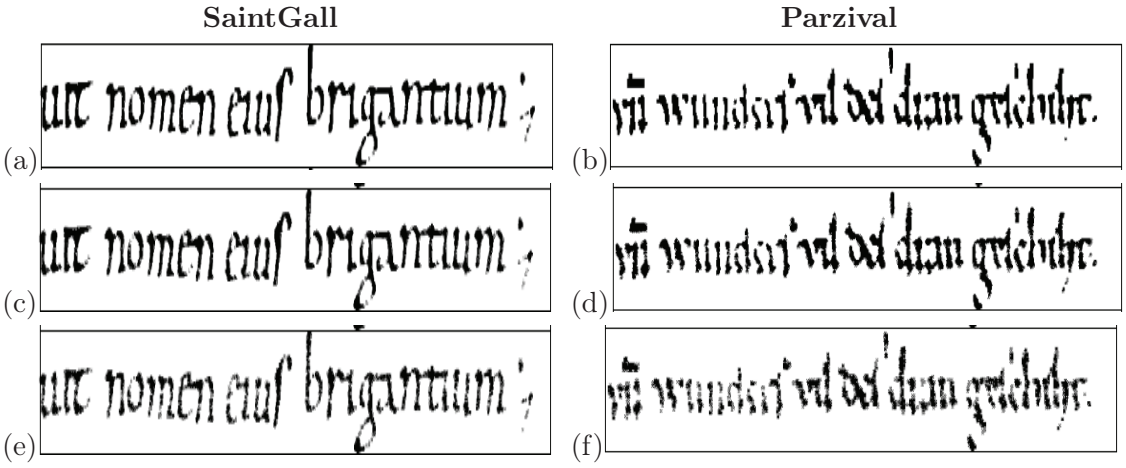


FIGURE 5.21: Images exemples issues de l'ensemble "bruit-Kanungo" : (a) et (b) bas niveau de bruit, (c) et (d) niveau moyen, (e) et (f) haut niveau

– Ensemble 3 "Distorsion" : cet ensemble est généré par le modèle résumé dans la section 3.2.1.3 et détaillé dans [Liang *et al.*, 2008]. Nous sélectionnons deux surfaces développables paraboliques et sinusoïdales pour générer des images courbes. Le niveau de distorsion est contrôlé par l'amplitude a et la longueur d'onde λ ci-après :
  – Bas niveau : surface sinusoïdale avec $a = 10$ et $\lambda = 0.5$
  – Niveau Moyen : surface sinusoïdale avec $a = 15$ et $\lambda = 0.25$
  – Haut niveau : surface parabolique avec $a = 20$ et $\lambda = 0.2$

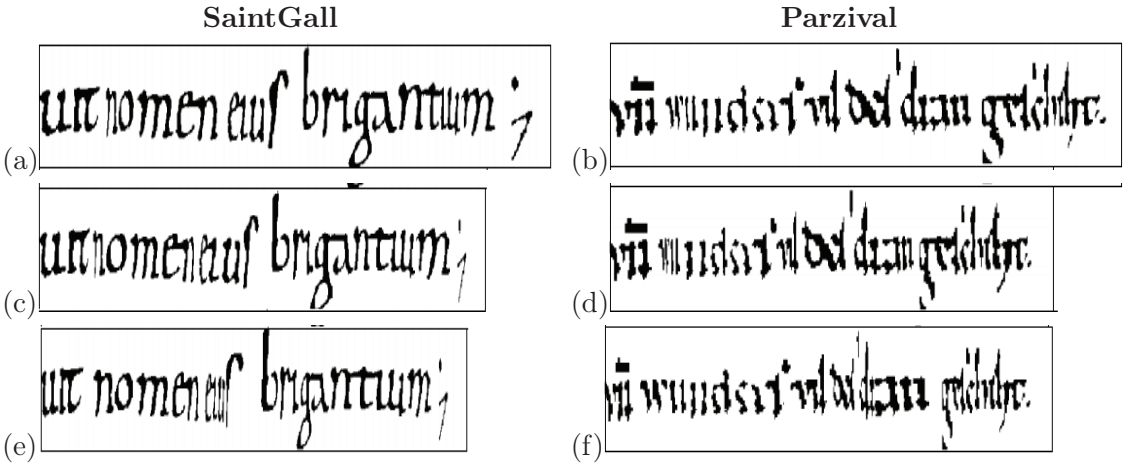


FIGURE 5.22: Images exemples issues de l'ensemble "distorsion" : (a) et (b) bas niveau de bruit, (c) et (d) niveau moyen, (e) et (f) haut niveau

– Ensemble 4 "Combinaison" : cet ensemble combine les images de deux ensembles parmi les trois ensembles précédents. Nous obtenons donc 6 combinaisons possibles.

Pour chaque image de la base d'origine, nous générons donc trois versions semi-synthétiques générées en utilisant l'une des trois dégradations, et 6 versions semi-synthétiques générées en combinant deux des trois dégradations. Ces images servent à entrainer un moteur de reconnaissance d'écriture ancienne afin d'améliorer les résultats de reconnaissance en utilisant simplement les images réelles. Ce moteur et le jeu de paramètres sont détaillés dans la sous-section suivante.

5.3.1.2 Moteur de reconnaissance et jeu de paramètres

Nous utilisons un moteur de reconnaissance d'écriture basé sur un modèle de Markov caché (HMM). Le moteur est détaillé dans les articles [Marti et Bunke, 2001, Fischer *et al.*, 2012].

Un des vecteurs est calculé à partir de descripteurs extraits en utilisant une fenêtre qui se déplace horizontalement sur l'image. La fenêtre permet d'extraire 9 descripteurs : le centre de gravité, les moment géométriques, la proportion de pixels noirs, des informations sur le contour, le nombre de pixels blanc, etc. Les descripteurs sont détaillés dans [Marti et Bunke, 2001]. La vérité terrain est utilisée pour la phase d'apprentissage. Les auteurs utilisent l'algorithme Baum-Welch [Rabiner, 1989]. Le processus de reconnaissance est basé

sur l'optimisation de deux fonctions (algorithme de Viterbi [Marti et Bunke, 2001]) : une fonction permet de trouver la meilleure imagette étiquetée correspondant à l'image de test, l'autre capture la séquence de mots qui satisfait un modèle de langage de type bi-gramme. En parallèle, un dictionnaire est utilisé (5,762 mots pour la base Saint Gall et l'autre de 4,934 mots pour la base Parzival) avec l'algorithme détaillé dans [Zimmermann et Bunke, 2004, Kneser et Ney, 1995] pour corriger les erreurs de reconnaissance de lettres.

Chaque ensemble d'images (réelles ou semi-synthétiques) est réparti en trois sous ensembles. Une partie de chaque ensemble est utilisé pour l'apprentissage, une autre partie pour la validation, et la dernière partie pour les tests. La table 5.5 détaille le nombre d'images de chaque sous ensemble.

TABLE 5.5: Répartition des images pour les tests

	SaintGall	Parzival
Base d'apprentissage	468	2237
Base de validation	235	912
Base de test	707	1328

5.3.1.3 Analyse des résultats

La table 5.6 présente le pourcentage de mots correctement reconnus pour chaque ensemble d'images (bruit-Kanungo, bruit-local, distorsion, et combinaison de défauts). Entre parenthèses est indiquée la différence entre le taux de reconnaissance obtenu avec les images synthétiques par rapport au test réalisé uniquement sur des images réelles.

Dans 7 cas sur 8, les résultats sont significativement améliorés (résultat obtenu avec un test de Student et un risque de première espèce de 5%).Enrichir une base d'apprentissage avec des images semi-synthétiques permet d'améliorer le taux de reconnaissance (+1,82 pour SaintGall et +3,23 pour Parzival). Ces résultats prouvent clairement l'intérêt qu'il y a à intégrer des images synthétiques dans une base, dans l'optique d'améliorer l'apprentissage.

TABLE 5.6: Taux de reconnaissance des mots pour chaque ensemble d'images

	SaintGall	Parzival
Référence	88,99	83,89
Bruit-Kanungo	90,15 (+1,16)	86,95 (+3,06)
Bruit-local	90,42 (+1.43)	85,37 (+1.48)
Distorsion	89,25 (+0.26)	85,66 (+1.77)
Combinaison	**90,81 (+1.82)**	**87,12 (+3.23)**

La table 5.7 présente, pour chaque type de dégradation, le niveaux de dégradation qui a donné les meilleurs résultats (1 : bas niveau, 2 : niveau moyen, 3 : haut niveau).

TABLE 5.7: Le niveau de dégradation plus efficace

	SaintGall	Parzival
Bruit-Kanungo	1	1
Bruit-local	2	1
Distorsion	1	3

Cette étude confirme les conclusions des études menées dans [Mori *et al.*, 2000, Varga et Bunke, 2003a]. Ce sont généralement les images faiblement dégradées qui permettent d'améliorer le plus la reconnaissance. Nos tests ont également montré que générer des images synthétiques en combinant des dégradations et de les intégrer dans une base d'apprentissage permettait d'améliorer significativement la capacité de reconnaissance du système.

5.3.2 Enrichissement d'une base d'apprentissage d'un système de prédiction des résultats de binarisation

Les études de [Baird, 2000, Mori *et al.*, 2000, Varga et Bunke, 2003a] et notre étude précédente ont montré que les images semi-synthétiques peuvent améliorer l'apprentissage. Néanmoins, il reste deux questions importantes à aborder. La première est de savoir dans quelle mesure les images semi-synthétiques utilisées lors d'un apprentissage doivent être similaires aux images de la base réelle testée ? La seconde question est celle relative à la proportion d'images semi-synthétiques/réelles à respecter pour améliorer l'apprentissage ?

L'algorithme testé est présenté dans [Rabeux *et al.*, 2013]. Il permet de prédire, pour n'importe quelle image de document, l'erreur que produirait l'application de 11 algorithmes différents de binarisation. De ce fait, l'algorithme prédictif détaillé dans [Rabeux *et al.*, 2013] permet de choisir la méthode de binarisation (parmi 11) la plus adaptée à un document donné.

La base originale choisie est la base DIBCO [Pratikakis *et al.*, 2011], qui est la base la plus utilisée pour l'évaluation de performance d'algorithmes de binarisation et qui contient moins de 50 images. Utiliser une base aussi peu fournie peut rendre difficile l'utilisation de méthodes utilisant une validation statistique. Ainsi, pouvoir enrichir une base réelle annotée manuellement avec des images semi-synthétiques permettrait de pallier ce problème. C'est justement l'objectif de cette expérimentation. Nous chercherons plus particulièrement à évaluer le biais introduit par la quantité de données ajoutées à la base d'origine.

5.3.2.1 Prédiction de taux d'erreur de binarisation

L'erreur de binarisation, pour un document et un algorithme donné, est très corrélée à l'état de dégradation de l'image analysée. Les auteurs de [Rabeux *et al.*, 2013] créent pour chaque méthode de binarisation une fonction de prédiction $E_m = f_m(Q_I)$, avec m la méthode de binarisation testée, E_m le taux d'erreur prédit, et Q_I un vecteur caractérisant la qualité de l'image. Q_I est mesuré sur la base du calcul de caractéristiques globales telles que la distribution de l'histogramme de niveaux de gris, la corrélation entre la moyenne des niveaux de gris des pixels considérés comme étant des pixels de dégradation et celle des pixels appartenant à la couche d'encre ou du fond. Q_I est également calculé à partir de caractéristiques mesurant la localisation et la forme des dégradations. Un ensemble d'images pour l'apprentissage (et sa vérité terrain associée) est nécessaire pour générer chaque fonction f_m. Les 11 algorithmes testés sont Bernsen, Kittler, Li, Niblack, Ramesh, Ridler, Shanbag, Kapur, Otsu, Sauvola, et White. Ils sont référencés dans [Rabeux *et al.*, 2013].

5.3.2.2 Protocole de test et analyse de résultats

Les images de documents semi-synthétiques sont générées à partir de 30% de la base originale (manuellement annotée) utilisée par [Rabeux *et al.*, 2013]. Ces 30% d'images réelles ne sont ensuite plus utilisées dans la suite de nos tests. La vérité terrain d'une image réelle est utilisée comme vérité terrain de l'image semi-synthétique qu'elle a permis de générer. Soit T_s l'ensemble des images semi-synthétiques utilisé pour l'apprentissage. Les 70%

d'images réelles sont utilisées pour cet apprentissage selon le découpage suivant : l'ensemble T_o correspond à 50% de ces images sélectionnées aléatoirement et est utilisé pour compléter la base d'apprentissage ; l'ensemble V correspond aux 50% restants et est utilisé pour l'étape de validation statistique. Enfin, nous définissons T comme étant l'union de T_o et T_s.

Puisque le processus de sélection des images générant ces ensembles est aléatoire, les performances calculées lors de l'étape de validation statistique peuvent varier. Ainsi, afin d'obtenir une évaluation objective de l'intérêt qu'il y a à ajouter des images semi-synthétiques à une base composée uniquement d'images réelles, l'ensemble du processus est répété plusieurs fois. Afin de répondre également à la question du nombre adéquat d'images semi-synthétiques à ajouter, nous effectuons des tests où le nombre d'images ajoutées augmente. Enfin, l'ensemble de ce protocole est fait une seconde fois, mais cette fois-ci en utilisant uniquement un bruit classique poivre et sel pour la génération d'images semi-synthétiques. Ceci nous permet de savoir si le modèle de bruit présenté dans la section 4.2 est pertinent et si son aspect visuellement réaliste permet de générer des images plus pertinentes qu'avec un modèle de dégradation moins typique de documents pour une application au ré-apprentissage..

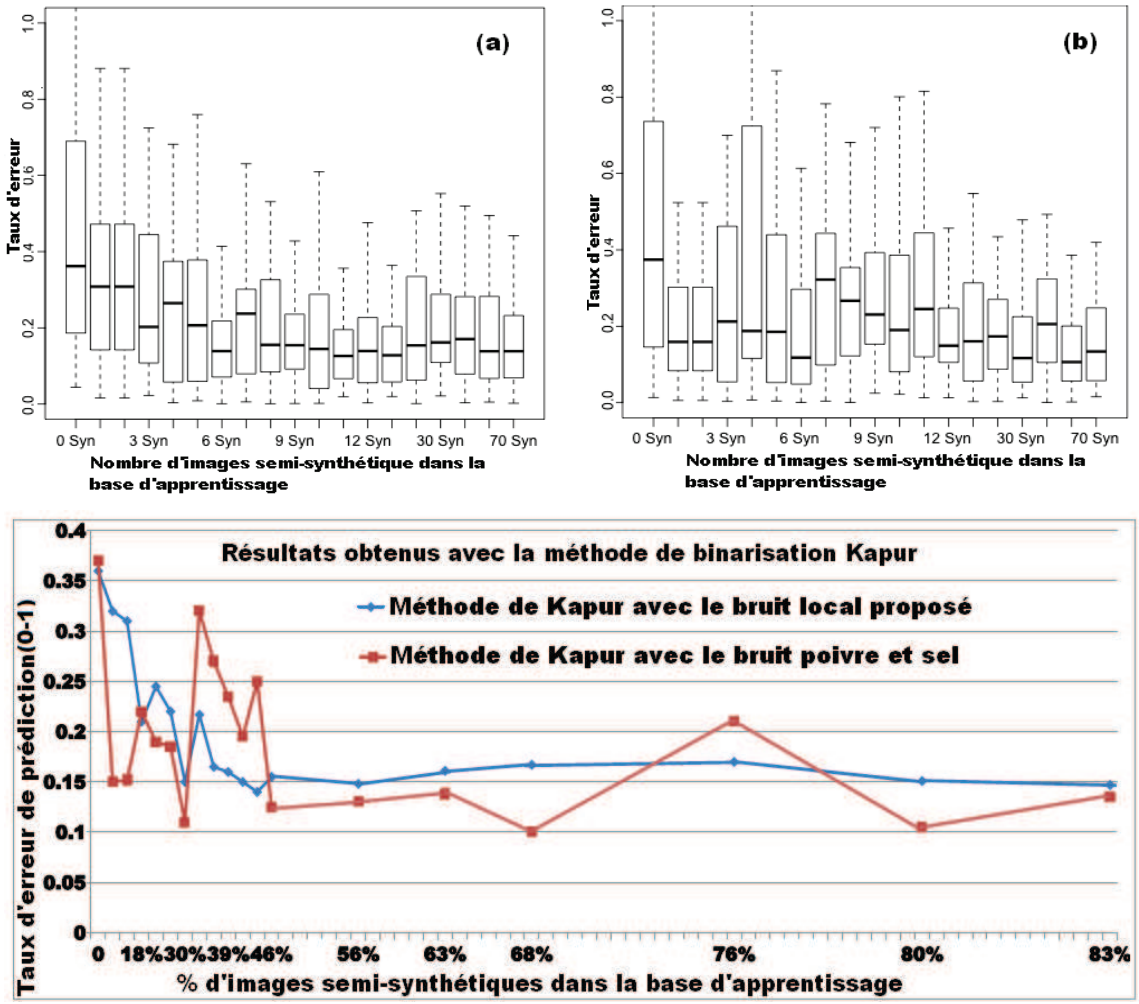


FIGURE 5.23: Le taux d'erreur de la méthode de binarisation Kapur testé avec des images semi-synthétiques dégradées par (a) le modèle de bruit local et (b) le modèle de bruit poivre et sel

Les résultats répertoriés dans la Figure 5.23-a montrent l'erreur moyenne de prédiction obtenue avec la méthode de binarisation de Kapur. Ils illustrent la tendance observée pour chaque modèle de prédiction. Ainsi, plus le nombre d'images semi-synthétiques ajouté à l'ensemble des images pour l'apprentissage est important, plus l'erreur moyenne diminue. On observe que ce taux d'erreur converge à partir de 25 images (soit 50% d'images semi-synthétiques par rapport à la base réelle initiale). En moyenne, l'utilisation d'images de documents semi-synthétiques a permis de diminuer de 15% le taux d'erreur de prédiction.

A titre de comparaison, les résultats obtenus avec des ensembles d'apprentissage

composés d'images dégradées avec un simple bruit poivre et sel sont visibles dans la Figure 5.23-b. On observe clairement que le taux d'erreur ne converge pas spécialement vers une valeur plus faible. Par exemple pour 6 images synthétiques le taux d'erreur est équivalent au taux d'erreur obtenu avec 0 image synthétique. En moyenne, les résultats obtenus avec notre modèle de dégradation sont meilleurs que ceux obtenus avec le bruit poivre et sel. Selon nous, cela vient du fait que nos déformations sont plus réalistes que celles obtenues avec le bruit poivre et sel. De ce fait, l'apprentissage correspond mieux aux types de dégradations réelles apparaissant dans les images utilisées pour l'étape de validation statistique.

Nous avons également réalisé un test de Student afin de vérifier que le ré-apprentissage avec des images semi-synthétiques ajoutées à une base d'images réelles permet d'améliorer significativement la méthode de prédiction. Tout d'abord, la méthode de prédiction est entrainée avec une base de 14 images réelles et testée sur une base réelle de 36 images. En parallèle, la méthode de prédiction est entrainée avec une base de 14 images réelles et 12 images semi-synthétiques puis testée avec une base réelle de 36 images. Le résultat du test de Student montre que notre hypothèse est acceptée (t-test(70) $= 3.648$, p-value < 0.01).

5.4 Conlusion

Nous avons présenté dans ce chapitre comment il était possible d'utiliser des images semi-synthétiques pour deux tâches : l'évaluation de performances d'algorithmes d'analyse de documents et l'enrichissement de bases d'apprentissage. Nous avons montré que pour un objectif d'évaluation de performances, les images semi-synthétiques permettent d'évaluer des algorithmes d'analyse de documents, non seulement grâce à un processus de génération rapide, mais surtout parce que certaines de ces dégradations générées peuvent être contrôlées par l'utilisateur et lui permettent de générer un contenu varié. Ces dégradations variées permettent de savoir précisément à quelle dégradation un algorithme est robuste, ou non. Nous avons également montré que pour un objectif d'apprentissage, ajouter des images semi-synthétiques dans une base d'apprentissage contenant uniquement des images de documents réels permet d'améliorer la performance d'algorithmes d'analyse de documents. Ceci est particulièrement intéressant pour des bases au nombre limité d'images.

Au travers de plusieurs expériences menées en collaboration avec d'autres chercheurs, nous avons pu avancer sur des questions relatives à l'utilisation d'images synthétiques. Si nous pensons que travailler avec des document réels est indispensable, nous pensons également que l'utilisation d'images semi-synthétiques est pertinent dans différents

contextes. Ces images permettent en particulier d'avoir un retour rapide sur les performances d'un algorithme ou d'avoir des données disponibles pour utiliser un algorithme basé sur une étape d'apprentissage, même si le nombre d'images réelles à disposition est faible.

En analysant les résultats de nos tests, nous avons également pu avancer notre réflexion relative à la manière d'utiliser nos images. Ainsi, nous conseillons de ne dégrader que légèrement les images d'origine dans un but de ré-apprentissage. De trop grosses déformations ont tendance à générer des images trop synthétiques non représentatives des dégradations réelles. Nous conseillons également de générer des images synthétiques pour chaque nouveau corpus. Enfin nous conseillons, pour le ré-apprentissage, de construire une base mélangeant à parts égales images réelles et images synthétiques (contenant elles-même des combinaisons de dégradations à bas niveaux).

Il est à noter que nous continuons à développer des collaborations avec d'autres laboratoires. Nous collaborons actuellement avec des chercheurs du laboratoire DIVA de l'université de Fribourg en Suisse pour générer une base d'images de documents anciens synthétiques afin d'évaluer les performances d'algorithmes d'analyse de structure de documents (évaluation et ré-apprentissage).

Une autre perspective de ces travaux est de faire évoluer nos modèles afin de pouvoir les appliquer sur des images en couleur.

Chapitre 6

Conclusion et perspectives

Ce manuscrit présente plusieurs propositions relatives à la constitution de bases semi-synthétiques d'images de documents anciens. L'enjeu est effectivement de taille pour la communauté scientifique travaillant sur le traitement et l'analyse d'images de documents. Si la majorité des tests présentés dans la littérature ont été réalisés sur des données réelles, cette solution nécessitant le plus souvent d'annoter manuellement un volume conséquent de données est complexe à mettre en place. En effet, plusieurs difficultés sont à appréhender : comment faire pour choisir un ensemble d'images représentatif de la réalité ? Combien d'images dois-je annoter sachant que, selon le type de vérité terrain à saisir, le travail peut être colossal ? Quelle est la part de subjectivité que je vais introduire dans la vérité terrain sachant que je peux également être l'auteur de la méthode qui sera testée ? Comment faire, à des fins de reproduction de tests ou de comparaison de performances avec l'état de l'art, pour rendre publiques mes données annotées ?

Nous ne remettons pas en cause l'intérêt qu'il y a à travailler sur des données réelles manuellement annotées. Des travaux de qualité (présentés dans le second chapitre de ce manuscrit) ont d'ailleurs permis de mettre en place des outils performants à l'origine de bases de qualité largement utilisées par la communauté. Notre position est que l'utilisation des données semi-synthétiques ou synthétiques offre une alternative intéressante aux données réelles manuellement annotées. Les données synthétiques offrent en particulier la possibilité intéressante de pouvoir générer rapidement un volume conséquent et varié d'images annotées sans nécessiter un travail d'étiquetage lourd. Il est ainsi possible, très tôt dans un processus de création d'algorithme d'analyse d'image, de pouvoir tester sa méthode ou d'alimenter une phase d'apprentissage. Nous avons détaillé dans ce manuscrit plusieurs références bibliographiques témoignant de l'intérêt de la communauté pour la génération de documents semi (ou totalement) synthétiques (modèles de dégradations à appliquer sur des images réelles, logiciels interactifs de création de documents, ...). Depuis plus de 20 ans, plusieurs expérimentations effectuées sur des données synthétiques ont permis de montrer que leur utilisation était pertinente dans le cadre de l'évaluation de performances ou de génération de données d'apprentissage. A de rares exceptions près, les travaux existants se limitent à la génération de dégradations binaires et à la création de documents synthétiques contemporains (articles, journaux, ...). Le cadre de ces travaux de thèse étant celui des documents anciens, l'utilisation telle quelle des travaux existants limitait leur pertinence. En effet, les documents anciens présentent des dégradations qui leur sont propres (transparence, taches d'encres, déformations du papier, ...) et que l'on retrouve plus rarement dans les images de documents contemporains. Nous avons donc créé deux modèles de dégradation, en niveaux

de gris, reproduisant des défauts typiques des documents anciens, sans toutefois avoir la prétention de reproduire tous les défauts possibles. Au travers de nombreuses expérimentations, nous avons évalué l'intérêt pratique de nos modèles. Enfin, nous avons également développé un logiciel permettant de créer une grande variété de documents semi-synthétiques (choix des fontes, du texte, de la mise en page, du fond, des dégradations, ...). A chaque image générée, est associée sa vérité terrain.

En ce qui concerne les modèles proposés, notre premier apport est un modèle de dégradation reproduisant les déformations (en 3 dimensions) apparaissant couramment dans les images de documents anciens (pliures, coins cornés, petites déchirures du papier, ...). Pour cela, nous faisons l'acquisition, en 3D, d'ouvrages anciens réels. Cette numérisation 3D nous permet d'obtenir un maillage sur lequel nous plaquons, à l'aide d'un algorithme que nous avons créé, une image 2D. Le second modèle permet, quant à lui, de dégrader les caractères. Que ce soit à proximité des caractères (ajouts de taches d'encre) ou à l'intérieur (on simule l'effacement de l'encre), nous arrivons à reproduire des dégradations typiques des documents anciens. Les annexes de ce manuscrit présentent plusieurs exemples d'images dégradées à l'aide de nos deux modèles.

Afin de valider la pertinence de nos modèles, nous avons mené plusieurs tests en collaboration avec d'autres chercheurs. Ce manuscrit présente quatre de ces collaborations. Nous avons réalisé deux campagnes d'évaluation de performances. La première a été réalisée dans le cadre d'un concours de segmentation de lignes de portées musicales dans des documents semi-synthétiques anciens (compétition ICDAR 2013). Nous avons généré les images, défini les métriques utilisées pour l'évaluation et effectué l'analyse des résultats. La seconde campagne d'évaluation de performances a été réalisée dans le cadre du projet ANR DIGIDOC. Nous avons généré une base permettant d'évaluer les performances d'algorithmes de segmentation basés sur l'analyse de texture. Ces algorithmes ont pour but de pouvoir segmenter le contenu d'une page. En parallèle, nous avons mené deux expérimentations visant à utiliser des images semi-synthétiques dans une phase d'apprentissage. Ces deux expérimentations nous ont permis d'apporter des éléments de réponse relatifs à l'utilisation de données semi-synthétiques lors d'un apprentissage. Les images semi-synthétiques générées avec nos modèles de dégradation apportent une plus-value significative statistiquement en termes de performances. Ces images permettent de compléter des bases d'apprentissage de faible volume. Elles permettent également d'obtenir de meilleurs taux de classification. Ces expérimentations ont également montré que les algorithmes de traitement et d'analyse d'images de documents ont, sur nos images semi-synthétiques, des comportements proches

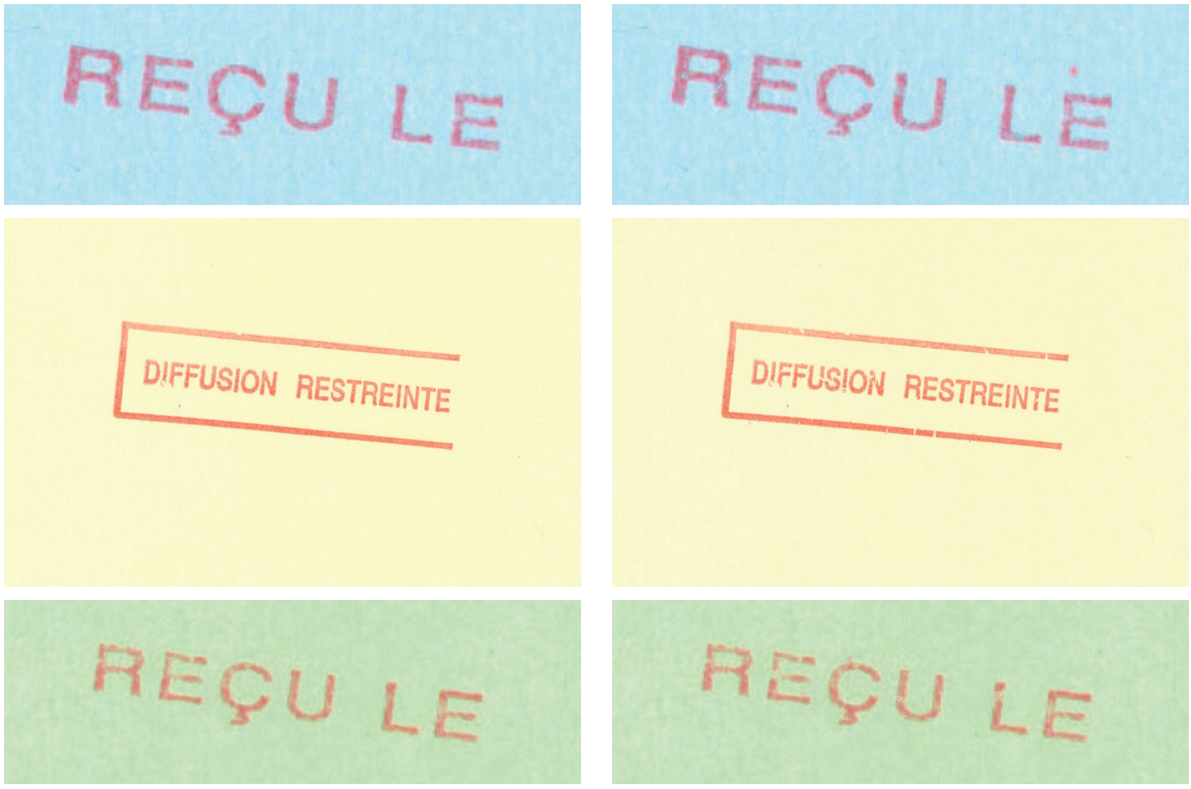
de ceux observés sur les images réelles.

Ces travaux de recherche ouvrent la voie à trois perspectives principales.

Lors des expérimentations, nous avons pu observer que deux questions particulièrement cruciales inhérentes à la création ou à l'utilisation de nos données semi-synthétiques étaient complexes à gérer. Dans le cas d'un ré-apprentissage, le choix de la quantité d'images semi-synthétiques à intégrer dans la base réelle a une influence non négligeable sur la qualité des résultats obtenus. L'autre difficulté est de pouvoir contrôler le niveau de dégradation des images afin de générer des images qui ne soient pas trop dégradées, car nous avons observé que cela pouvait faire chuter les résultats de l'étape de reconnaissance. La première perspective de ces travaux est donc de réfléchir à une méthode automatique qui permettrait de détecter si les images générées avec nos modèles (2D ou 3D) passent un seuil critique de dégradation au-delà duquel l'image n'est plus pertinente pour une tache de ré-apprentissage. De même, nous chercherons à mettre en place un protocole qui, dans la perspective d'un ré-apprentissage, pourra estimer plus finement la proportion idéale d'images semi-syntétiques/réelles à utiliser dans une base d'apprentissage en fonction du type d'approches utilisées.

La deuxième perspective est celle visant à adapter nos modèles de dégradation aux images couleurs. Même si peu d'algorithmes de traitement ou d'analyse d'images de documents anciens utilisent l'information couleur, il nous semble intéressant de pouvoir évaluer la difficulté d'adapter nos algorithmes. Nous avons mené un première expérimentation dans le cadre d'une collaboration avec le service de recherche et développement de l'entreprise Gestform [1]. Nos premiers tests ont consisté à expérimenter une chaine de traitements toute simple : l'image couleur est binarisée pour identifier les points de dégradations (comme dans l'algorithme original). En parallèle, nous utilisons un algorithme de clustering capable d'associer à chaque pixel de l'image couleur une classe (fond/texte). Sur la version d'image composée uniquement des pixels de fond nous appliquons un algorithme d'inpainting permettant de faire disparaitre totalement le blanc laissé par la suppression de la couche d'encre. Sur la version d'image composée uniquement des pixels d'encre, une approche par patch nous permet d'aller chercher dans l'image d'origine des configurations de pixels de fond proches des caractères présentant des dégradés de couleur. Ces patchs sont appliqués sur cette version de l'image aux points de dégradation calculés lors de la première étape. Finalement les

1. Merci à Olivier Augereau, Jean-Marc Nahon et Seif Bernoussi pour le travail réalisé. `http://www.gestform.com`



(a) (b)

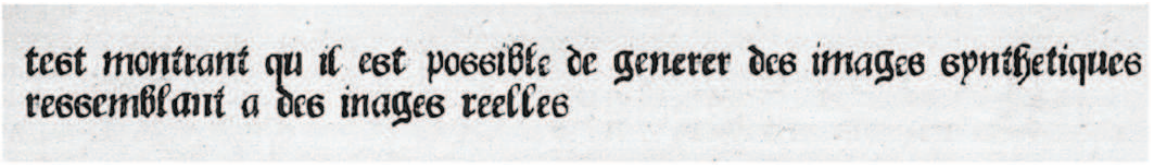
FIGURE 6.1: Deuxième perspective de nos travaux de recherche : pouvoir utiliser le modèle de dégradation de l'encre sur des documents couleur. Colonne (a) image originale (b) image dégradée

deux versions de l'image d'origine (celle ne contenant que le fond après inpainting avec celle contenant uniquement l'encre dégradée) sont fusionnées (opérateur d'union). La figure 6.1 présente des résultats obtenus sur des images de tampons. Nous avons testé plusieurs types de fond et plusieurs couleurs d'encre.

Ces premiers tests tendent à montrer qu'il est compliqué de reproduire (dans les zones que l'on dégrade) le dégradé naturel correspondant à la pression opérée sur la feuille avec le tampon. Il est également compliqué de dégrader une image qui contiendrait un grand nombre de couleurs (par exemple une partie de la zone est surlignée avec un marqueur de couleur différent de celui de l'encre, présence de fonds colorés, ...)

La troisième perspective de nos travaux est de continuer à développer notre plate-forme de génération de documents afin d'offrir la possibilité à un utilisateur expert de créer (pour les

(a)



(b)

FIGURE 6.2: Troisième perspective de nos travaux de recherche : pouvoir générer des images de documents totalement synthétiques à partir d'exemples d'images réelles.(a) image originale (b) image totalement synthétique générée à partir de l'image (a)

usages les plus courants de la communauté) des bases complètement synthétiques qui soient aussi similaires que possible (volume, variabilité des contenus, état de dégradation...) à ses des bases réelles d'images de documents de son choix. Afin de générer synthétiquement des images réalistes à partir d'exemples de documents, nous proposons d'extraire les différents éléments composant ces documents (fond, fontes, caractères, ...), d'apprendre les règles d'édition (positionnement des caractères, mise en page, illustrations, ...) et de recomposer à volonté des documents synthétiques. La Figure 6.2 illustre le type de résultats attendus sur un document ancien.

Le principal verrou scientifique associé à cette perspective est celui de l'apprentissage des modèles de documents. Il s'agit d'apprendre à reproduire, de manière synthétique mais aussi fidèlement que possible, des documents réels. Nous envisageons d'intégrer des étapes d'apprentissage interactif au cœur même du processus de génération d'images synthétiques. Ainsi, plutôt que de demander à un utilisateur de donner une définition exhaustive et de composer entièrement manuellement des documents synthétiques, notre objectif est de concevoir un système basé sur un apprentissage semi-supervisé interactif. Cet apprentissage, à partir d'une base d'entraînement et des retours de l'utilisateur, aura pour vocation d'extraire de manière semi-automatique les modèles de documents, contenant notamment des informations relatives aux règles d'édition (caractères, fontes, positionnement, illustrations et fonds sur lesquels sont imprimés les caractères). *In fine* nous générerons des documents synthétiques (avec vérité terrain associée) ayant des caractéristiques similaires à un ou plu-

sieurs documents de la base d'entraînement.

Ces bases pourront être mises à disposition de la communauté scientifique et des utilisateurs finaux d'outils de traitement ou d'analyse d'images (par exemple des utilisateurs de moteurs d'OCR), afin d'enrichir l'entrainement, de faciliter le paramétrage, ou encore d'évaluer/comparer finement les performances des outils de traitement ou d'analyse d'images utilisés.

Bibliographie

[Tob, 2007] (2007). *The Legacy Tobacco Document Library (LTDL)*. University of California, San Francisco. [23]

[Allier *et al.*, 2006] Allier, B., Bali, N. et Emptoz, H. (2006). Automatic Accurate Broken Character Restoration for Patrimonial Documents. *Int. J. Doc. Anal. Recognit.*, 8(4):246–261. [21]

[Antonacopoulos, 2011] Antonacopoulos, A. (2011). Impact final conference – case study : Scanning parameters. [25, 52, 55]

[Antonacopoulos *et al.*, 2006] Antonacopoulos, A., Karatzas, D. et Bridson, D. (2006). Ground truth for layout analysis performance evaluation. *In Document Analysis Systems VII*, volume VII, pages 302–311, Nelson, New Zealand. Springer. [13]

[Baird, 1990] Baird, H. S. (1990). Document Image Defect Models. *In IAPR workshop on Syntatic and Structural Pattern Recognition*, pages 13–15. Murray Hill, NJ. [30, 32, 62, 72, 97, 118, 121]

[Baird, 1993] Baird, H. S. (1993). Document Image Defect Models and Their Uses. *In Proc. of 2 nd ICDAR 1993*, pages 62–67, Tsukuba City, Japan. [118]

[Baird, 2000] Baird, H. S. (2000). The State of the Art of Document Image Degradation Modeling. *In Proc. of 4 th IAPR International Workshop on Document Analysis Systems 2000*, pages 1–16, Rio de Janeiro, Brazil. [13, 26, 118, 122, 148, 155]

[Bal *et al.*, 2009] Bal, G., Agam, G., Frieder, O. et Frieder, G. (2009). Interactive degraded document enhancement and ground truth generation. *In Document Analysis and Recognition (ICDAR), 2009 10th International Conference on*, pages 743–747. [25, 33]

[Barney Smith, 1998] Barney Smith, E. H. (1998). Characterization of Image Degradation Caused by Scanning. *Pattern Recogn. Lett.*, 19(13):1191–1197. [118, 119, 121]

[Barney Smith, 2000] Barney Smith, E. H. (2000). Estimating scanning characteristics from corners in bilevel images. volume 4307, pages 176–183. [77, 118, 119]

[Barney Smith, 2009] Barney Smith, E. H. (2009). New metric describes edge noise in bilevel images. *SPIE Online Newsroom*. [118]

[Beusekom *et al.*, 2008] Beusekom, J. v., Shafait, F. et Breuel, T. (2008). Automated ocr ground truth generation. *In Document Analysis Systems, 2008. DAS '08. The Eighth IAPR International Workshop on*, pages 111–117. [25, 27, 30]

[Blando *et al.*, 1995] Blando, L., Kanai, J. et Nartker, T. (1995). Prediction of OCR Accuracy Using Simple Image Features. *In Proc. of the Third ICDAR*, volume 1, pages 319–322, Montreal, Quebec. [25, 51, 52, 53, 112]

[Blostein et Baird, 1992] Blostein, D. et Baird, H. S. (1992). *Structured Document Image Analysis*, chapitre A critical survey of music image analysis, pages 405–434. Springer Verlag. [126]

[Breuel, 2008] Breuel, T. (2008). The ocropus open source ocr system. *IS-T/SPIE 20th Annual Symposium*, 2008. [25]

[Cardoso et Rebelo, 2010] Cardoso, J. et Rebelo, A. (2010). Robust staffline thickness and distance estimation in binary and gray-level music scores. *In 20th International Conference on Pattern Recognition (ICPR)*, pages 1856–1859. [138]

[Clausner *et al.*, 2014] Clausner, C., Pletschacher, S. et Antonacopoulos, A. (2014). Efficient ocr training data generation with aletheia. *In Proceedings of the DAS 2014*, Tours, France. International Association for Pattern Recognition (IAPR). [23]

[Curtis *et al.*, 1997] Curtis, C. J., Anderson, S. E., Seims, J. E., Fleischer, K. W. et Salesin, D. H. (1997). Computer-generated Watercolor. *In Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 421–430, New York, NY, USA. [74]

[Dalitz *et al.*, 2008] Dalitz, C., Droettboom, M., Pranzas, B. et Fujinaga, I. (2008). A comparative study of staff removal algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(5):753–766. [126, 137]

[Delalandre *et al.*, 2010] Delalandre, M., Valveny, E., Pridmore, T. et Karatzas, D. (2010). Generation of Synthetic Documents for Performance Evaluation of Symbol Recognition & Spotting Systems. *Int. J. Doc. Anal. Recognit.*, 13(3):187–207. [23, 24, 34]

[Diem *et al.*, 2013] Diem, M., Kleber, F. et Sablatnig, R. (2013). Text line detection for heterogeneous documents. *In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 743–747. [58]

[dos Santos Cardoso *et al.*, 2009] dos Santos Cardoso, J., Capela, A., Rebelo, A., Guedes, C. et Pinto da Costa, J. (2009). Staff detection with stable paths. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6):1134–1139. [126, 138]

[Dutta *et al.*, 2010] Dutta, A., Pal, U., Fornés, A. et Lladós, J. (2010). An Efficient Staff Removal Approach from Printed Musical Documents. pages 1965–1968, Istanbul, Turkey. [58, 138]

[Eikvil, 1993] Eikvil, L. (1993). Ocr - optical character recognition. [52]

[Elliman, 2002] Elliman, D. (2002). Tif2vec, an algorithm for arc segmentation in engineering drawings. *In Selected Papers from the Fourth International Workshop on Graphics Recognition Algorithms and Applications*, GREC '01, pages 350–358, London, UK, UK. Springer-Verlag. [120]

[Fischer *et al.*, 2011] Fischer, A., Frinken, V., Fornés, A. et Bunke, H. (2011). Transcription alignment of latin manuscripts using hidden markov models. *In Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*, HIP '11, pages 29–36, New York, NY, USA. ACM. [149]

[Fischer *et al.*, 2010] Fischer, A., Indermuhle, E., Bunke, H., Viehhauser, G. et Stolz, M. (2010). Ground truth creation for handwriting recognition in historical documents. *In Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, DAS, pages 3–10, New York, NY, USA. ACM. [149]

[Fischer *et al.*, 2012] Fischer, A., Keller, A., Frinken, V. et Bunke, H. (2012). Lexicon-free Handwritten Word Spotting Using Character HMMs. *Pattern Recognition Letters*, 33(7):934–942. [149, 153]

[Fischer *et al.*, 2013] Fischer, A., Visani, M., Kieu, V. C. et Suen, C. Y. (2013). Generation of Learning Samples for Historical Handwriting Recognition Using Image Degradation. *In Proceedings of the 2nd HIP*, pages 73–79, Washington DC, Usa. [26]

[Fornés *et al.*, 2011] Fornés, A., Dutta, A., Gordo, A. et Llados, J. (2011). The ICDAR 2011 Music Scores Competition : Staff Removal and Writer Identification. *In Proc. of ICDAR*, pages 1511 –1515, Beijing, China. [13, 32, 58, 80, 118, 120, 121, 126, 128, 143]

[Fornés *et al.*, 2012] Fornés, A., Dutta, A., Gordo, A. et Lladós, J. (2012). Cvc-muscima : a ground truth of handwritten music score images for writer identification and staff removal. *IJDAR*, 15(3):243–251. [23, 24, 126, 127, 129]

[Fujinaga et Adviser-Pennycook, 1997] Fujinaga, I. et Adviser-Pennycook, B. (1997). *Adaptive optical music recognition*. McGill University. [137]

[Fujita, 2013] Fujita, K. (2013). Extract an essential skeleton of a character as a graph from a character image. *International Journal of Computer Science Issues (IJCSI)*, 10(5). [60]

[Gang Zi, 2005] Gang Zi (2005). GroundTruth Generation and Document Image Degradation. Rapport technique LAMP-TR-121,CAR-TR-1008,CS-TR-4699,UMIACS-TR-2005-08, University of Maryland, College Park. [25, 27]

[Gouinaud *et al.*, 2011] Gouinaud, H., Gavet, Y., Debayle, J. et Pinoli, J.-C. (2011). Color correction in the framework of color logarithmic image processing. *In Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on*, pages 129–133. [21]

[Graves, 2013] Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv :1308.0850*. [30, 31]

[Grosicki *et al.*, 2009] Grosicki, E., Carré, M., Brodin, J.-M. et Geoffrois, E. (2009). Results of the second RIMES evaluation campaign for handwritten mail processing. *In Proceedings of the International Conference on Document Analysis and Recognition*. [23]

[Hale et Barney Smith, 2007] Hale, C. et Barney Smith, E. (2007). Human image preference and document degradation models. *In Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 1, pages 257–261. [50, 77, 118, 119]

[Héroux *et al.*, 2007] Héroux, P., Barbu, E., Adam, S. et Trupin, E. (2007). Automatic ground-truth generation for document image analysis and understanding. *In Document Analysis and Recognition, ICDAR 2007. Ninth International Conference on*, pages 476–480, Curitiba, State of Parana, Brazil. [27, 30]

[Ho et Baird, 1995] Ho, T. K. et Baird, H. S. (1995). Evaluation of OCR Accuracy Using Synthetic Data. *In Proceedings of the 4th Annual Symposium on Document Analysis and Information Retrieval*, pages 413–422, Nevada, USA. [13, 63, 118, 119]

[Ishidera et Nishiwaki, 2003] Ishidera, E. et Nishiwaki, D. (2003). A study on top-down word image generation for handwritten word recognition. *In Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2*, ICDAR '03, pages 1173–, Washington, DC, USA. IEEE Computer Society. [27, 30, 31]

[Jacobi, 2011] Jacobi, J. (2011). Abbyy finereader 10 professional edition review. [25]

[Jain et Yu, 1998] Jain, A. et Yu, B. (1998). Document representation and its application to page decomposition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(3):294–308. [58]

[Jayant Kumar *et al.*, 2013] Jayant Kumar, Peng Ye et Doermann, D. (2013). A Dataset for Quality Assessment of Camera Captured Document Images. *In International Workshop on CBDAR 2013*, pages 39–44, Washington DC, USA. [112]

[Jenkins et Kanai, 1994] Jenkins, F. et Kanai, J. (1994). Use of Synthesized Images to Evaluate the Performance of Optical Character Recognition Devices and Algorithms. *In Proc. of SPIE, Document Recognition 1994*, volume 2181, San Jose, CA, USA. [13, 118]

[Jiuzhou, 2005] Jiuzhou, Z. (2005). Creation of Synthetic Chart Image Database with Ground Truth. Rapport technique, National University of Singapore. [27]

[Jung, 2004] Jung, K. (2004). Text information extraction in images and video : a survey. *Pattern Recognition*, 37(5):977–997. [60]

[Kanungo et Haralick, 1996] Kanungo, T. et Haralick, R. (1996). Automatic generation of character groundtruth for scanned documents : a closed-loop approach. *In Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 3, pages 669–675 vol.3. [25, 27]

[Kanungo *et al.*, 2000] Kanungo, T., Haralick, R., Baird, H., Stuezle, W. et Madigan, D. (2000). A statistical, Nonparametric Methodology for Document Degradation Model Validation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1209 – 1223. [107, 108]

[Kanungo *et al.*, 1993] Kanungo, T., Haralick, R. M. et Phillips, I. (1993). Global and Local Document Degradation Models. *In Proc. of the ICDAR*, pages 730–734, Tsukuba Science City, Japan. [30, 32, 37, 54, 58, 60, 63, 64, 65, 72, 73, 76, 77, 80, 90, 91, 92, 97, 120, 126, 149, 151]

[Kanungo *et al.*, 1998] Kanungo, T., Marton, G. E. et Bulbul, O. (1998). Performance evaluation of two arabic ocr products. *In Proc. of AIPR Workshop on Advances in Computer Assisted Recognition*, volume 3584, pages 14–16, Washington DC, USA. SPIE. [25, 51, 52, 54]

[Kieu *et al.*, 2012] Kieu, V., Visani, M., Journet, N., Domenger, J. P. et Mullot, R. (2012). A Character Degradation Model for Grayscale Ancient Document Images. *In Proc. of the ICPR*, pages 685–688, Tsukuba Science City, Japan. [150]

[Kise *et al.*, 1998] Kise, K., Sato, A. et Iwata, M. (1998). Segmentation of page images using the area voronoi diagram. *Computer Vision and Image Understanding*, 70(3):370–382. [58]

[Kneser et Ney, 1995] Kneser, R. et Ney, H. (1995). Improved backing-off for m-gram language modeling. *In Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, volume 1, pages 181–184 vol.1. [154]

[Lazzara et Géraud, 2014] Lazzara, G. et Géraud, T. (2014). Efficient multiscale sauvola's binarization. *International Journal on Document Analysis and Recognition (IJDAR)*, 17(2):105–123. [21, 56]

[Lazzara *et al.*, 2011] Lazzara, G., Levillain, R., Géraud, T., Jacquelet, Y., Marquegnies, J. et Crépin-Leblond, A. (2011). The SCRIBO module of the Olena platform : a free software framework for document image analysis. *In Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR)*, Beijing, China. International Association for Pattern Recognition (IAPR). [22, 23]

[Lee *et al.*, 2000] Lee, K.-H., Choy, Y.-C. et Cho, S.-B. (2000). Geometric structure analysis of document images : a knowledge-based approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1224–1240. [58]

[Lévy *et al.*, 2002] Lévy, B., Petitjean, S., Ray, N. et Maillot, J. (2002). Least Squares Conformal Maps for Automatic Texture Atlas Generation. *ACM Trans. Graph.*, 21(3): 362–371. [82, 83]

[Li *et al.*, 1996] Li, Y., Lopresti, D., Nagy, G. et Tomkins, A. (1996). Validation of Image Defect Models for Optical Character Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(2):99–108. [107]

[Liang *et al.*, 2008] Liang, J., DeMenthon, D. et Doermann, D. S. (2008). Geometric Rectification of Camera-Captured Document Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(4):591–605. [13, 21, 25, 32, 37, 39, 54, 66, 69, 70, 80, 91, 92, 118, 120, 121, 149, 152]

[Libenzi,] Libenzi, D. Ras2vec 1.2 freeware. online. [120]

[Lienhardt, 1988] Lienhardt, P. (1988). Extension of the notion of map and subdivisions of a 3d space. *In symposium on Theoretical Aspects in Computer Science, Bordeaux, France, LNCS 294, pp. 301-311.* [82]

[Mao *et al.*, 2003] Mao, S., Rosenfeld, A. et Kanungo, T. (2003). Document structure analysis algorithms : a literature survey. volume 5010, pages 197–207. [57, 58]

[Marti et Bunke, 2001] Marti, U.-V. et Bunke, H. (2001). Using A Statistical Language Model to Improve the Performance of An HMM-based Cursive Handwriting Recognition System. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 15:65–90. [153, 154]

[Marti et Bunke, 2002] Marti, U.-V. et Bunke, H. (2002). The iam-database : an english sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5(1):39–46. [123]

[Mehri *et al.*, 2013] Mehri, M., Gomez-Krämer, P., Héroux, P., Boucher, A. et Mullot, R. (2013). Texture Feature Evaluation for Segmentation of Historical Document Images. *In Pro. of the 2nd HIP*, pages 102–109, Washington DC, USA. [144, 145, 147]

[Moghaddam et Cheriet, 2009] Moghaddam, R. F. et Cheriet, M. (2009). Low Quality Document Image Modeling and Enhancement. *In Int. J.Doc. Anal. Recognit*, volume 11, pages 183–201, Berlin, Heidelberg. Springer. [13, 21, 32, 37, 70, 71, 72, 120, 121]

[Mori *et al.*, 2000] Mori, M., Suzuki, A., Shio, A. et Ohtsuka, S. (2000). Generating New Samples from Handwritten Numerals Based on Point Correspondence. *In Proc. 7th Int. Workshop on Frontiers in Handwriting Recognition*, pages 281–290, Amsterdam, Netherlands. [13, 26, 33, 122, 123, 124, 125, 148, 155]

[Nakagawa *et al.*, 2010] Nakagawa, K., Fujiyoshi, A. et Suzuki, M. (2010). Ground-truthed dataset of chemical structure images in japanese published patent applications. *In Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, DAS '10, pages 455–462, New York, NY, USA. ACM. [23, 24]

[Oja et Yuan, 2006] Oja, E. et Yuan, Z. (2006). The fastica algorithm revisited : Convergence analysis. *Neural Networks, IEEE Transactions on*, 17(6):1370–1381. [120]

[Otsu, 1975] Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27. [21, 56]

[Pal *et al.*, 2014] Pal, K., Schüller, C., Panozzo, D., Sorkine-Hornung, O. et Weyrich, T. (2014). Content-aware surface parameterization for interactive restoration of historical documents. *In Computer Graphics Forum*, volume 33, pages 401–409. Wiley Online Library. [82]

[Perez *et al.*, 2009] Perez, D., Tarazon, L., Serrano, N., Castro, F., Terrades, O. R. et Juan, A. (2009). The germana database. *In Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*, ICDAR '09, pages 301–305, Washington, DC, USA. IEEE Computer Society. [23]

[Phillips *et al.*, 1993] Phillips, I., Ha, J., Haralick, R. et Dori, D. (1993). The implementation methodology for a cd-rom english document database. *In Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on*, pages 484–487. [25]

[Phillips, 1999] Phillips, I. T. (1999). How to extend and bootstrap an existing data set with real-life degraded images. *In ICDAR*, pages 689–692. IEEE Computer Society. [25, 32]

[Phong, 1975] Phong, B. T. (1975). Illumination for Computer Generated Pictures. *Commun. ACM*, 18(6):311–317. [87]

[Pratikakis *et al.*, 2011] Pratikakis, I., Gatos, B. et Ntirogiannis, K. (2011). ICDAR 2011 Document Image Binarization Contest (DIBCO 2011). *In Pro. of the 11th ICDAR*, pages 1506–1510, Beijing, China. [23, 24, 156]

[Rabeux *et al.*, 2011] Rabeux, V., Journet, N. et Domenger, P. (2011). Document Recto-verso Registration Using a Dynamic Time Warping Algorithm. *In Proc. of the ICDAR*, pages 1230–1234, Beijing, China. [13]

[Rabeux *et al.*, 2013] Rabeux, V., Journet, N., Vialard, A. et Domenger, J.-P. (2013). Quality Evaluation of Degraded Document Images for Binarization Result Prediction. *IJDAR*, pages 1–13. [156]

[Rabiner, 1989] Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286. [153]

[Rebelo et Cardoso, 2013] Rebelo, A. et Cardoso, J. (2013). Staff line detection and removal in the grayscale domain. *In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 57–61, Washington DC, USA. [138]

[Rebelo *et al.*, 2012] Rebelo, A., Fujinaga, I., Paszkiewicz, F., Marcal, A., Guedes, C. et Cardoso, J. (2012). Optical music recognition : state-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, 1(3):173–190. [126]

[Rice *et al.*, 1996] Rice, S. V., Jenkins, F. et Nartker, T. A. (1996). The fifth annual test of ocr accuracy. *In 1996 Annual Research Report*. Information Science Research Institute. [25, 51, 52]

[Rice *et al.*, 1993] Rice, S. V., Kanai, J. et Nartker, T. A. (1993). An evaluation of ocr accuracy. *In 1993 Annual Research Report*, pages 9–20. Information Science Research Institute. [25, 51, 52, 53]

[Roy *et al.*, 2011] Roy, P., Ramel, J. et Ragot, N. (2011). Word retrieval in historical document using character-primitives. *In Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 678–682, Bejing, China. [22, 23, 24]

[Shahab *et al.*, 2010] Shahab, A., Shafait, F., Kieninger, T. et Dengel, A. (2010). An open approach towards the benchmarking of table structure recognition systems. *In Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, DAS '10, pages 113–120, New York, NY, USA. ACM. [22, 23]

[Shinagawa *et al.*, 1991] Shinagawa, Y., Kunii, T. et Kergosien, Y. (1991). Surface coding based on morse theory. *Computer Graphics and Applications, IEEE*, 11(5):66–78. [82]

[SLIMANE *et al.*, 2009] SLIMANE, F., INGOLD, R., KANOUN, S., ALIMI, A. M. et HENNEBERT, J. (2009). A New Arabic Printed Text Image Database and Evaluation Protocols. *In Document Analysis and Recognition (ICDAR), 2009 10th International Conference on*, pages 743–747. [33]

[Smith, 2008] Smith, E. B. (2008). Modeling Image Degradations for Improving OCR. *In 16th European Signal Processing Conference (EUSIPCO)*, pages 1–5, Lausanne, Switzerland. [32, 72, 76, 97, 118]

[Smith et Qiu, 2003] Smith, E. B. et Qiu, X. (2003). Statistical image differences, degradation features, and character distance metrics. *Document Analysis and Recognition*, 6(3):146–153. [119]

[Smith, 2001] Smith, E. H. B. (2001). Uniqueness of bilevel image degradations. *In Electronic Imaging 2002*, pages 174–180. International Society for Optics and Photonics. [118]

[Smith et Andersen, 2005] Smith, E. H. B. et Andersen, T. (2005). Text degradations and ocr training. *2013 12th International Conference on Document Analysis and Recognition*, 0:834–838. [13, 118, 119]

[Srivastava *et al.*, 2009] Srivastava, R., Parthasarthy, H., Gupta, J. R. P. et Choudhary, D. (2009). Image restoration from motion blurred image using pdes formalism. *In Advance Computing Conference, 2009. IACC 2009. IEEE International*, pages 61–64. [21]

[Su *et al.*, 2012a] Su, B., Lu, S., Pal, U. et Tan, C. (2012a). An effective staff detection and removal technique for musical documents. *In Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on*, pages 160–164. [58, 59]

[Su *et al.*, 2012b] Su, B., Lu, S., Pal, U. et Tan, C. L. (2012b). An effective staff detection and removal technique for musical documents. *In IAPR International Workshop on Document Analysis Systems (DAS)*, pages 160–164. [137]

[Thrin, 2003] Thrin, E. (2003). *De la numérisation à la consultation des documents anciens.* Thèse de doctorat, Université de Lyon. [21]

[Tong *et al.*, 2004] Tong, H., Li, M., Zhang, H. et Zhang, C. (2004). Blur detection for digital images using wavelet transform. *In Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, volume 1, pages 17–20 Vol.1. [20, 21]

[Varga et Bunke, 2003a] Varga, T. et Bunke, H. (2003a). Effects of Training Set Expansion in Handwriting Recognition Using Synthetic Data. *In Proc. 11th Conf. of the Int. Graphonomics Society*, pages 200–203, Scottsdale, AZ, USA. Citeseer. [13, 26, 122, 123, 125, 148, 155]

[Varga et Bunke, 2003b] Varga, T. et Bunke, H. (2003b). Generation of Synthetic Training Data for an HMM-based Handwriting Recognition System. *In Proc. 7th ICDAR*, pages 618–622, Edinburgh, Scotland. [33, 123]

[Wenyin et Dori, 1994] Wenyin, L. et Dori, D. (1994). Incremental arc segmentation algorithm and its evaluation. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 20(1):424–431. [120]

[Wenyin et Dori, 1997] Wenyin, L. et Dori, D. (1997). A protocol for performance evaluation of line detection algorithms. *Mach. Vision Appl.*, 9(5-6):240–250. [58]

[Wenyin *et al.*, 2002] Wenyin, L., Zhai, J. et Dori, D. (2002). Extended summary of the arc segmentation contest. *In Selected Papers from the Fourth International Workshop on Graphics Recognition Algorithms and Applications*, GREC '01, pages 343–349, London, UK, UK. Springer-Verlag. [120]

[Yalniz et Manmatha, 2011] Yalniz, I. et Manmatha, R. (2011). A fast alignment scheme for automatic ocr evaluation of books. *In Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 754–758. [22, 23]

[Yin *et al.*, 2013] Yin, F., Wang, Q.-F. et Liu, C.-L. (2013). Transcript mapping for handwritten chinese documents by integrating character recognition model and geometric context. *Pattern Recogn.*, 46(10):2807–2818. [33]

[Zhai *et al.*, 2003] Zhai, J., Wenyin, L., Dori, D. et Li, Q. (2003). A Line Drawings Degradation Model for Performance Characterization. *In Proc. 7th ICDAR*, pages 1020–1024, Edinburgh, Scotland. [30, 32, 72, 75, 76, 97, 118, 119, 121]

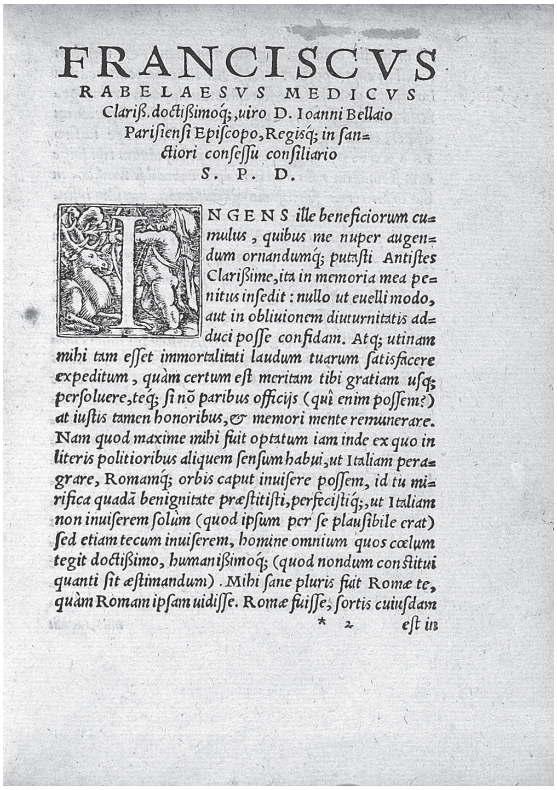
[Zhao et Pope, 2007] Zhao, W. et Pope, A. (2007). Image restoration under significant additive noise. *Signal Processing Letters, IEEE*, 14(6):401–404. [21]

[Zimmermann et Bunke, 2004] Zimmermann, M. et Bunke, H. (2004). Tv-gram language models for offline handwritten text recognition. *In Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on*, pages 203–208. [154]

Annexe A

Modèle de distorsion 3D : les maillages scannés et leurs images d'exemple

Pour dégrader des images de documents en utilisant le modèle de distorsion 3D, nous avons fait l'acquisition de 12 maillages. Pour cela nous avons numérisé (en 3D) des ouvrages réels qui contiennent des dégradtions que nous pensons représentatives des cas les plus fréquemment rencontrés dans la réalité. Dans cette annexe, nous présentons les 12 maillages 3D et le plaquage associé. L'image originale de la figure A.1 est plaquée sur 12 maillages pour générer des images dégradées.



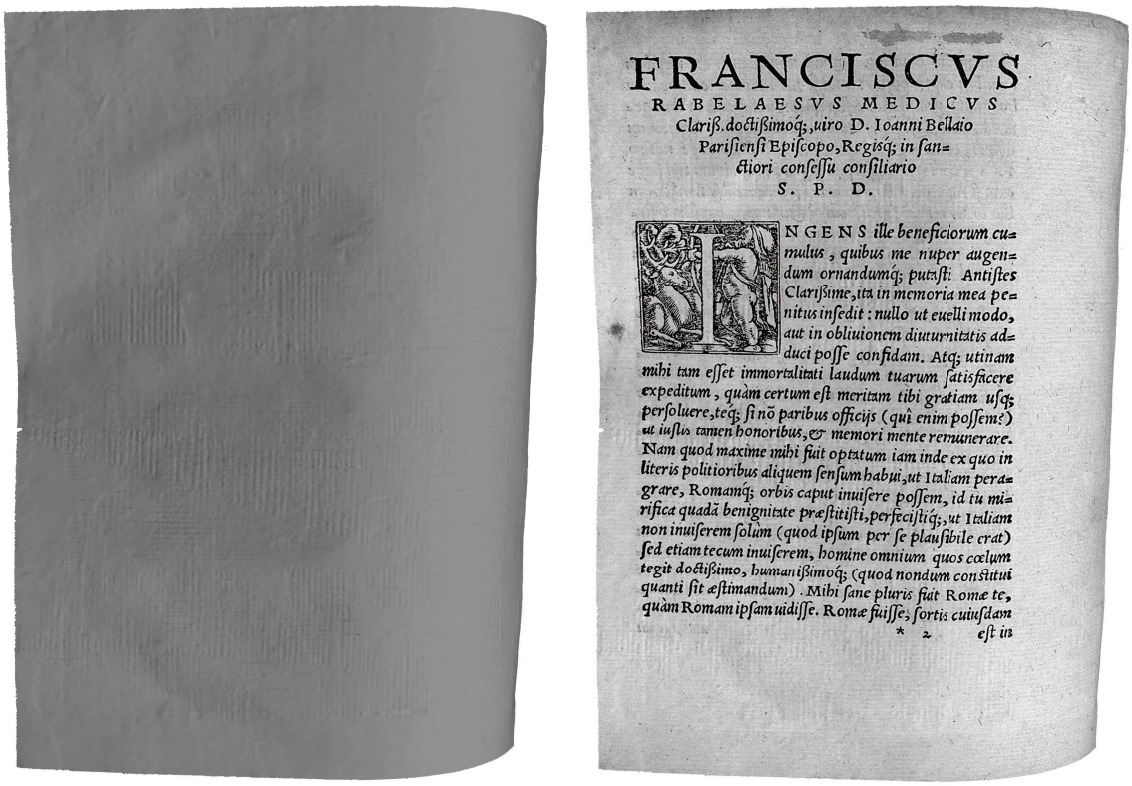
FRANCISCVS
RABELAESVS MEDICVS
Clariß. doctiſſimoq;, uiro D. Ioanni Bellaio
Pariſienſi Epiſcopo, Regiſq; in ſan-
ctiori conſeſſu conſiliario
S. P. D.

INGENS ille beneficiorum cu-
mulus, quibus me nuper augen-
dum ornandumq; putaſti Antiſtes
Clariſsime, ita in memoria mea pe-
nitus inſedit: nullo ut euelli modo,
aut in obliuionem diuturnitatis ad-
duci poſſe confidam. Atq; utinam
mihi tam eſſet immortalitati laudum tuarum ſatisfacere
expeditum, quàm certum eſt meritam tibi gratiam uſq;
perſoluere, teq; ſi nõ paribus officijs (quî enim poſſem?)
at iuſtis tamen honoribus, & memori mente remunerare.
Nam quod maxime mihi fuit optatum iam inde ex quo in
literis politioribus aliquem ſenſum habui, ut Italiam pera-
grare, Romamq; orbis caput inuiſere poſſem, id tu mi-
rifica quadã benignitate præſtitiſti, perfeciſtiq;, ut Italiam
non inuiſerem ſolùm (quod ipſum per ſe plauſibile erat)
ſed etiam tecum inuiſerem, homine omnium quos cœlum
tegit doctiſsimo, humaniſsimoq; (quod nondum conſtitui
quanti ſit æſtimandum). Mihi ſane pluris fuit Romæ te,
quàm Romam ipſam uidiſſe. Romæ fuiſſe, ſortis cuiuſdam

* 2 eſt in

FIGURE A.1: L'image originale fournie par la BNF est plaquée sur 12 maillages pour générer des images dégradées

Maillage 1 Image



FRANCISCVS
RABELAESVS MEDICVS
Clariß. doctiſſimoq;, viro D. Ioanni Bellaio
Pariſienſi Epiſcopo, Regiſq; in ſan-
ctiori conſeſſu conſiliario
S. P. D.

INGENS ille beneficiorum cumulus, quibus me nuper augendum ornandumq; putaſti Antiſtes Clariſsime, ita in memoria mea penitus inſedit: nullo ut euelli modo, aut in obliuionem diuturnitatis adduci poſſe confidam. Atq; utinam mihi tam eſſet immortalitati laudum tuarum ſatisfacere expeditum, quàm certum eſt meritam tibi gratiam uſq; perſoluere, teq; ſi nõ paribus officijs (quì enim poſſem?) ut iuſtus tamen honoribus, & memori mente remunerare. Nam quod maxime mihi fuit optatum iam inde ex quo in literis politioribus aliquem ſenſum habui, ut Italiam peragrare, Romamq; orbis caput inuiſere poſſem, id tu mirifica quadã benignitate præſtitiſti, perfeciſtiq;, ut Italiam non inuiſerem ſolum (quod ipſum per ſe plauſibile erat) ſed etiam tecum inuiſerem, homine omnium quos cœlum tegit doctiſsimo, humaniſsimoq; (quod nondum conſtituì quanti ſit æſtimandum). Mihi ſane pluris fuit Romæ te, quàm Romam ipſam uidiſſe. Romæ fuiſſe, ſortis cuiuſdam

* 2 eſt in

FIGURE A.2: Maillage numéro 1 : contient un distorsion globale

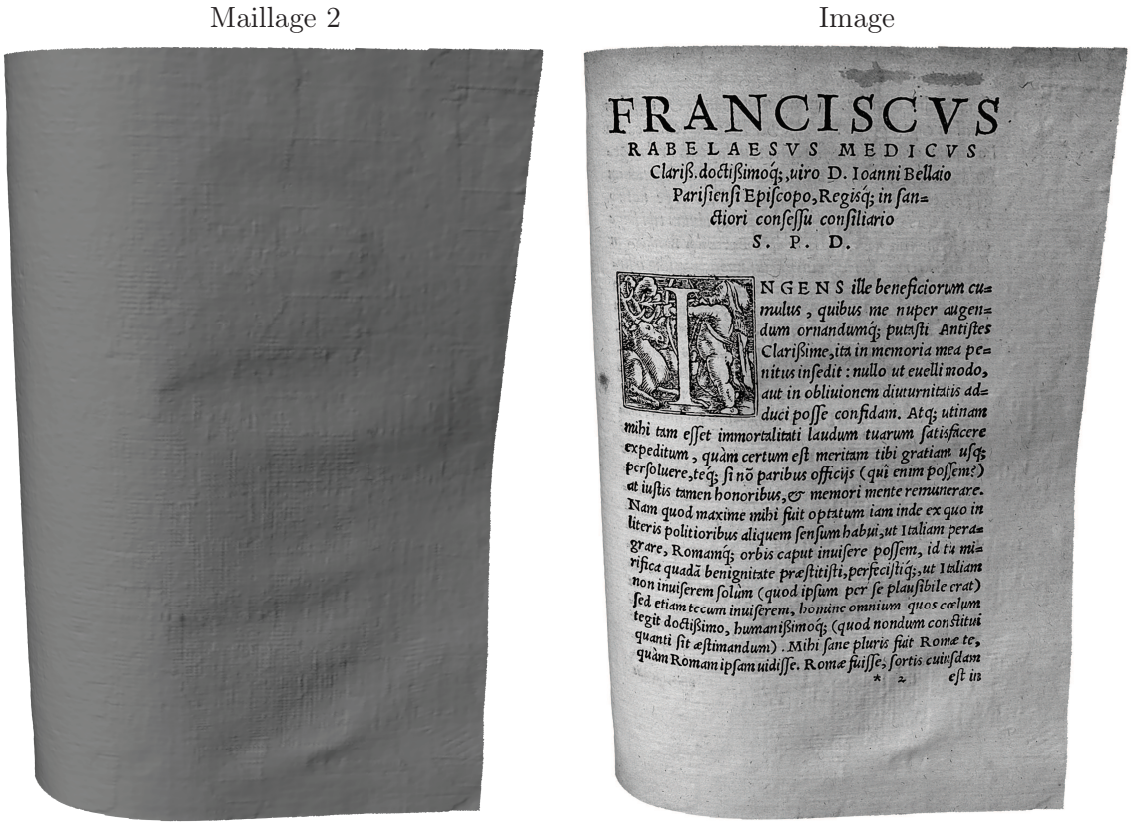
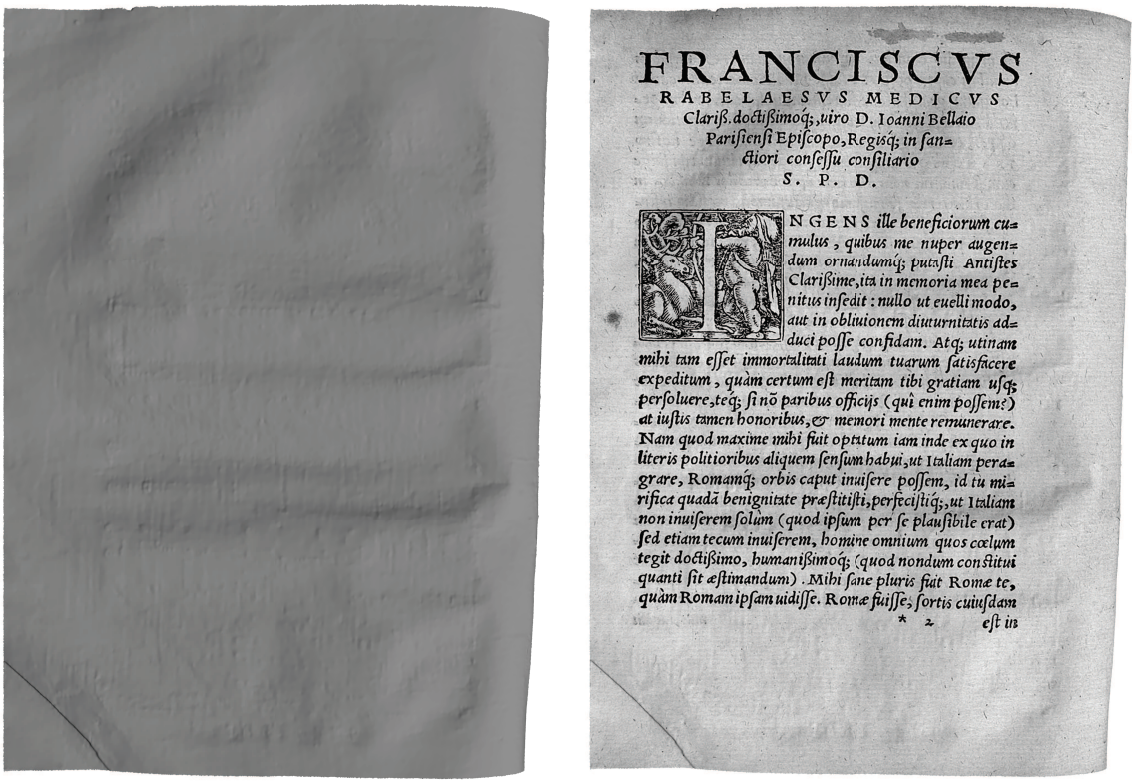


Figure A.3: Maillage numéro 2 : contient une distorsion globale

Maillage 3 Image



FRANCISCVS
RABELAESVS MEDICVS
Clariß. doctißimoq;, uiro D. Ioanni Bellaio
Pariſienſi Epiſcopo, Regiſq; in ſan-
ctiori conſeſſu conſiliario
S. P. D.

INGENS ille beneficiorum cu-
mulus, quibus me nuper augen-
dum ornandumq; putaſti Antiſtes
Clariſſime, ita in memoria mea pe-
nitus inſedit: nullo ut euelli modo,
aut in obliuionem diuturnitatis ad-
duci poſſe confidam. Atq; utinam
mihi tam eſſet immortalitati laudum tuarum ſatisfacere
expeditum, quàm certum eſt meritam tibi gratiam uſq;
perſoluere, teq; ſi nõ paribus officijs (quì enim poſſem?)
at iuſtis tamen honoribus, & memori mente remunerare.
Nam quod maxime mihi fuit optatum iam inde ex quo in
literis politioribus aliquem ſenſum habui, ut Italiam pera-
grare, Romamq; orbis caput inuiſere poſſem, id tu mi-
rifica quadã benignitate præſtitiſti, perfeciſtiq;, ut Italiam
non inuiſerem ſolum (quod ipſum per ſe plauſibile erat)
ſed etiam tecum inuiſerem, homine omnium quos cœlum
tegit doctißimo, humanißimoq; (quod nondum conſtitui
quanti ſit æſtimandum). Mihi ſane pluris fuit Romæ te,
quàm Romam ipſam uidiſſe. Romæ fuiſſe, ſortis cuiuſdam

* 2 eſt in

FIGURE A.4: Maillage numéro 3 : contient des distorsions locales

Maillage 4 Image

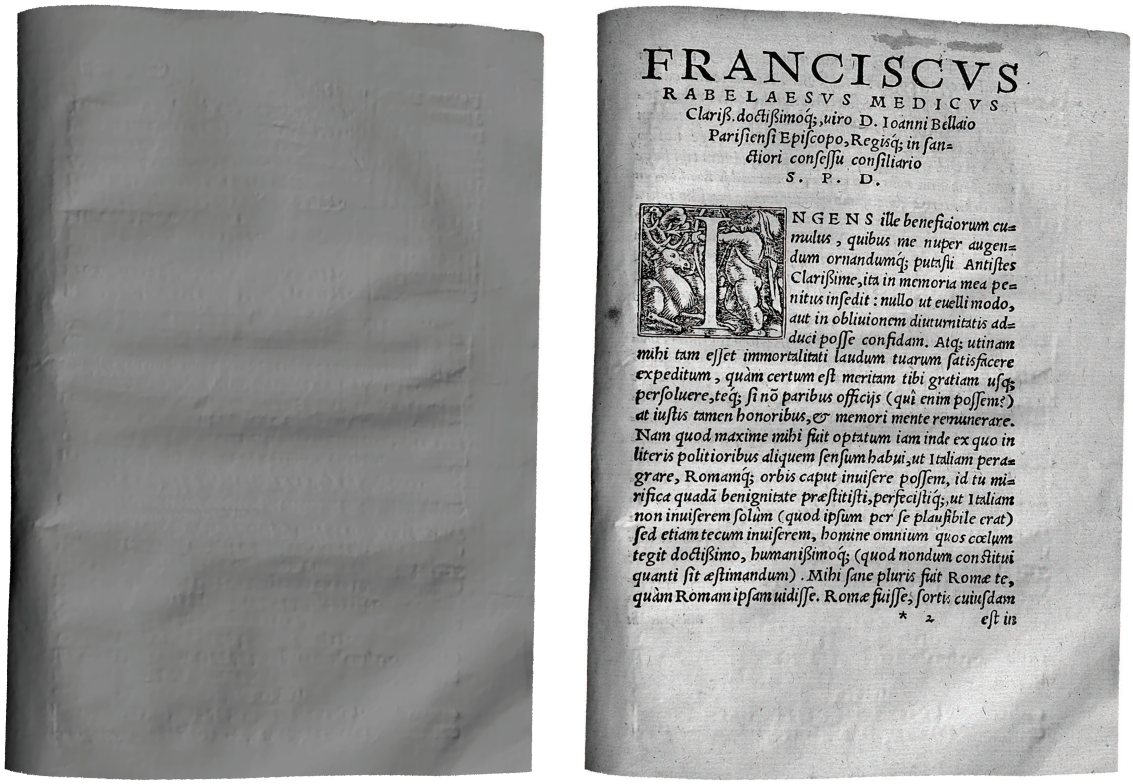
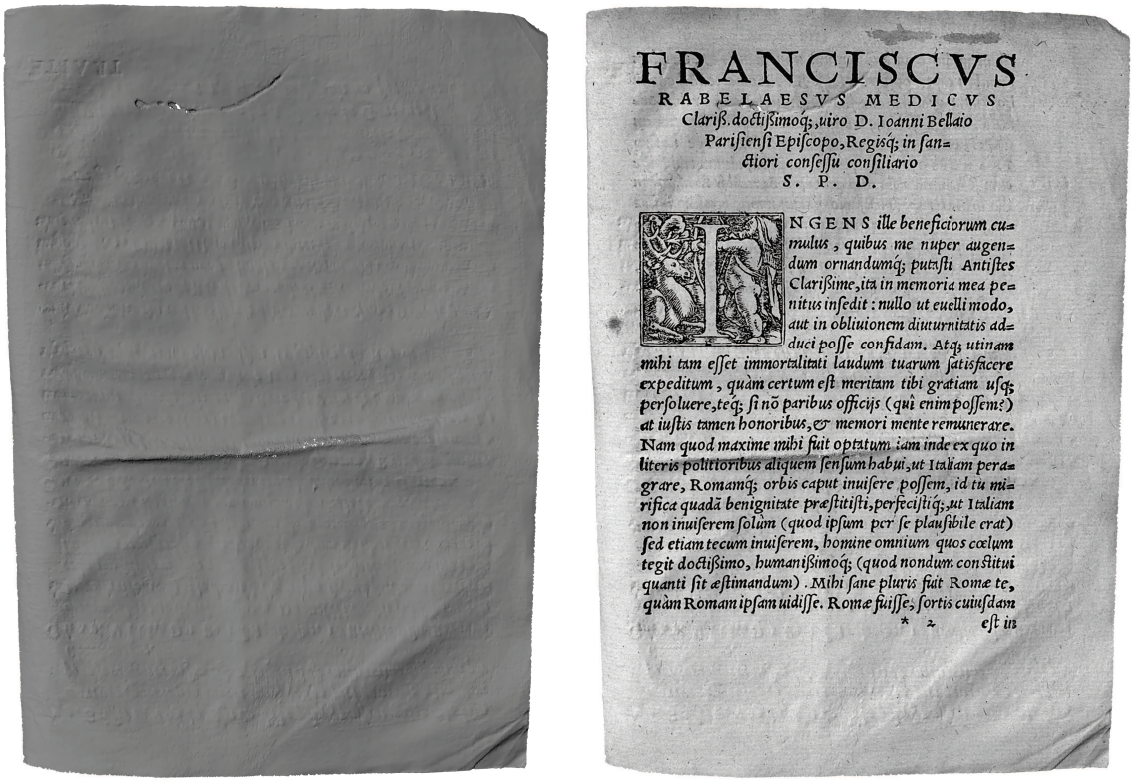


FIGURE A.5: Maillage numéro 4 : contient des distorsions locales

Maillage 5 Image

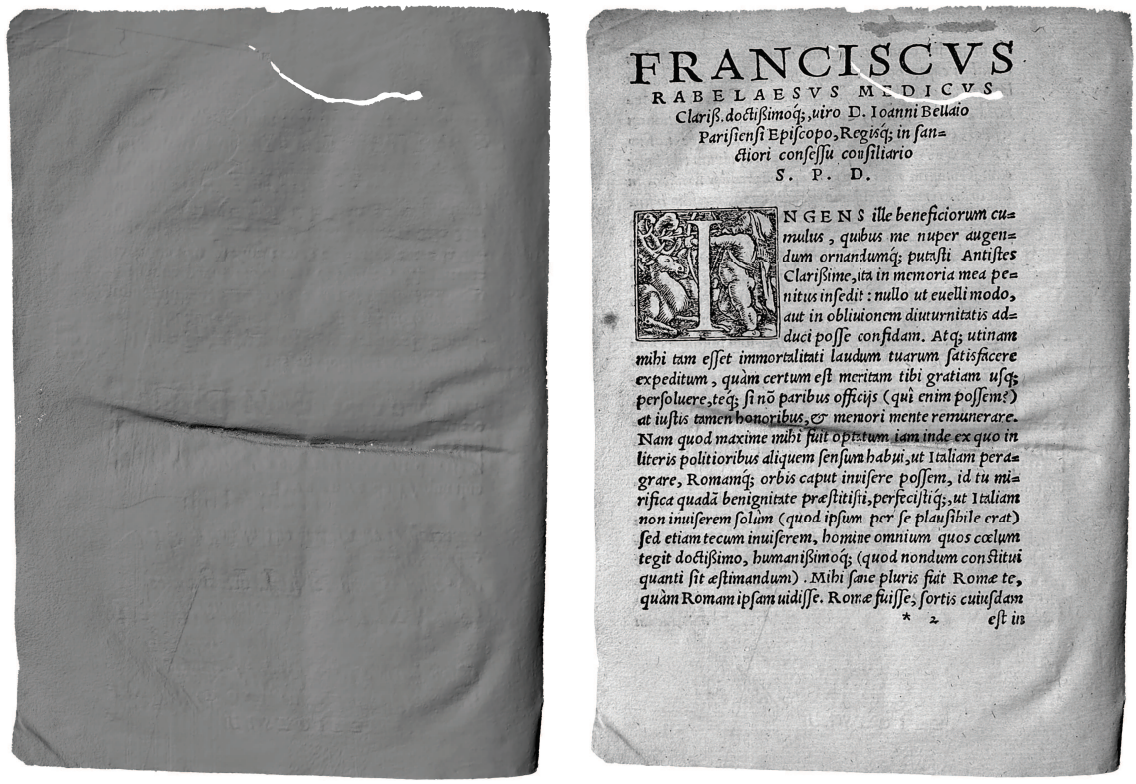


FRANCISCVS
RABELAESVS MEDICVS
Clariß. doctißimoq;, uiro D. Ioanni Bellaio
Pariſienſi Epiſcopo, Regiſq; in ſan=
ctiori conſeſſu conſiliario
S. P. D.

INGENS ille beneficiorum cu=
mulus, quibus me nuper auge=
dum ornandumq; putaſti Antiſtes
Clariſſime, ita in memoria mea pe=
nitus inſedit: nullo ut euelli modo,
aut in obliuionem diuturnitatis ad=
duci poſſe confidam. Atq; utinam
mihi tam eſſet immortalitati laudum tuarum ſatisfacere
expeditum, quàm certum eſt meritam tibi gratiam uſq;
perſoluere, teq; ſi nõ paribus officijs (quî enim poſſem?)
at iuſtis tamen honoribus, & memori mente remunerare.
Nam quod maxime mihi fuit optatum, iam inde ex quo in
literis politioribus aliquem ſenſum habui, ut Italiam pera=
grare, Romamq; orbis caput inuiſere poſſem, id tu mi=
rifica quadã benignitate præſtitiſti, perfeciſtiq;, ut Italiam
non inuiſerem ſolùm (quod ipſum per ſe plauſibile erat)
ſed etiam tecum inuiſerem, homine omnium quos cœlum
tegit doctißimo, humaniſſimoq; (quod nondum conſtitui
quanti ſit æſtimandum). Mihi ſane pluris fuit Romæ te,
quàm Romam ipſam uidiſſe. Romæ fuiſſe, ſortis cuiuſdam
* 2 eſt in

FIGURE A.6: Maillage au numéro 5 : contient des distorsions locales (concavités, trous, plis etc.)

Maillage 6 Image



FRANCISCVS
RABELAESVS MEDICVS
Clariß. doctißimoq; ,uiro D. Ioanni Bellaio
Pariſienſi Epiſcopo, Regisq; in ſan=
ctiori conſeſſu conſiliario
S. P. D.

INGENS ille beneficiorum cu=
mulus , quibus me nuper augen=
dum ornandumq; putaſti Antiſtes
Clariſsime, ita in memoria mea pe=
nitus inſedit : nullo ut euelli modo,
aut in obliuionem diuturnitatis ad=
duci poſſe confidam. Atq; utinam
mihi tam eſſet immortalitati laudum tuarum ſatisfacere
expeditum , quàm certum eſt meritam tibi gratiam uſq;
perſoluere, teq; ſi nõ paribus officijs (quì enim poſſem?)
at iuſtis tamen honoribus, & memori mente remunerare.
Nam quod maxime mihi fuit optatum iam inde ex quo in
literis politioribus aliquem ſenſum habui, ut Italiam pera=
grare, Romamq; orbis caput inuiſere poſſem, id tu mi=
rifica quadã benignitate præſtitiſti, perfeciſtiq;, ut Italiam
non inuiſerem ſolùm (quod ipſum per ſe plauſibile erat)
ſed etiam tecum inuiſerem, homine omnium quos cœlum
tegit doctißimo, humaniſsimoq; (quod nondum conſtitui
quanti ſit æſtimandum) . Mihi ſæpe pluries fuit Romæ te,
quàm Romam ipſam uidiſſe. Romæ fuiſſe, ſortis cuiuſdam

* 2 eſt in

FIGURE A.7: Maillage numéro 6 : contient des distorsions locales, partie déchirée

Maillage 7 Image

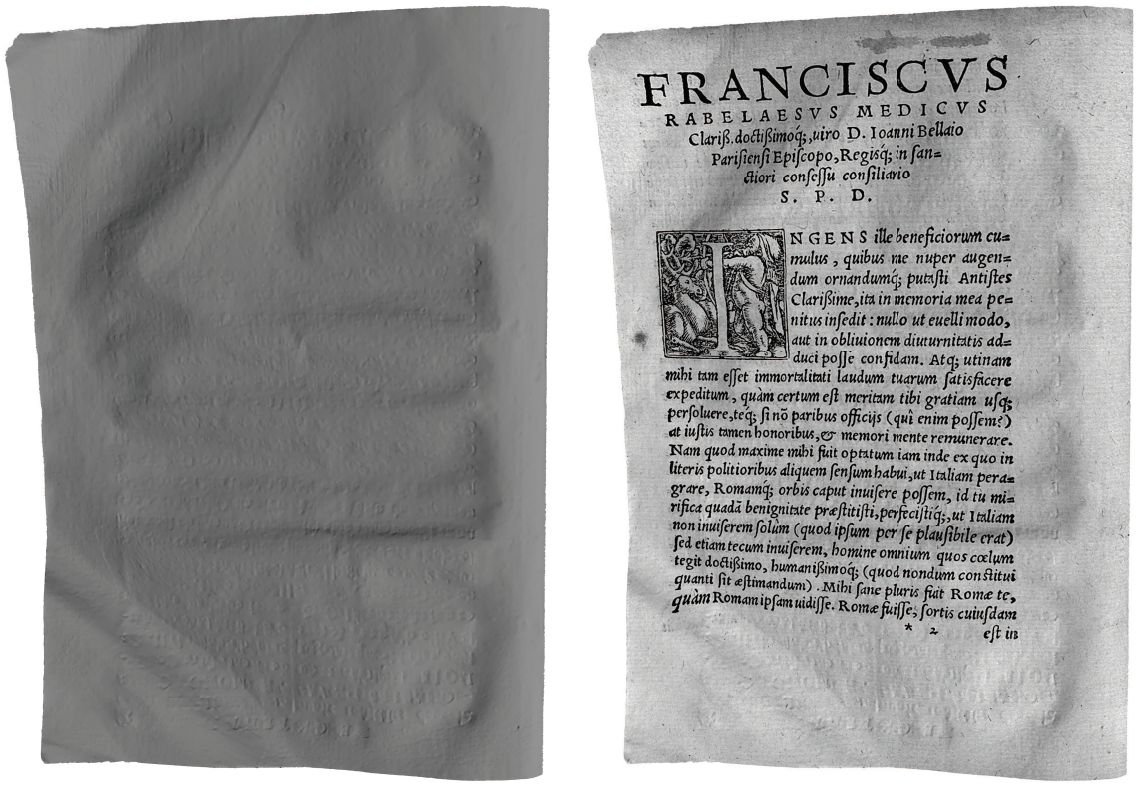


FIGURE A.8: Maillage numéro 7 : contient des distorsions globales et locales

Maillage 8 Image

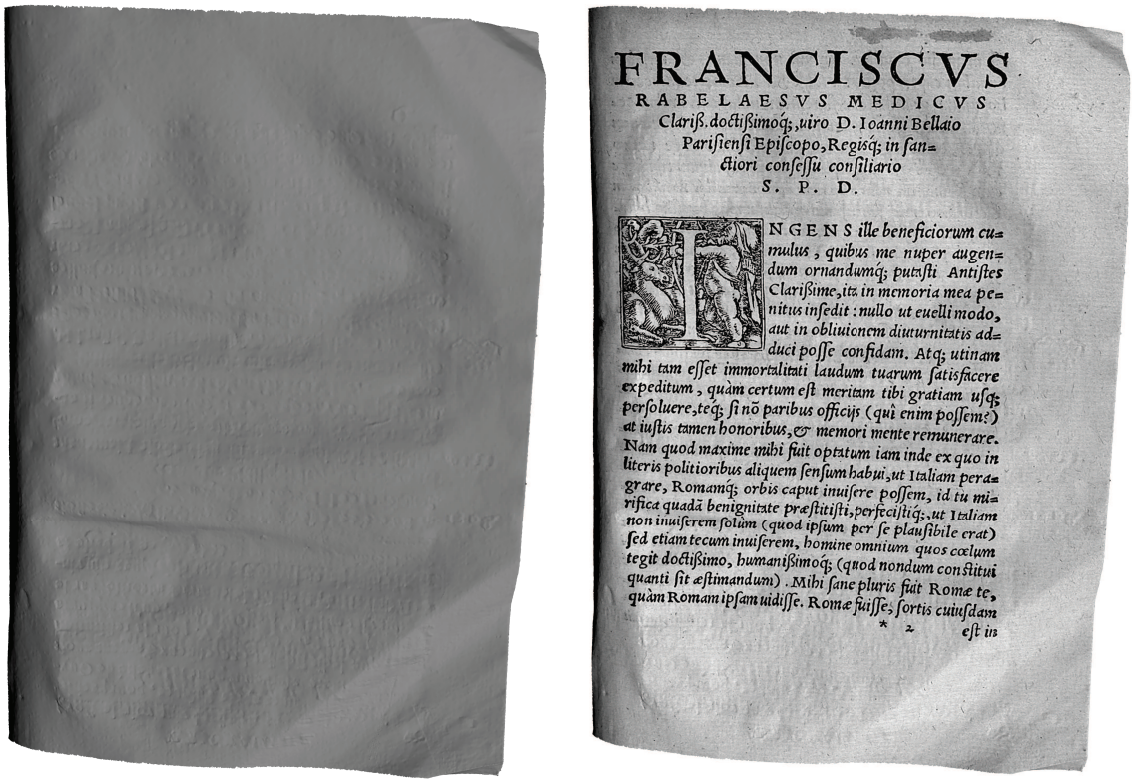
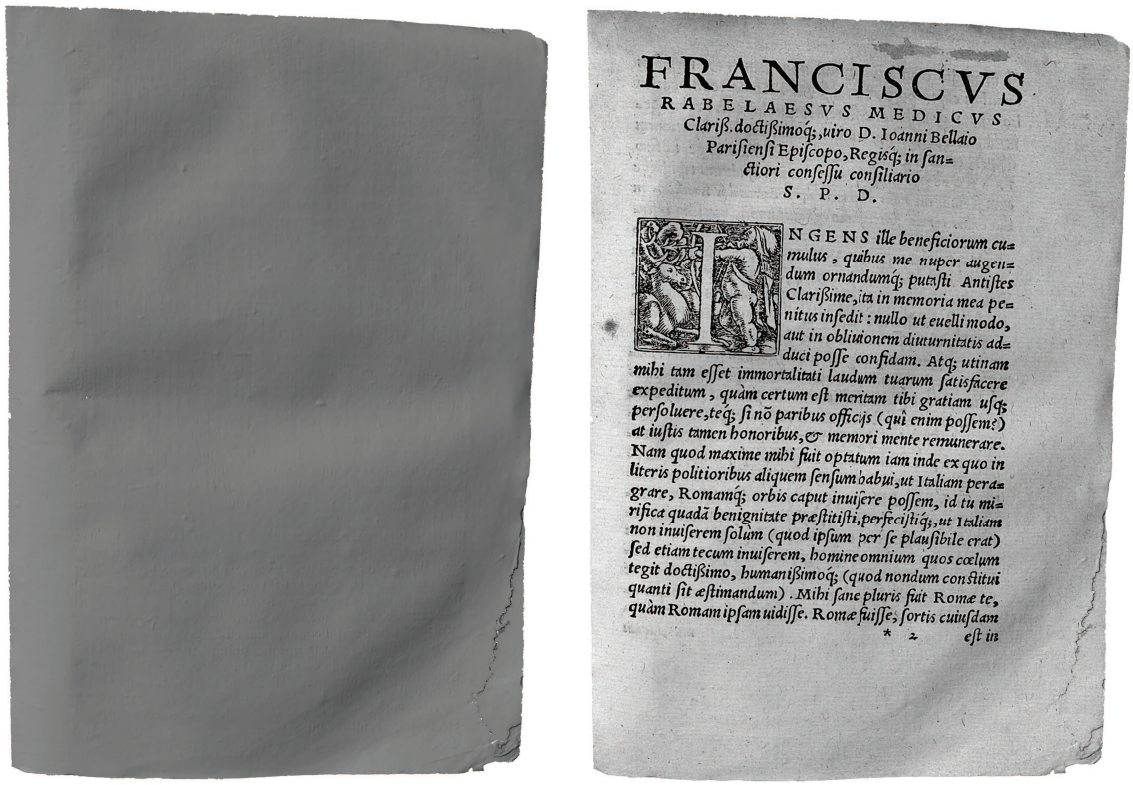


FIGURE A.9: Maillage numéro 8 : contient des distorsions locales

Maillage 9 Image



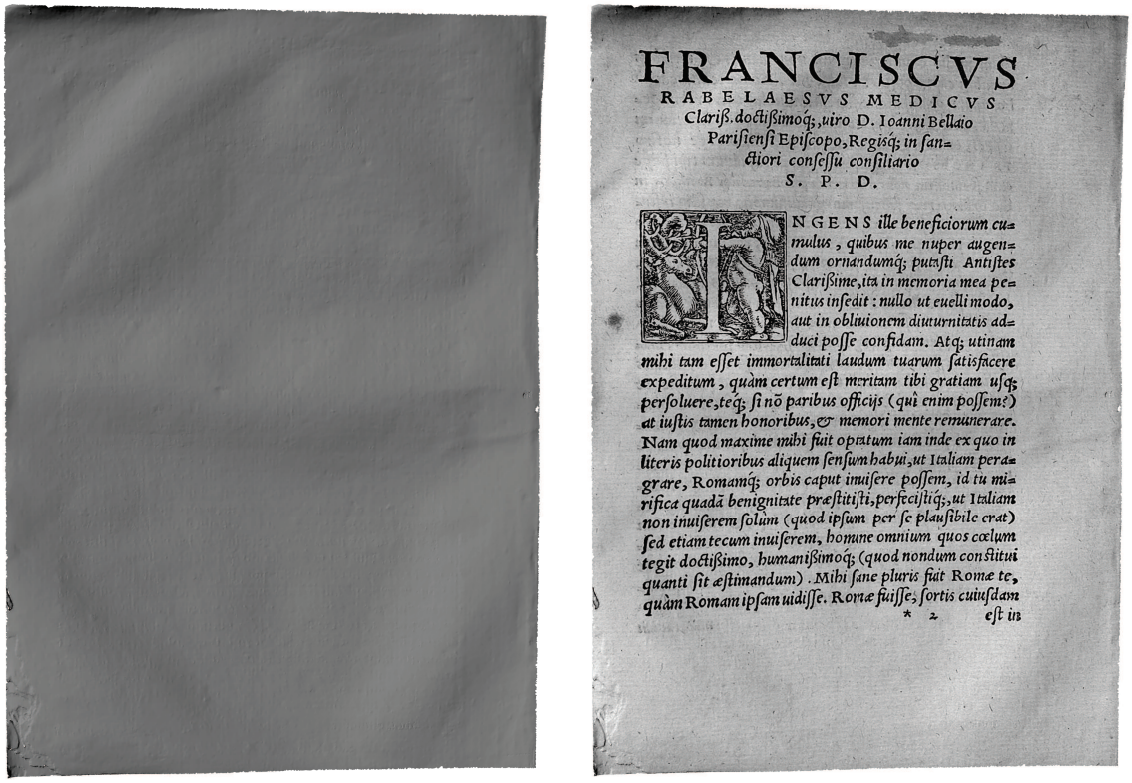
FRANCISCVS
RABELAESVS MEDICVS
Clariſs. doctiſſimoq;, uiro D. Ioanni Bellaio
Pariſienſi Epiſcopo, Regiq; in ſan-
ctiori conſeſſu conſiliario
S. P. D.

INGENS ille beneficiorum cu-
mulus, quibus me nuper augen-
dum ornandumq; putaſti Antiſtes
Clariſsime, ita in memoria mea pe-
nitus inſedit: nullo ut euelli modo,
aut in obliuionem diuturnitatis ad-
duci poſſe confidam. Atq; utinam
mihi tam eſſet immortalitati laudum tuarum ſatisfacere
expeditum, quàm certum eſt meritam tibi gratiam uſq;
perſoluere, teq; ſi nõ paribus offic;js (quì enim poſſem?)
at iuſtis tamen honoribus, & memori mente remunerare.
Nam quod maxime mihi fuit optatum iam inde ex quo in
literis politioribus aliquem ſenſum habui, ut Italiam pera-
grare, Romamq; orbis caput inuiſere poſſem, id tu mi-
rifica quadã benignitate præſtitiſti, perfeciſtiq;, ut Italiam
non inuiſerem ſolum (quod ipſum per ſe plauſibile erat)
ſed etiam tecum inuiſerem, homine omnium quos cœlum
tegit doctiſſimo, humaniſſimoq; (quod nondum conſtitui
quanti ſit æſtimandum). Mihi ſane pluris fuit Romæ te,
quàm Romam ipſam uidiſſe. Romæ fuiſſe, ſortis cuiuſdam

* 2 eſt in

FIGURE A.10: Maillage numéro 9 : contient des distorsions locales et des parties déchirées

Maillage 10 Image



FRANCISCVS
RABELAESVS MEDICVS
Clariſs. doctiſsimoq;, uiro D. Ioanni Bellaio
Pariſienſi Epiſcopo, Regiſq; in ſan=
ctiori conſeſſu conſiliario
S. P. D.

INGENS ille beneficiorum cu=
mulus, quibus me nuper augen=
dum ornandumq; putaſti Antiſtes
Clariſsime, ita in memoria mea pe=
nitus inſedit: nullo ut euelli modo,
aut in obliuionem diuturnitatis ad=
duci poſſe confidam. Atq; utinam
mihi tam eſſet immortalitati laudum tuarum ſatisfacere
expeditum, quàm certum eſt meritam tibi gratiam uſq;
perſoluere, teq; ſi nõ paribus officijs (quì enim poſſem?)
at iuſtis tamen honoribus, & memori mente remunerare.
Nam quod maxime mihi fuit optatum iam inde ex quo in
literis politioribus aliquem ſenſum habui, ut Italiam pera=
grare, Romamq; orbis caput inuiſere poſſem, id tu mi=
rifica quadã benignitate præſtitiſti, perfeciſtiq;, ut Italiam
non inuiſerem ſolùm (quod ipſum per ſe plauſibile erat)
ſed etiam tecum inuiſerem, homine omnium quos cœlum
tegit doctiſsimo, humaniſsimoq; (quod nondum conſtitui
quanti ſit æſtimandum). Mihi ſane pluris fuit Romæ te,
quàm Romam ipſam uidiſſe. Romæ fuiſſe, ſortis cuiuſdam

* 2 eſt in

FIGURE A.11: Maillage numéro 10 : contient des distorsions locales

Maillage 11 Image

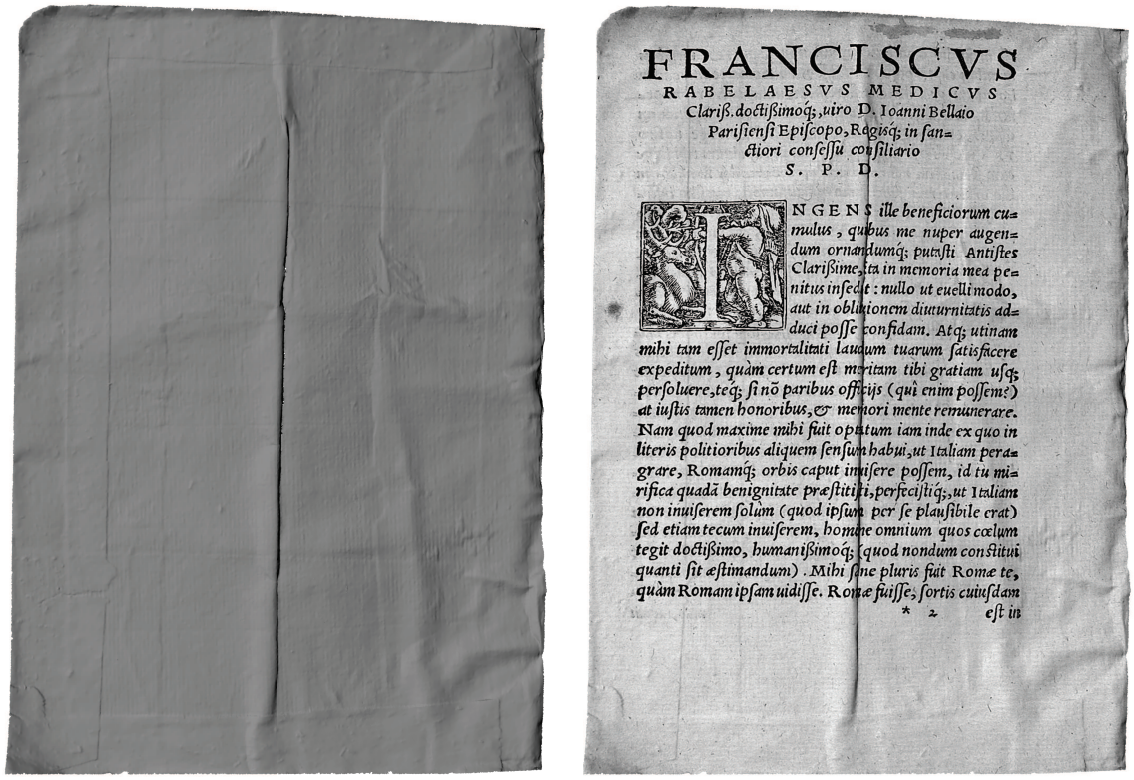
FRANCISCVS
RABELAESVS MEDICVS
Clariſs. doctiſsimoq;, uiro D. Ioanni Bellaio
Pariſienſi Epiſcopo, Regiſq; in ſan-
ctiori conſeſſu conſiliario
S. P. D.

INGENS ille beneficiorum cu-
mulus, quibus me nuper augen-
dum ornandumq; putaſti Antiſtes
Clariſsime, ita in memoria mea pe-
nitus inſedit: nullo ut euelli modo,
aut in obliuionem diuturnitatis ad-
duci poſſe confidam. Atq; utinam
mihi tam eſſet immortalitati laudum tuarum ſatisfacere
expeditum, quàm certum eſt meritam tibi gratiam uſq;
perſoluere, teq; ſi nõ paribus officijs (quì enim poſſem?)
at iuſtis tamen honoribus, & memori mente remunerare.
Nam quod maxime mihi fuit optatum iam inde ex quo in
literis politioribus aliquem ſenſum habui, ut Italiam pera-
grare, Romamq; orbis caput inuiſere poſſem, id tu mi-
rifica quadã benignitate præſtitiſti, perfeciſtiq;, ut Italiam
non inuiſerem ſolum (quod ipſum per ſe plauſibile erat)
ſed etiam tecum inuiſerem, homine omnium quos cœlum
tegit doctiſsimo, humaniſsimoq; (quod nondum conſtitui
quanti ſit æſtimandum). Mihi ſane pluris fuit Romæ te,
quàm Romam ipſam uidiſſe. Romæ fuiſſe, ſortis cuiuſdam

* 2 eſt in

FIGURE A.12: Maillage numéro 11 : contient des distorsions locales et des parties déchirées

Maillage 12 Image



FRANCISCVS
RABELAESVS MEDICVS
Clariß. doctißimoq;, uiro D. Ioanni Bellaio
Pariſienſi Epiſcopo, Regisq; in ſan-
ctiori conſeſſu conſiliario
S. P. D.

INGENS ille beneficiorum cumulus, quibus me nuper augendum ornandumq; putaſti Antiſtes Clariſsime, ita in memoria mea penitus inſedit: nullo ut euellimodo, aut in oblitionem diuturnitatis adduci poſſe confidam. Atq; utinam mihi tam eſſet immortalitati laudum tuarum ſatisfacere expeditum, quàm certum eſt meritam tibi gratiam uſq; perſoluere, teq; ſi nō paribus officijs (quì enim poſſem?) at iuſtis tamen honoribus, & memori mente remunerare. Nam quod maxime mihi fuit optatum iam inde ex quo in literis politioribus aliquem ſenſum habui, ut Italiam peragrare, Romamq; orbis caput inuiſere poſſem, id tu mirifica quadā benignitate præſtitiſti, perfeciſtiq;, ut Italiam non inuiſerem ſolum (quod ipſum per ſe plauſibile erat) ſed etiam tecum inuiſerem, homine omnium quos cœlum tegit doctißimo, humaniſsimoq; (quod nondum conſtitui quanti ſit æſtimandum). Mihi ſane pluris fuit Romæ te, quàm Romam ipſam uidiſſe. Romæ fuiſſe, ſortis cuiuſdam

* 2 eſt in

FIGURE A.13: Maillage numéro 12 : contient des distorsions locales