



HAL
open science

Segmentation de documents administratifs en couches couleur

Elodie Carel

► **To cite this version:**

Elodie Carel. Segmentation de documents administratifs en couches couleur. Traitement des images [eess.IV]. Université de La Rochelle, 2015. Français. NNT : 2015LAROS014 . tel-01373416

HAL Id: tel-01373416

<https://theses.hal.science/tel-01373416v1>

Submitted on 28 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ DE LA ROCHELLE

ÉCOLE DOCTORALE S2IM

LABORATOIRE : L3I

THÈSE présentée par :

Élodie CAREL

soutenue le : **8 octobre 2015**

pour obtenir le grade de : **Docteur de l'université de La Rochelle**

Discipline : **Informatique et Applications**

**Segmentation de documents administratifs
en couches couleur.**

JURY :

Rolf INGOLD

Josep LLADÓS

Rafael DUEIRE LINS

Véronique EGLIN

Vincent POULAIN D'ANDECY

Jean-Marc OGIER

Jean-Christophe BURIE

Vincent COURBOULAY

Professeur, Université de Fribourg, Rapporteur

Professeur, Université Autonome de Barcelone, Rapporteur

Professeur, Université Fédérale de Pernambuco, Examineur

Maître de conférences, HDR, INSA de Lyon, Examinatrice

Responsable Recherche et Technologies, Itesoftware, Invité

Professeur, Université de La Rochelle, Directeur de thèse

Professeur, Université de La Rochelle, co-encadrant de thèse

Maître de conférences, HDR, Université de La Rochelle, co-encadrant de thèse

Remerciements

It is good to have an end to journey toward, but it is the journey that matters in the end.

Ursula K. Le Guin

Ces travaux sont le fruit d'une collaboration avec mes encadrants et avec un partenaire industriel, mais aussi de multiples rencontres et d'échanges qui m'ont aidé à les mener à bien.

Je tiens à remercier sincèrement mes encadrants, Jean-Marc Ogier, Jean-Christophe Burie, et Vincent Courboulay, pour leur disponibilité, en particulier pendant ce sprint final, et malgré un emploi du temps chargé, pour avoir toujours su me remotiver, et m'encourager quand j'en avais besoin. Merci à Vincent Poulain d'Andecy, Saddok Kebairi et Aurélie Joseph de nous avoir accompagné dans la réalisation de ce projet. Je voudrais remercier également les membres de mon jury d'avoir accepté d'évaluer ces travaux : Rolf Ingold, Josep Lladós, Rafael Dueire Lins, Véronique Eglin et Vincent Poulain d'Andecy.

During these years first as a master student and then as a PhD student, I had the opportunity to exchange in several occasions with many people about Research. I would like to thank them all for having shared with me their valuable experience, their time and for having advised me. A special thanks to Koichi Kise from Osaka Prefecture University, as well to Josep Lladós, from the CVC of the Universitat Autònoma of Barcelona for having welcomed me in their labs during my research stays. I would also like to thank Dimosthenis Karatzas, who has supervised me, Marçal Russiñol and Hongxing who have worked with me. Sincere thanks to all the members of these two labs for their warm welcome. It has been a real pleasure to work with you. I am also grateful to Véronique Eglin for her essential help when I was deadlocked. Thank you to Do Thi Luyen for her active participation to this project during her master internship.

Merci également à tous les collègues de l'Université de La Rochelle que j'ai eu la chance de côtoyer pendant ces années de thèse, au sein du labo, de l'administration, de l'ADocs et pendant le service DCACE. Je voudrais aussi remercier mes camarades relecteurs pour leurs remarques pertinentes qui ont contribué à l'amélioration de ce manuscrit.

Une dédicace spéciale aux deux passionnés de vélo (*Raciines!*) qui ont eu cette idée un peu folle d'entraîner un certain nombre d'entre nous dans un grand périple (*Piééééton ... avec chien!*), petite pause et grand bol d'air frais bienvenus pendant la Rédaction (*Chicaaanes!*) J'en profite pour remercier chaleureusement tous les amis qui ont apporté leur bonne humeur dans les bons et les mauvais moments, et qui m'ont supporté et soutenu tout au long de ces années de thèse.

Enfin, un grand merci à ma famille pour son soutien et en particulier à mes parents grâce à qui j'ai pu poursuivre ces études sereinement. Je conclurai ces remerciements par une dernière contribution à un jeu commencé en 2012 sur un tableau de la 121 bis, mais qui n'est pas dédiée uniquement à ses seuls habitants. Merci à tous!



Table des matières

1	Introduction générale	1
1.1	Introduction à la dématérialisation	1
1.2	Problématique et enjeux	3
1.3	Extraction de couches couleur	5
1.4	Organisation du mémoire	8
I	Etat-de-l'art	11
2	Analyse de document	13
2.1	Introduction	13
2.2	Caractérisation des images de documents	14
2.3	Segmentation et applications	15
2.3.1	Stratégies d'analyse	19
2.3.2	Techniques usuelles de segmentation	21
2.3.3	Choix de la méthode	30
2.4	Problèmes liés à la binarisation	31
2.5	Conclusion	33
3	Traitement de l'information couleur	37
3.1	Introduction	37
3.2	Perception de la couleur	38
3.2.1	Phénomène physique	38
3.2.2	Vision humaine	39
3.2.3	Mécanismes de la vision	40
3.3	Espaces couleurs numériques	41
3.3.1	Systèmes de primaires	42
3.3.2	Systèmes luminance-chrominance	42
3.3.3	Systèmes perceptuels	43
3.3.4	Systèmes d'axes indépendants	43
3.3.5	Systèmes hybrides	43
3.3.6	Modèles d'apparence couleur	43
3.3.7	Choix d'un espace couleur	44
3.4	Traitement de l'information colorimétrique	44
3.4.1	Stratégies	44
3.4.2	Segmentation couleur	46
3.5	Quantification couleur	52
3.5.1	Approches classiques de réduction couleur	52
3.5.2	Approches de nommage de couleurs	53
3.6	Analyse de documents couleur	57

3.6.1	Caractéristiques spécifiques aux documents couleur	57
3.6.2	Segmentation de documents administratifs couleur	58
3.7	Conclusion	68
II	Segmentation d'images de documents en couches couleur	71
4	Jeu de données pour validations expérimentales	73
4.1	Introduction	73
4.2	Evaluation de la segmentation en analyse de documents	74
4.3	Base de documents réels	76
4.3.1	Données	76
4.3.2	Evaluation sur des données réelles	77
4.4	Constitution d'une base de vérité-terrain de documents synthétiques	78
4.4.1	Données	78
4.4.2	Simuler les dégradations causées par le bruit	80
4.4.3	Mesures d'évaluation	82
4.5	Vers un modèle de dégradation des caractères applicable aux images de documents couleur	83
4.5.1	Introduction	83
4.5.2	Modèle de dégradation des caractères	84
4.5.3	Résultats et discussion	91
4.5.4	Conclusion sur le modèle de bruit	91
4.6	Conclusion sur le jeu de données	94
5	Segmentation colorimétrique	97
5.1	Introduction	97
5.2	Contributions	98
5.2.1	Pré-traitements	98
5.2.2	Approche de quantification basée sur des mesures psychovi- suelles	99
5.2.3	Approche basée sur une classification hiérarchique descendante	100
5.3	Expérimentations	105
5.3.1	Approche de quantification basée sur des mesures psychovi- suelles	105
5.3.2	Approche basée sur une classification hiérarchique descendante	105
5.4	Résultats et discussion	106
5.4.1	Documents réels	109
5.4.2	Images synthétiques	110
5.5	Conclusion	114
6	Segmentation spatio-colorimétrique	117
6.1	Introduction	117
6.2	Contributions	118
6.2.1	Pré-traitements	118

6.2.2	SLIC original	119
6.2.3	SLIC Adaptatif	121
6.2.4	Procédure de regroupement	124
6.3	Expérimentations	126
6.4	Résultats et discussion	127
6.4.1	Documents réels	127
6.4.2	Images synthétiques	128
6.5	Conclusion	131
7	Analyse multi-résolution	133
7.1	Introduction	133
7.2	Contributions	134
7.2.1	Pré-traitements	135
7.2.2	Procédure globale	136
7.2.3	Procédure de fusion	138
7.2.4	Procédure d'étiquetage	139
7.3	Expérimentations	142
7.4	Résultats et discussion	144
7.4.1	Documents réels	144
7.4.2	Images synthétiques	145
7.5	Conclusion	152
8	Conclusion générale	153
8.1	Conclusion	153
8.2	Perspectives scientifiques	156
8.2.1	Affectation des pixels à plusieurs couches	156
8.2.2	Interaction avec l'utilisateur	157
8.2.3	Apprentissage et paramétrage automatique	157
8.2.4	Dégradés de couleurs et textures	158
8.3	Applications industrielles futures possibles	158
	Bibliographie	161
	A Techniques de segmentation bas-niveaux	169
	B Exemples de documents administratifs réels	171
	C Publications	175

Table des figures

1.2	Illustration d'une solution de GED.	3
1.3	La chaîne de dématérialisation classique : traitement des images en noir et blanc.	4
1.6	Bruit apparaissant sur une image réelle : apparition de fausses couleurs, artéfacts de compression.	7
2.1	Exemple de document administratif réel complexe du fait de son hétérogénéité en termes de contenus, de structures, et de couleurs.	15
2.4	Niveaux d'abstraction auxquels une image peut être vue. Crédits : [Dombre 2003] (Figure 3.1)	20
2.5	Profils de projection. Crédits : [Kise 2014] (Figure 5.10)	22
2.6	Recursive XY cut. Crédits : [Kise 2014] (Figure 5.11)	22
2.11	Pyramide par réduction. Crédits : [Dombre 2003] (Figure 4.8)	29
2.12	Approche proposée par [Barlas 2014] pour la segmentation en blocs. Crédits : [Barlas 2014] (Figure 5)	30
2.13	Effets du choix de différentes valeurs de seuil de binarisation sur une image, et le résultat fourni par un outil de reconnaissance de texte de type ROC. Crédits : [Dengel 2014] (Figure 6.7)	31
3.1	Dispersion de la lumière à travers un prisme.	38
3.2	Spectre du domaine visible.	39
3.3	Mécanisme de la vision humaine.	40
3.5	Les principales familles d'espaces couleurs. Crédits : [Vandenbroucke 2000] (Figure 2.12).	45
3.7	Structure hiérarchique de type hexagonal de l'algorithme du CSC. Crédits : [Rehrmann 1998] (Figure 1)	51
3.8	Images segmentées par la méthode SLIC en superpixels de tailles 64, 256 et 1 024 (approximativement). Crédits : [Achanta 2012](Figure 1)	52
3.9	Découpage du cube RVB par la méthode <i>Median-Cut</i> . Crédits : [Montagne 2005] (Figure 6.21)	53
3.10	Schéma des conditions d'expérimentation de nommage de couleurs. Les expériences ont été réalisées dans des conditions contrôlées, dans une chambre noire, pour s'assurer que les échantillons ne seraient pas illuminés par d'autres sources que celle placée en haut du dispositif. Crédits : [Benavente 2006]	55
3.11	Partition de l'espace couleur RGB par une approche nommage de couleur. Ce schéma a été obtenu grâce à des données obtenues à partir d'un outil d'annotation mis en ligne sur l'Internet. Crédits : [Munroe 2010]	56

3.12	Lorsque des éléments se chevauchent, des effets de mélange créent de nouvelles couleurs.	58
3.13	Bruit de saturation. Crédits : [Ouji 2012](Figure I.1)	59
3.14	Bruit de teinte. Crédits : [Ouji 2012] (Figure I.2)	59
3.18	Couches issues de la segmentation colorimétrique proposée par [Ouji 2012]. Crédits : [Ouji 2012] (Figure I.4)	64
3.21	Détection des rectangles. Crédits : [Hammami 2014] (Figure 4)	68
4.1	Vérité-terrain constituée de régions de forme complexe qui ont été extraites grâce au logiciel <i>Aletheia</i> . Crédits : [Antonacopoulos 2009a] (Figure 4)	75
4.2	Images synthétiques et couches associées aux couleurs rose, vert et marron.	79
4.6	Processus complet de dégradation des caractères appliqué sur des images de documents couleur. Crédits : Do Thi Luyen	86
4.7	Exemple de sélection de germes (points rouges), sur trois couches couleur extraites à partir d'une image synthétique. Crédits : Do Thi Luyen	88
4.9	Zone de dégradation elliptique. Crédits : [Kieu 2012]	90
5.3	Classification hiérarchique descendante.	102
5.5	La numérisation et la compression introduisent du bruit sur les images sous forme d'artéfacts et de fausses couleurs qui apparaissent notam- ment près des contours.	110
6.3	Fenêtre de recherche du SLIC. Crédits : [Achanta 2012] (Figure 2b)	121
6.4	SLIC Adaptatif - initialisation des centres.	123
7.2	Analyse multi-résolution.	137
7.7	Fenêtre de recherche utilisée pendant la phase d'étiquetage.	142
A.1	Taxonomie des algorithmes de segmentation bas-niveaux. Crédits : [Vantaram 2012]	170

Liste des tableaux

2.1	Les éléments de base d'un document et leurs principales propriétés. . .	17
2.2	Caractéristiques des différentes surcharges d'un document.	18
4.1	Ensemble de documents fournis par notre partenaire industriel. . . .	77
5.1	Performance de la quantification sur la base d'images synthétiques. . .	109
5.2	Performance de la classification hiérarchique sur la base d'images synthétiques.	114
6.1	Mesure de performance de l'approche basée sur le SLIC original et sur le SLIC adaptatif appliqués sur la base d'images synthétiques. . .	131
6.2	Performance de la classification hiérarchique sur la base d'images synthétiques.	131
7.1	Valeurs des paramètres utilisés pendant l'étape d'étiquetage appliquée lors de l'analyse multi-résolution pour chaque niveau de résolution.	144
7.2	Valeurs des paramètres utilisés lors de l'application du SLIC adaptatif sur l'image en haute résolution dans un but de comparaison. L'étiquetage s'est déroulé en trois étapes dans le but de faire varier la taille de la fenêtre de recherche.	144
7.3	Performance de l'extraction des couches couleur extraites à partir d'une segmentation basée sur le SLIC adaptatif suivie par l'application de deux approches de fusion : une approche basée sur la représentativité globales des couleurs (V1), et une autre basée sur l'étude des couleurs présentes dans le voisinage de chaque pixel (V2).	145
7.4	Comparaison de l'approche basée sur une analyse multi-résolution, avec un processus appliqué à une résolution unique.	151
7.5	Nombre de labels extraits à chaque étape du processus pour le document réel, figure 7.11a.	151
8.1	Tableau récapitulatif des performances des approches évaluées lors de ces travaux de recherche.	156

Introduction générale

Sommaire

1.1	Introduction à la dématérialisation	1
1.2	Problématique et enjeux	3
1.3	Extraction de couches couleur	5
1.4	Organisation du mémoire	8

1.1 Introduction à la dématérialisation

Depuis plusieurs décennies, la dématérialisation, processus qui permet de passer d'un support physique à une copie numérique, est devenue un sujet d'actualité. En effet, les enjeux culturels et économiques sont importants. Entre autres projets, la numérisation de fonds d'archives est de plus en plus fréquente dans un but de préservation et de diffusion. Par exemple, une partie des registres paroissiaux et d'Etat-civil français a été mise à disposition du grand public à travers le moteur de recherche RECEL (Registres de l'Etat-Civil En Ligne)¹. Gallica², la bibliothèque numérique de la BnF (Bibliothèque nationale de France) propose un accès gratuit à différents documents (livres, presse, cartes et plans, etc). Les documents papiers étant souvent fragiles, l'utilisation de copies numériques en facilite la manipulation.

Plusieurs grands projets impliquant différents laboratoires de recherche français et internationaux ont ainsi vu le jour. Leur but est de travailler sur des problématiques de valorisation de documents historiques. De 2007 à 2010, le projet NAVIDOMASS³ a eu pour objectif la création d'outils permettant d'enrichir ces images avec des données supplémentaires liées à leurs contenus. L'enjeu était de les classer et de les étudier à travers des bibliothèques numériques. Entre 2011 et 2014, DIGIDOC⁴ s'est concentré la mise au point de scanner "intelligents" s'adaptant aux caractéristiques de documents anciens et fragiles pour améliorer la qualité de la numérisation. Entre 2008 et 2012, le projet IMPACT⁵ a conduit à l'implémentation de solutions

1. <http://recel.geonames.org/>

2. <http://gallica.bnf.fr/>

3. Navigation dans des Masses de Documents (NAVIGATION into DOcument MASSes en anglais) : <http://navidomass.univ-lr.fr>

4. Document Image diGitisation with Interactive DescriptiOn Capability en anglais : <http://13i.univ-larochelle.fr/DIGIDOC>

5. IMProving ACces to Text en anglais : <http://www.impact-project.eu>

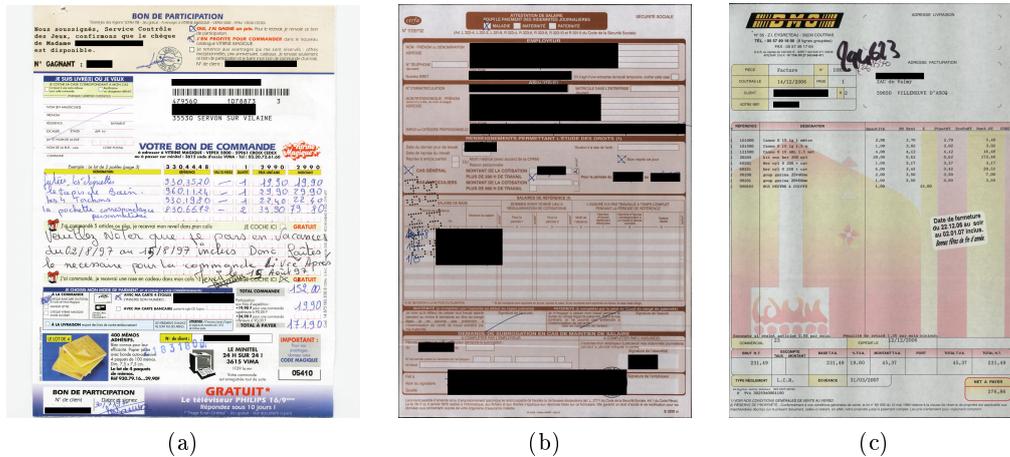


FIGURE 1.1 – Exemple de documents administratifs : factures et bons de commande.

de reconnaissance automatique de texte dans des documents historiques. Enfin depuis 2013, le projet transcriptorium⁶, toujours en cours, cherche à développer des systèmes de transcription d’images de documents historiques manuscrits pour aider à leur indexation et à la recherche de contenu.

Dans ce mémoire, nous nous intéresserons à la dématérialisation de documents administratifs, qui, même s’ils ont des caractéristiques similaires à celles des documents historiques, diffèrent en termes de contenu, et de contraintes de traitement. Ces travaux s’inscrivent dans un contexte industriel fort avec un partenaire privé et pour ambition d’avoir, je l’espère, un impact économique. Les enjeux seront donc moins la préservation ou la restauration de ces documents que la mise en place de solutions efficaces et applicables dans ce contexte particulier.

De nos jours, l’administration publique et les industriels doivent traiter quotidiennement de gros volumes de documents papiers de toutes sortes (Figure 1.1). Une étude réalisée par EY et Dimo Gestion [EY 2014] montre l’impact de la dématérialisation dans les entreprises. Le temps de traitement moyen d’une facture fournisseur papier serait d’environ 14 jours et coûterait en moyenne 15 euros. Le processus implique des opérations de tri, de validation, d’archivage, etc. Tout l’intérêt du processus de dématérialisation est alors la mise en place d’outils capables d’extraire automatiquement des données métiers. Dans ce contexte, nous parlerons de systèmes de LAD/RAD (Lecture/Reconnaissance Automatique de Documents). Les informations textuelles sont “lues” par des systèmes de reconnaissance de texte à partir d’une image préalablement numérisée. Il existe ainsi les techniques de ROC (Reconnaissance Optique de Caractères, OCR en anglais), ou de RIC (Reconnaissance Intelligente de Caractères, ICR en anglais), plus perfectionnées. Pour finir,

6. <http://transcriptorium.eu>

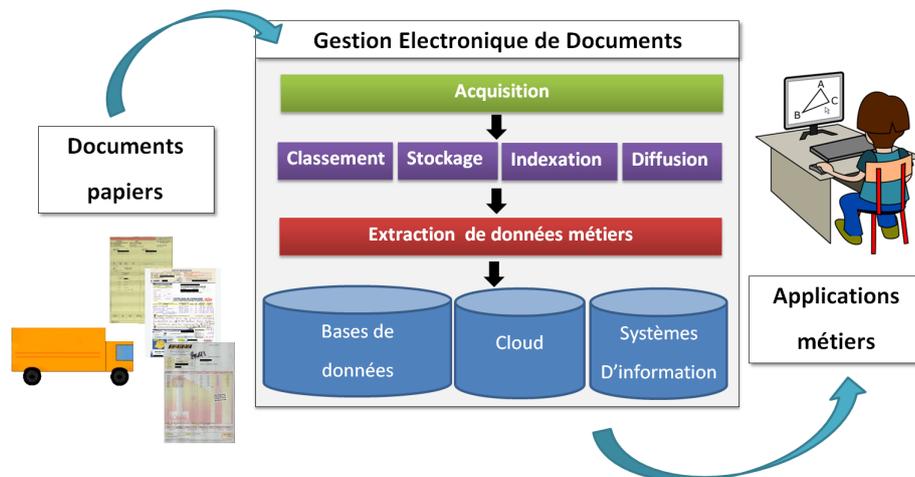


FIGURE 1.2 – Illustration d'une solution de GED.

toutes les données ainsi extraites sont injectées dans des bases de données et vont alimenter des systèmes d'informations.

L'automatisation de certaines tâches réduit fortement les coûts en temps et en ressources humaines jusque-là nécessaires. D'une part, les erreurs entraînées par la saisie manuelle sont limitées et l'ensemble des processus est optimisé. D'autre part, l'utilisation de copies numériques centralisées, et surtout traçables, offre un meilleur suivi des documents et des communications, aussi bien au sein de l'entreprise qu'avec l'extérieur. Les délais de réponse sont ainsi beaucoup plus courts. Enfin, d'un point de vue plus écologique, la numérisation des documents n'entraîne pas forcément une suppression des supports papiers, mais permet une nette diminution du nombre de copies. L'ensemble des applications impliquées dans la chaîne de dématérialisation constitue une solution de GED (Gestion Electronique des Documents). La figure 1.2 illustre les fonctionnalités d'une GED.

1.2 Problématique et enjeux

Les industriels peuvent être amenés à traiter des documents de tous types qui sont généralement très riches en termes de contenus, de structures, et de couleurs (Figure 1.1). Le processus historique implique une étape de simplification pour réduire la complexité des traitements et les coûts de stockage. Par conséquent, les images couleur sont généralement numérisées en noir et blanc grâce à une étape de binarisation, ou éventuellement converties en niveaux de gris. Elles sont ensuite analysées pour en extraire des zones d'intérêt qui sont transmises à un système de reconnaissance de texte (Figure 1.3) afin d'en extraire les informations utiles aux applications métiers.

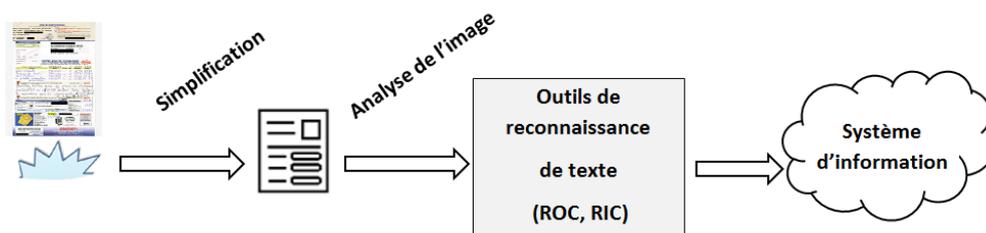
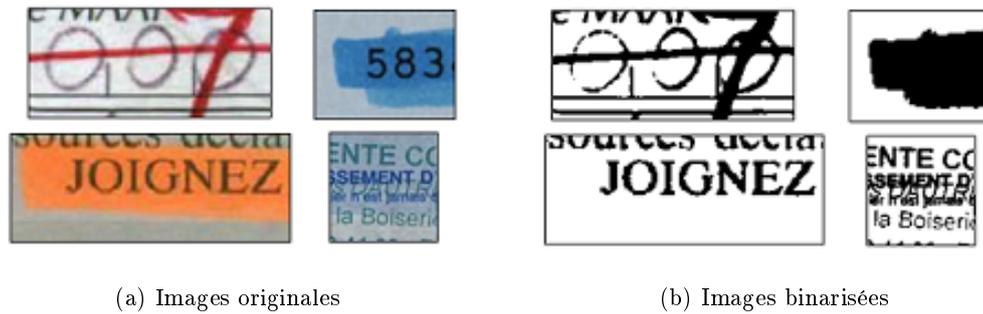


FIGURE 1.3 – La chaîne de dématérialisation classique : traitement des images en noir et blanc.

L'approche classique génère des problèmes du fait des erreurs liées à l'étape de simplification (Figure 1.4). La performance de reconnaissance est dégradée en cas d'échec de la segmentation, lors de la superposition d'objets graphiques et textuels notamment. Par exemple, des éléments de couleurs différentes ne pourront plus être différenciés. Certaines informations textuelles peuvent disparaître comme dans le cas d'un surlignement. De plus, la couleur est généralement porteuse de sens sur ce type d'images : surlignement, correction... Or ici, toute l'information sémantique apportée par la couleur disparaît du fait de l'étape de binarisation préalable. Des post-traitements peuvent éventuellement corriger certains problèmes. Cependant, ils nécessitent souvent une phase de paramétrage, et d'interaction avec un utilisateur. En revanche, le traitement direct de l'information colorimétrique sans binarisation pourrait améliorer le processus actuel de dématérialisation.

Certaines approches usuelles, spécifiques à l'étude de documents en noir et blanc, ont été transposées au traitement en niveaux de gris. L'amélioration des technologies et l'augmentation des capacités de stockage ont levé les verrous scientifiques qui empêchaient jusque là la manipulation des images couleur dans un contexte industriel. Cependant, malgré un intérêt croissant, encore peu de travaux s'intéressent à l'utilisation directe de la couleur pour la segmentation de documents. La plupart des méthodes sont dédiées à des tâches particulières, et à des catégories spécifiques d'images, partageant les mêmes caractéristiques structurelles. Elles sont aussi souvent basées sur l'utilisation de modèles permettant l'introduction de connaissances *a priori*.

Ces travaux de recherches sont focalisés sur la segmentation bas-niveau des documents administratifs couleur. Ils se situent donc avant l'étape de reconnaissance de texte dans la chaîne de dématérialisation (Figure 1.3). Ils ont pour objectif d'utiliser l'information colorimétrique pour améliorer les résultats de la segmentation. Du fait du contexte industriel, notre approche devra être générique et non supervisée, afin de pouvoir analyser tous types de documents administratifs couleur sans paramétrage et sans interaction avec l'utilisateur. Le contenu des documents et leurs couleurs étant très variables, notre approche devra aussi être adaptative.



(a) Images originales

(b) Images binarisées

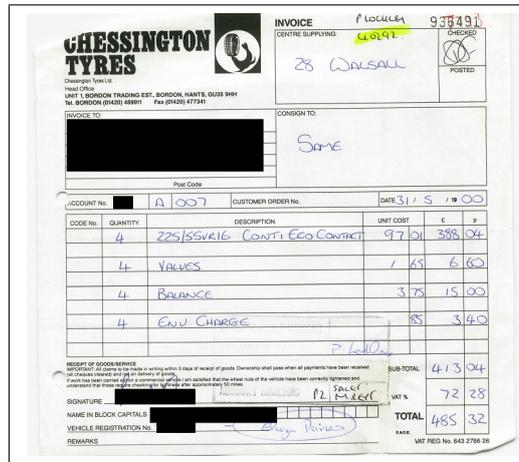
FIGURE 1.4 – Erreurs liées à l'étape de binarisation.

1.3 Extraction de couches couleur

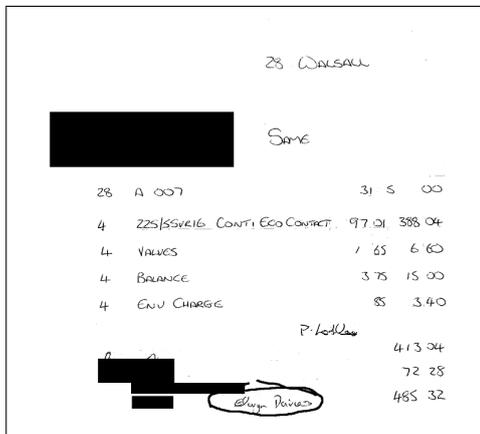
Dans ces travaux, nous proposons la segmentation des documents en couches couleur comme une alternative à l'étape de binarisation qui est source d'erreurs. Pour pouvoir manipuler facilement l'information colorimétrique, le nombre de couleurs présentes sur l'image doit être réduit en regroupant tous les éléments ayant des teintes et des niveaux de saturation très proches. C'est à dire que seules les couleurs dominantes globalement ou localement, lorsqu'elles constituent des régions connexes compactes, seront conservées. Cependant, la segmentation logique des régions du document doit aussi être respectée (ex : caractères, mots, graphiques) même si de petites variations colorimétriques apparaissent au niveau pixel. Une couche couleur cohérente peut donc être définie comme un ensemble de pixels ayant à la fois une apparence colorimétrique très proche et présentant une unité logique ou sémantique.

Afin de respecter la représentation historique, chaque couche est associée à une image binaire. Pour une couche donnée, un pixel apparaît noir sur un fond blanc lorsqu'il appartient à cette couche. Les industriels ont ainsi la possibilité de réinjecter ces images binaires dans la chaîne classique de dématérialisation sans avoir à remplacer les traitements traditionnellement utilisés. De plus, la segmentation en couches couleur permet l'ajout d'informations supplémentaires concernant les caractéristiques colorimétriques de l'image. Il devient alors possible de décrire les éléments de l'image en fonction de leur couleur, de différencier ceux ayant des couleurs différentes, ou bien de ne traiter que certaines couches spécifiques. Des exemples de couches couleur souhaitées sont présentées sur la figure 1.5. Elles ont été obtenues à partir d'une annotation manuelle sur un document réel relativement simple en termes de structure et de caractéristiques couleur.

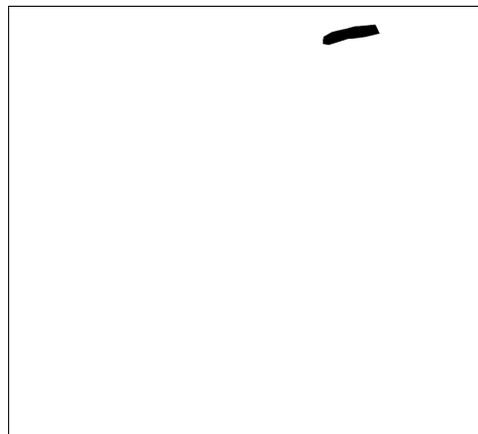
La principale difficulté de ces travaux provient de la complexité et de la densité des informations à traiter. Considérons tout d'abord la couleur, la numérisation et la compression introduisent du bruit sous forme d'artéfacts et de fausses couleurs qui dégradent l'homogénéité des éléments de l'image (Figure 1.6). Numériquement,



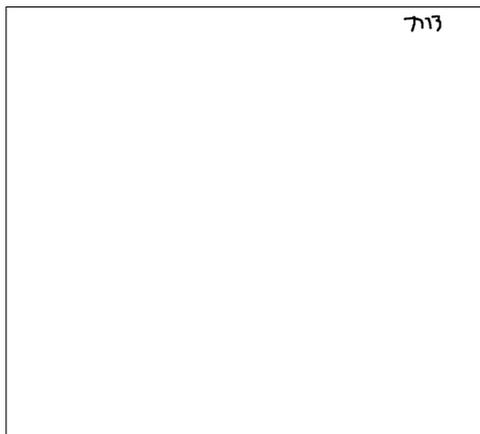
(a) Image originale



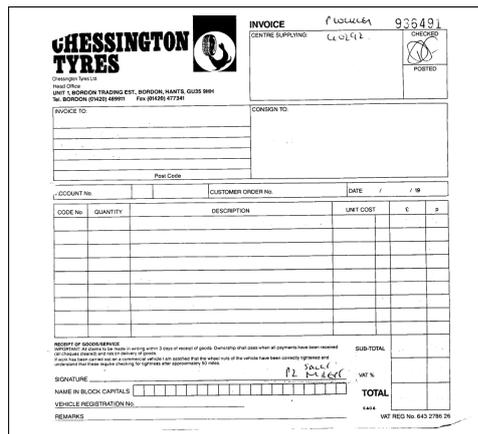
(b) Couche bleue



(c) Couche jaune



(d) Couche rouge



(e) Couche noire

FIGURE 1.5 – Exemple de quatre couches couleur annotées manuellement sur un document administratif réel. Afin de respecter la représentation traditionnelle, pour une couche donnée, un pixel apparaît noir sur un fond blanc lorsqu'il appartient à cette couche. La couche de fond n'est pas affichée ici et les données personnelles ont été masquées.



FIGURE 1.6 – Bruit apparaissant sur une image réelle : apparition de fausses couleurs, artefacts de compression.

l'information colorimétrique s'exprime dans un espace couleur dans lequel les couleurs dominantes peuvent théoriquement être extraites en recherchant les nuages de points. En réalité, ces points formeront plutôt des trainées qui peuvent être difficilement exploitables, et ce d'autant plus avec des documents complexes, contenant par exemple des dégradés de couleurs, et des nuances très proches visuellement (Figure 1.7). L'une des problématiques consistera donc à trouver automatiquement le nombre de couleurs présentes dans l'image.

Notre objectif étant l'extraction de n'importe quelle structure présente sur l'image et non pas uniquement la reconnaissance des éléments textuels, nous sommes alors confrontés à un problème relevé par le paradoxe de Sayre [Sayre 1973]⁷. Ce paradoxe met en évidence l'interdépendance des problématiques de reconnaissance et de segmentation. Pour pouvoir être exploitées, nos couches couleur doivent conserver une certaine unité sémantique par rapport aux régions réelles de l'image. La prise de décision relative à l'appartenance à une couche couleur ou à une autre doit donc être suffisamment flexible pour regrouper les éléments en prenant en compte à la fois l'information couleur mais aussi l'identification de la structure logique du document. Par conséquent, la difficulté majeure introduite par le contexte de nos travaux est la segmentation de contenus très hétérogènes sans l'aide de modèles et sans l'ajout de connaissances *a priori*.

Finalement, le cadre de nos travaux peut être défini comme suit :

- contraintes industrielles fortes ;
- base de données génériques ;
- segmentation non-supervisée ;
- multi-contenu ;
- détection automatique du nombre de couleurs ;
- conservation de l'information sémantique attachée à la couleur.

7. Paradoxe de Sayre : il faut segmenter pour reconnaître, et il faut reconnaître pour segmenter [Sayre 1973].

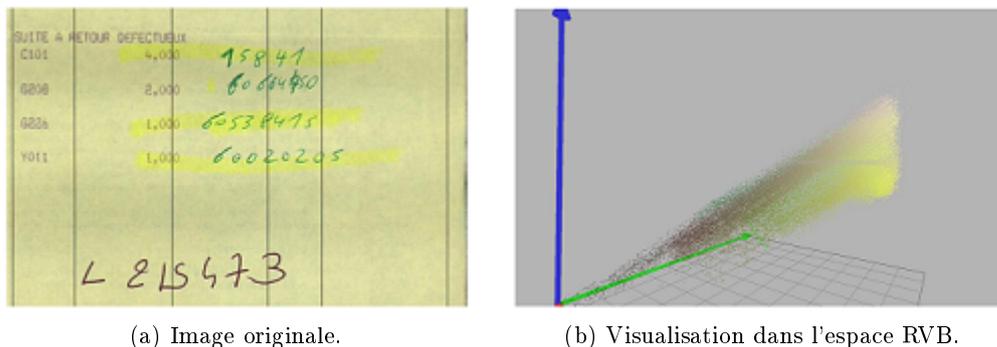


FIGURE 1.7 – Difficultés liées à la visualisation des couleurs dominantes dans un espace couleur donné.

1.4 Organisation du mémoire

Ce mémoire est organisé de la manière suivante. La première partie sera dédiée à l'état-de-l'art qui portera, d'une part, sur les approches traditionnellement appliquées en analyse de documents, et d'autre part, sur le traitement de l'information couleur. Nous introduirons les grands mécanismes de gestion de la couleur en décrivant tout d'abord des approches appliquées dans le domaine de la vision par ordinateur. Cependant, les images de documents sont des images particulières qui sont faites spécifiquement pour transmettre de l'information. Par conséquent, elles diffèrent des images naturelles par leur caractère beaucoup plus structuré sur lequel les approches d'analyse peuvent s'appuyer. La deuxième partie sera consacrée à nos travaux de recherches concernant la segmentation d'images de documents en couches couleur.

Nous commencerons, dans le chapitre 2, par caractériser les images de documents, en mettant en évidence leur complexité en termes de structures, et l'hétérogénéité de leurs contenus. Du fait du contexte industriel, les méthodes traditionnellement employées simplifient les images de documents. L'objectif étant la plupart du temps une classification fond/premier-plan, ces images sont alors analysées en noir et blanc. Nous verrons en quoi ces méthodes sont restrictives et peuvent être sources d'erreurs lors de l'étape de segmentation.

Le chapitre 3 sera consacré au traitement de l'information couleur. Nous introduirons les principaux mécanismes qui ont été à l'origine des systèmes de codage de l'information numérique couleur. Nous verrons ensuite quelles sont les stratégies de segmentation impliquant la gestion de la couleur. Puis, nous orienterons plus particulièrement cet état-de-l'art sur l'étude des systèmes de réduction couleur qui en sont une composante essentielle. Pour terminer, nous ferons un inventaire des approches traitant de la segmentation de documents couleur, en mettant en évidence les problèmes spécifiques liées au traitement de la couleur.

Nous passerons ensuite à la description de nos travaux. Les différentes approches présentées ici ayant été implémentées de manière linéaire, nous présenterons avant toutes choses les données et les méthodes d'évaluation dans le chapitre 4. Nous avons expérimenté trois types d'approches : une approche colorimétrique (chapitre 5), une approche spatio-colorimétriques (chapitre 6), et une approche basée sur une analyse multi-résolution (chapitre 7).

Première partie

Etat-de-l'art

Analyse de document

Sommaire

2.1	Introduction	13
2.2	Caractérisation des images de documents	14
2.3	Segmentation et applications	15
2.3.1	Stratégies d'analyse	19
2.3.2	Techniques usuelles de segmentation	21
2.3.3	Choix de la méthode	30
2.4	Problèmes liés à la binarisation	31
2.5	Conclusion	33

2.1 Introduction

Les administrations publiques et les industriels sont amenés à traiter des documents administratifs de toutes sortes : factures, lettres, formulaires, etc. Avec le développement important des technologies, de nombreuses solutions de dématérialisation ont vu le jour. L'intérêt n'est plus seulement de numériser un document dans un but d'archivage, mais plutôt d'être capable d'en extraire des données métiers de manière automatique, pour optimiser la gestion de l'information d'entreprise. Ces données alimentent des systèmes d'informations, apportant ainsi une plus-value significative à l'ensemble des processus liés au cycle de vie des documents.

Bien que généralement structurées par des règles plus ou moins implicites, les images de documents administratifs sont cependant complexes à analyser, car très hétérogènes en terme de contenus. Les solutions proposées sont par conséquent souvent dédiées à une catégorie spécifique d'images, et à l'extraction d'un type d'information particulier. De plus, elles peuvent nécessiter une étape de paramétrage, ou l'injection de connaissance *a priori*, impliquant éventuellement une interaction avec un utilisateur.

A travers ce chapitre, nous réaliserons un bref état-de-l'art des principales techniques d'analyse de documents, et plus particulièrement des documents administratifs. Nous commencerons par caractériser ce type d'images en mettant en évidence leur complexité. Puis, nous présenterons les techniques historiques de segmentation qui ont été développées spécifiquement pour le traitement des images en noir et

blanc. En effet, le contexte industriel nécessite une simplification des images qui passe la plupart du temps par une étape de séparation fond/premier-plan. Notre objectif est ici de présenter les processus d'analyse traditionnels et de montrer en quoi cette étape de binarisation peut être source d'erreurs. Le chapitre 3 sera dédié aux traitements spécifiques de l'information couleur.

2.2 Caractérisation des images de documents

Au cours de ces travaux nous nous sommes intéressés aux documents administratifs de tous types. Ceux-ci sont bien sur très hétérogènes en termes de contenus, de structures, et de couleurs. Cependant, un document étant fait spécifiquement pour transmettre de l'information, cette dernière doit par conséquent être facilement identifiable et lisible. Il est donc possible d'extraire un certain nombre de caractéristiques communes que nous allons maintenant présenter.

Tout d'abord, un document est construit selon un ensemble de règles explicites ou implicites qui codifient le placement des différents éléments. Nous parlerons alors d'images *structurées*, voire *semi-structurées*. Par exemple, un bloc d'adresse se situe principalement dans la partie supérieure d'une lettre. Un tableau est souvent placé dans la partie centrale d'une facture. Ou encore, les éléments ont généralement une orientation horizontale. Les images de documents sont néanmoins complexes du fait de la densité d'information qu'elles contiennent, et de la richesse de leur contenu (Figure 2.1).

Nous pouvons classer ce contenu en deux catégories :

- les éléments de base, généralement observés sur les documents administratifs. Ces informations sont liées au support (tableau 2.1) ;
- les surcharges, qui correspondent le plus souvent à l'ajout d'information de la part d'un utilisateur du document (tableau 2.2).

Éléments de base : Ce sont globalement les mêmes pour tous les documents d'une même famille (Figure 2.2). Leurs caractéristiques et leur position sur le document sont généralement bien identifiées et relativement homogènes. Leur orientation est souvent rectiligne. Elles ne présentent que peu de variations. Le tableau 2.1 synthétise leurs principales propriétés.

Surcharges : Elles sont pour la plupart propres à un document donné (Figure 2.3). Leur position et leur orientation sont aléatoires, et peuvent présenter des variations importantes. Elles sont souvent placées dans les marges du document, dans des zones dans lesquelles elles sont facilement visibles. Elles se superposent parfois à d'autres éléments de l'image. Leur couleur n'est pas homogène et peut être dégradée par des effets de mélange de teinte en cas de chevauchements. Le tableau 2.2 synthétise leurs principales propriétés.

BON DE PARTICIPATION
 "Grand jeu des régions" (28/01/98) - Jeu gratuit - A renvoyer à VITRINE MAGIQUE - VERTIX 2000 - 5961 CREUX CEDEX.

Nous soussignés, Service Contrôle des Jeux, confirmons que le chèque de Madame [redacted] est disponible.

N° GAGNANT : [redacted]

JE SUIS LIVRÉ(E) OU JE VEUX
 JE COCHE LA CASE CORRESPONDANT A MON CAS
 J'ai une adresse fixe
 Je change de domicile
 [redacted] - [redacted] - [redacted]
 NOM EN MAJUSCULES
 PRÉNOM
 RESIDENCE - ÉTAGE - BÂTIMENT
 ESCALIER - ÉTAGE - APP. N°
 N° ET NOM DE LA RUE
 NOM DE LA RUE - N°
 COMMUNE - CODE POSTAL

VOTRE BON DE COMMANDE
 à adresser à VITRINE MAGIQUE - VERTIX 2000 - 5961 CREUX CEDEX
 ou à passer sur minitel : 3615 code d'accès VIMA - Tél : 03.20.72.61.00

Exemple - (le lot de 2 pochettes (page 3))	3 3 0 4 4 4 8	1	2 9 9 0	2 9 9 0
DESCRIPTION	RÉFÉRENCE	QUANTITÉ	PREX UNITAIRE	MONTANT
lutes les digitales	930.35.20	1	19.90	19.90
le tapis de bain	930.11.24	1	29.90	29.90
les 4 Tachans	930.19.80	1	22.40	22.40
la pochette correspondance personnalisée	930.66.73	2	39.90	79.80

JE COCHE ICI GRATUIT
 J'ai commandé 5 articles ou plus, je recevrai mon relevé dans mon colis.

Je tiens à noter que je pars en vacances du 02/8/97 au 15/8/97 inclus. Donc faites le nécessaire pour la commande à l'ère après le 15 Août 97.

JE COCHE ICI GRATUIT
 J'ai commandé, je recevrai une trace en cadeau dans mon colis.

RECHERCHER MON MODÈLE DE PAIEMENT ET COCHER LA CASE CORRESPONDANTE

A LA COMMANDE
 AVEC MA CARTE BANCAIRE (indiquer son numéro) - Participer sans frais de disposition : 19.90 F pour une commande supérieure à 100.00 F
 AVEC MON CHÉQUE - Participer sans frais de disposition : 19.90 F pour une commande inférieure à 100.00 F

A LA LIVRAISON (à cocher si vous ne pouvez pas recevoir votre commande)
 À domicile - Participer sans frais de disposition : 19.90 F pour une commande inférieure à 100.00 F
 À la poste - Participer sans frais de disposition : 19.90 F pour une commande inférieure à 100.00 F

LE LOT DE 4 400 MÉMOS ADHÉSIFS.
 Bien utiles pour leur utilisation. Papier gelé avec bande auto-adhésive. 4 pochettes de 100 mémos. Dim. 7,5 x 7,5 cm. Le lot de 4 pochettes de mémos. Réf 930.79.16...29.90F

N° de client : [redacted]

LE MINITEL 24 H SUR 24 ! 3615 VIMA
 1024 n°
 Votre commande est enregistrée tout de suite.

IMPORTANT :
 Pour vos commandes, écrivez votre CODE MAGIQUE 05410

BON DE PARTICIPATION
 N° de client [redacted] Date et signature [redacted]

GRATUIT*
 Le téléviseur PHILIPS 16/9***
 Répondez sous 10 jours !
 *Tirage à l'échelle de la page 3. Voir document de participation.

FIGURE 2.1 – Exemple de document administratif réel complexe du fait de son hétérogénéité en termes de contenus, de structures, et de couleurs.

La *structure physique* d'un document décrit l'organisation des différents éléments sur ce document. Sa *structure logique* est liée à la sémantique associée à son contenu. Prenons l'exemple d'un caractère. Un caractère physique correspond au symbole lui-même (ex : tracé, caractéristiques géométriques, etc), alors que le caractère logique fait directement référence à sa signification. Un même contenu logique peut avoir une organisation physique suivant le médium utilisé.

Dans la section 2.3, nous décrivons les approches de segmentation permettant l'extraction de ces différentes structures.

2.3 Segmentation et applications

I stand at the window and see a house, trees, sky. Theoretically I might say there were 327 brightness and nuances of colour. Do I have "327"? No. I have sky, house, and trees.

Max Wertheimer

La segmentation consiste à partitionner une image en régions exploitables pour un objectif donné. En fonction de la granularité souhaitée, des informations différentes seront extraites à plusieurs niveaux grâce à des processus de regroupement.

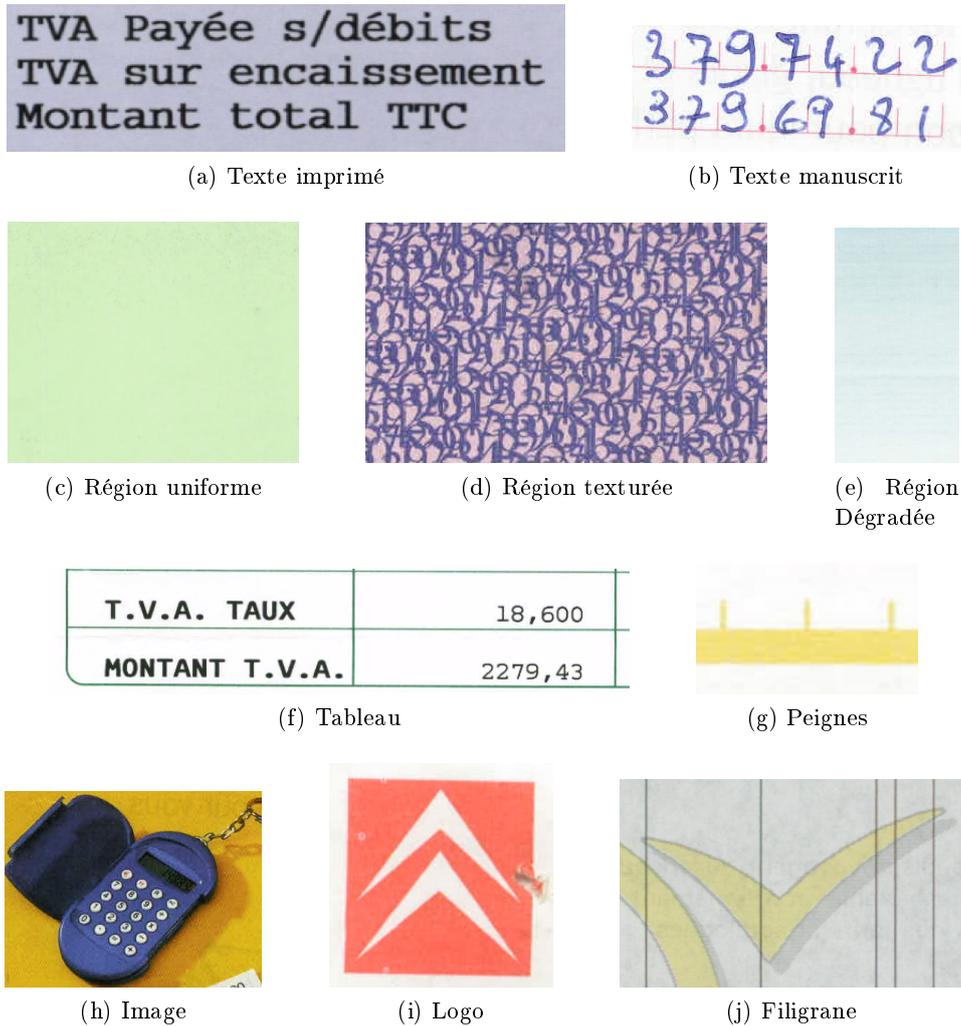


FIGURE 2.2 – Différents éléments basiques contenus par les documents administratifs.

La figure 2.4 présente les différents niveaux d'abstraction auxquels une image peut être vue. Segmenter est bien souvent un problème difficile. Le champ des techniques est très vaste. Les stratégies mises en place dépendent des images et de contraintes liées au domaine d'application.

Du fait du contexte industriel, le processus historique simplifie les images. Celles-ci sont donc traditionnellement traitées en noir et blanc. En effet, les capacités de stockage et les technologies existantes ne permettaient pas de traiter des données complexes telles que les images en couleurs. Nous nous limiterons ici aux techniques usuelles d'analyse de documents qui ont été développées spécifiquement pour ce type d'images. Pour plus de détails, le lecteur est invité à se référer aux états-de-l'art détaillés réalisés par [Baird 2014], [Dengel 2014] et [Kise 2014]. Au cours de cette revue, nous évoquerons aussi quelques travaux dédiés à l'analyse de docu-

Éléments de base	Propriétés
Texte imprimé (Figure 2.2a)	<ul style="list-style-type: none"> — longiligne — petites composantes connexes — homogène mais potentiellement dégradé — chevauchements potentiels
Texte manuscrit (Figure 2.2b)	<ul style="list-style-type: none"> — longiligne ou curviligne — petites composantes connexes — variations colorimétriques possibles — chevauchement potentiels
Région (Figures 2.2c, 2.2d et 2.2e)	<ul style="list-style-type: none"> — homogène, texturée, avec des dégradés — variable mais souvent de grande taille — généralement de forme géométrique
Structure graphique (Figures 2.2f et 2.2g)	<ul style="list-style-type: none"> — composée de lignes horizontales et verticales — taille variable — homogène mais potentiellement dégradée
Image (Figure 2.2h)	<ul style="list-style-type: none"> — structure complexe — taille variable — couleurs et textures variées
Logo (Figure 2.2i)	<ul style="list-style-type: none"> — souvent placé en haut ou en bas de l'image — structure relativement simple — variable mais souvent de petite taille — couleurs et textures relativement homogènes
Filigrane (Figure 2.2j)	<ul style="list-style-type: none"> — généralement placé au centre de l'image — structure souvent simple — variable mais souvent de grande taille — couleurs et textures variées — servent de fond — chevauchements avec d'autres éléments

TABLEAU 2.1 – Les éléments de base d'un document et leurs principales propriétés.

ments historiques. Même si ceux-ci ne sont pas soumis aux mêmes contraintes de traitement (notamment en termes de dégradations), certaines problématiques sont similaires ; et les processus appliqués pour les résoudre peuvent être facilement transposés aux documents administratifs. Pour terminer, nous aborderons les problèmes liés à l'étape de binarisation dans la section 2.4.

Le but premier de la segmentation de documents administratifs est l'extraction des données textuelles qui sont porteuses d'informations. Par conséquent, l'objectif est

Éléments de base	Propriétés
Étiquette (Figure 2.3a)	<ul style="list-style-type: none"> — orientation généralement horizontale mais potentiellement oblique — de taille moyenne — couleur la plupart du temps homogène — contient du texte généralement imprimé, mais avec potentiellement des éléments manuscrits — de forme rectangulaire — peu recouvrir des éléments de l'image
Surcharges imprimées (Figures 2.3b et 2.3c)	<ul style="list-style-type: none"> — orientation généralement horizontale mais potentiellement oblique — de taille moyenne — couleur souvent non homogène — contient du texte imprimé, et éventuellement des structures graphiques — chevauchements potentiels avec d'autres éléments
Surcharges manuscrites (Figure 2.3d)	<ul style="list-style-type: none"> — orientation généralement horizontale mais potentiellement oblique — possibles variations dans l'alignement — variable mais souvent de petite taille — couleur souvent non homogène — chevauchements potentiels
Surlignement (Figure 2.3e)	<ul style="list-style-type: none"> — très souvent superposé à du texte — orientation généralement horizontale mais potentiellement oblique — souvent de petite taille — couleur non homogène et potentiellement des mélanges de teintes avec les éléments surlignés — forme des régions souvent rectilignes
Surcharges correctives (Figure 2.3f)	<ul style="list-style-type: none"> — composée de traits ou de texte — superposé ou très proche d'éléments textuels — orientation généralement horizontale mais potentiellement oblique — possibles variations dans l'alignement — souvent petite taille — couleur homogène ou non avec potentiellement des mélanges de teintes — chevauchements potentiels

TABLEAU 2.2 – Caractéristiques des différentes surcharges d'un document.



FIGURE 2.3 – Différentes surcharges pouvant apparaître sur les images de documents.

la localisation et l'identification de régions d'intérêt qui sont ensuite transmises à un système de reconnaissance de texte de type ROC ou RIC. Cependant, nous avons vu dans la section 2.2 qu'un document ne se résume pas seulement à son contenu textuel. D'un point de vue global, la segmentation consiste alors à décomposer le document en ses composantes principales : blocs de texte, graphiques, tableaux, etc. Par ailleurs, nous pouvons remarquer une certaine hiérarchie entre ces éléments. Par exemple, un ensemble de caractères forme un mot ; plusieurs mots composent une ligne de texte ; plusieurs lignes forment un paragraphe.

2.3.1 Stratégies d'analyse

Les images de documents administratifs sont caractérisées par leur structure répondant à un ensemble de règles plus ou moins explicites, et sur lesquelles les approches de segmentation peuvent s'appuyer. Plus un document est structuré, plus il est facile d'en extraire des données de manière automatique. L'analyse de documents ayant une structure, dite de type Manhattan, constituée de zones rectangulaires, bien délimitées, et qui ont la même orientation, est considérée comme un problème

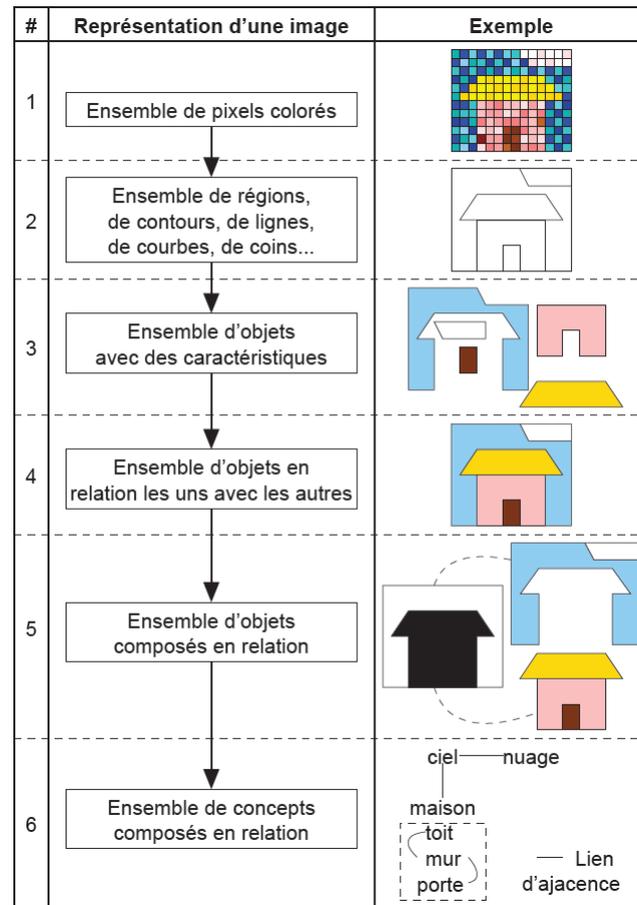


FIGURE 2.4 – Niveaux d'abstraction auxquels une image peut être vue. Crédits : [Dombre 2003] (Figure 3.1)

résolu ([Kise 2014]). L'extraction de certains éléments non homogènes par nature telle que l'écriture manuscrite peut néanmoins poser problème. Les verrous scientifiques identifiés sont liés au traitement des documents complexes, faiblement ou non structurés, et avec des éléments disposés de manière irrégulière sur l'image : espacement, orientation, alignement non constants, superpositions, etc. La qualité de l'image et la présence ou non de bruit influence grandement les résultats de la segmentation.

Une image de document peut être analysée avec une approche *descendante* (*top-down* en anglais), ou *ascendante* (*bottom-up* en anglais). Une approche descendante traite le document dans son ensemble et le divise récursivement en blocs de plus en plus fins jusqu'à l'atteinte d'un critère d'arrêt : bloc de texte, puis lignes, puis caractères. Lorsque la disposition des éléments est connue *a priori* pour un document donné, il est souvent intéressant de construire un *modèle* décrivant les relations entre les blocs élémentaires, et ainsi aider la phase de segmentation. Ce sont des approches

dites “guidées par le modèle”. Au contraire, une méthode ascendante considère tout d’abord les petits éléments et les agrège peu à peu pour constituer des éléments de plus haut niveau. Ce sont des stratégies “guidées par les données” qui se basent sur un ensemble de caractéristiques mesurées sur l’image sans prendre en compte les données contextuelles. Il est possible d’adopter une stratégie hybride. Parmi les méthodes de segmentation, certaines s’appuient sur l’analyse des formes des éléments présents sur l’image, alors que d’autres se basent sur les caractéristiques des régions de fond pour piloter la segmentation.

Afin de pallier les problèmes cités précédemment, la stratégie de certaines approches est de se spécialiser dans l’extraction d’éléments spécifiques, ou dans l’analyse d’un type de documents donné. L’avantage principal est alors la possibilité de contraindre un processus de traitement par un ensemble de paramètres spécialement adaptés à la réalisation d’une sous-tâche particulière. Cette stratégie requiert donc une étape de paramétrage, ou l’injection de connaissance *a priori* notamment avec une phase d’apprentissage. Dans le cas d’un apprentissage, il est nécessaire d’utiliser un ensemble de données intégrant le plus de variabilité pour éviter une trop grande spécialisation dans les traitements. En effet, les résultats seront fortement dépendants de cette base. Le danger de cette spécialisation est de rendre l’approche peu flexible aux variations éventuelles des caractéristiques prises en compte dans la segmentation.

2.3.2 Techniques usuelles de segmentation

Les approches de segmentation fournissent traditionnellement en sortie une image binaire avec les éléments extraits apparaissant en noir sur un fond blanc. [Kise 2014] classe les techniques usuelles en 4 catégories :

- les approches basées sur l’analyse de profils de projections ;
- les techniques de lissage (smearing en anglais) ;
- les approches étudiant les composantes connexes ;
- les approches analysant le fond ;
- les approches basées sur une analyse multi-résolution.

Analyse de profils de projections

Ces méthodes reposent sur le principe que les différents éléments d’une image de documents sont alignés horizontalement, et verticalement. Pour obtenir un profil de projections horizontales (respectivement verticales), il suffit de compter les pixels noirs contenus par chaque colonne (respectivement chaque ligne). Nous obtenons ainsi un histogramme à partir duquel il est possible de séparer les principaux blocs par un seuillage de l’espace entre deux pics (Figure 2.5). Une des techniques les plus communes basées sur l’analyse des profils de projections est le *XY Cut* qui a été proposé pour la première fois par [Nagy 1984]. L’idée est d’appliquer itérativement des séparations horizontales et verticales pour identifier les espaces inter-blocs (Figure 2.6). L’algorithme fournit en sortie un arbre qui permet de décrire les relations

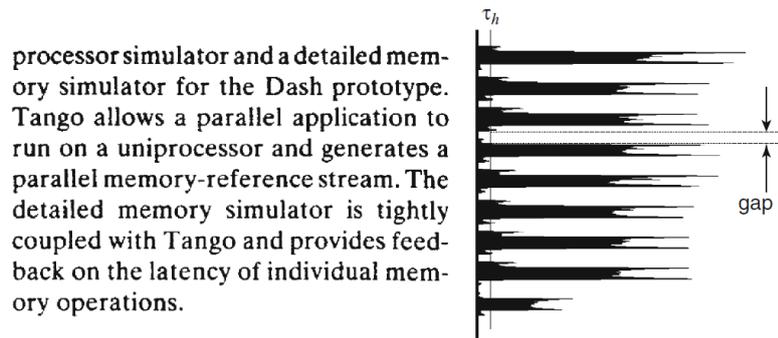


FIGURE 2.5 – Profils de projection. Crédits : [Kise 2014] (Figure 5.10)

hiérarchiques entre les blocs d'éléments de l'image. Le choix de l'ordre d'application des séparations horizontales et verticales peut influencer la qualité des résultats.

A l'origine, ces méthodes ont été proposées pour détecter des lignes de texte parallèles. En revanche, il peut être plus difficile voire impossible de trouver une bonne séparation lorsque le document contient des blocs plus complexes de tailles non uniformes telles que des images. Ils sont donc limités à la segmentation de documents avec des blocs de forme rectangulaire bien marqués (i.e. structure de type Manhattan). De plus, ces approches sont très sensibles aux variations d'alignements et à l'orientation générale du document.

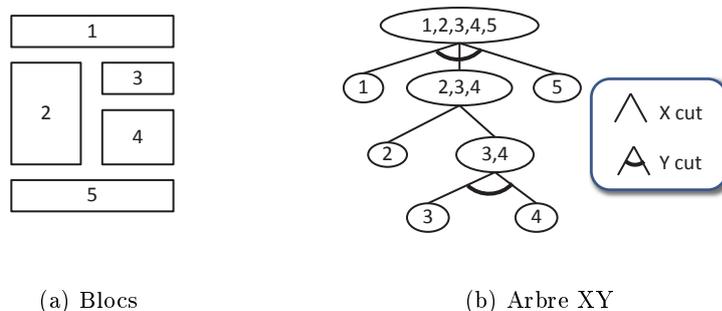


FIGURE 2.6 – Recursive XY cut. Crédits : [Kise 2014] (Figure 5.11)

Techniques de lissage

L'objectif des techniques de lissage est de noircir les plages blanches entre deux pixels noirs si leur longueur est inférieure à un seuil. Tous les éléments appartenant à un même bloc, et dont l'espacement est relativement faible, vont être regroupés de façon à extraire des blocs d'éléments similaires. La structure globale du document devient alors visible. Prenons le cas des régions textuelles. Les caractères vont

être fusionnés en mots, puis en lignes, et finalement, en paragraphes. Les blocs sont ensuite souvent exploités par des systèmes de classification qui vont identifier leur contenu en fonction des propriétés caractéristiques de leurs composantes connexes.

L'algorithme du *RLSA* (Run-Length Smearing Algorithm), issu de la morphologie mathématique, et proposé par [Wong 1982], est l'une des premières techniques développées selon ce principe. L'idée est d'appliquer un lissage horizontal, et un autre verticalement pour obtenir deux masques. Ces masques sont ensuite combinés au moyen d'un ET logique. Pour finir, l'image résultante est utilisée pour le filtrage et l'extraction des régions d'intérêt du document. Un ET logique avec l'image originale permet ensuite de ne conserver que les zones sélectionnées. Cette approche est devenue très populaire pour la segmentation d'images de documents pour sa rapidité et sa simplicité. [Yin 2001], entre autres, a appliqué cet algorithme. L'ensemble du processus est observable sur la figure 2.7. Les blocs sont ensuite classifiés en trois classes (texte, graphique, et image) selon un ensemble de règles heuristiques.

Les techniques de lissage sont maintenant communément utilisées sur des documents simples pour lesquels elles se sont révélées particulièrement efficaces. Ces approches sont rapides. Cependant, elles sont sensibles aux variations de l'orientation de l'image. La valeur des seuils peut influencer la qualité des résultats. Ceux-ci peuvent cependant être déterminés par une analyse statistique des caractéristiques des éléments traités (hauteurs ou largeurs moyennes). Ces méthodes sont dédiées aux documents qui ont une structure de type Manhattan. Elles échouent généralement à segmenter des blocs non rectangulaires, qui sont peu espacés, et non alignés.

Etude des composantes connexes

Les composantes connexes (CCs) d'une image de document peuvent être regroupées ou filtrées selon un ensemble de règles impliquant les propriétés caractéristiques de la composante : position relative, alignement, taille, forme, etc. Pour pouvoir identifier des blocs d'éléments, il est intéressant d'étudier la distance entre CCs. L'hypothèse de base est que la distance entre les CCs d'un même bloc est plus petite que la distance entre celles qui appartiennent à des blocs différents. Ce type de méthode est souvent utilisé pour la détection de texte du fait de la régularité des composantes textuelles en termes de taille, d'espacement et d'alignement.

Ces approches sont souvent basées sur une modélisation de la structure du document sous forme de graphe. Parmi ces techniques, l'*Arbre de recouvrement minimum* (*Minimum Spanning Tree* en anglais) connecte tous les noeuds (CCs de l'image), en minimisant la somme des pondérations de toutes les arêtes du graphe. Ce type de graphe a notamment été utilisé dans des travaux récents pour la segmentation de lignes de texte manuscrites ([Yin 2009], [Fernández-Mota 2014]). Il s'est révélé robuste à plusieurs types de contraintes liées à l'écriture manuscrite : variations d'alignement, de non-homogénéité des espaces inter-lignes, etc.

Utiliser un arbre de recouvrement minimum peut poser problème lorsque la distance entre deux blocs est plus petite que les distances entre les éléments à l'intérieur des blocs. Pour résoudre ce problème, l'approche *Docstrum* proposée par [O'Gorman 1993] utilise un graphe des *k plus proches voisins*. La valeur de k est généralement fixée à 5. En effet, il peut arriver sur certains documents complexes que des blocs d'éléments soient très proches en termes de distance, et que cette distance soit même inférieure à celle mesurée entre les CCs d'un même bloc. Il est alors assez difficile de trouver un seuil suffisamment discriminant pour pouvoir réaliser la séparation. *Docstrum* associe à chaque arête du graphe une paire de valeur (d, ϕ) , avec d la distance, et ϕ , la valeur de l'angle entre les deux CCs considérées. Une approche de partitionnement (clustering en anglais) évalue ensuite les valeurs d'espacement inter-lignes, entre les caractères d'une même ligne, et l'angle d'orientation des lignes. Le fait de considérer les K plus proches voisins de chaque CC augmente la quantité d'information disponible et rend la méthode plus robuste.

Le choix du paramètre k déterminant le nombre de voisins utilisés par l'approche *Docstrum* est délicat. Une valeur trop élevée produira trop de fausses arêtes (des CCs très éloignées seront connectées), et au contraire, une valeur trop petite fournira un nombre insuffisant d'information. [Kise 1998] ont donc proposé une segmentation basée sur la génération d'un diagramme de Voronoï. Cette méthode consiste tout d'abord à extraire un échantillon de points sur les contours des composantes connexes. Ce seront les germes pour la phase de segmentation. Puis une partition de Voronoï est réalisée. Durant cette phase, des *cellules* (i.e. régions) sont générées à partir des germes de telle façon que tous les points appartenant à cette cellule soient plus proches (en termes de distance) du germe correspondant que d'autres germes. Les résultats du partitionnement et de la segmentation peuvent être visualisés sur la figure 2.8. Ce mode de partitionnement est efficace pour les structures complexes car il réalise un pavement de l'image en zone non rectangulaires, qui s'adaptent aux blocs d'éléments à segmenter.

Les blocs d'éléments du document peuvent être associés à un ensemble de régions de Voronoï adjacentes. La tâche de segmentation consiste donc à déterminer quels sont les contours entre deux cellules qui correspondent à des frontières potentielles. Cette opération est réalisée par filtrage : distance entre CCs, ratio surface, etc. La valeur des paramètres dépend du contenu du document. Ils sont déterminés globalement par une étude des fréquences de distribution de la caractéristique considérée. [Agrawal 2009] ont proposé une version améliorée appelée Voronoï++. En effet, la méthode classique peut entraîner des problèmes de sur-segmentation qui surviennent notamment avec des éléments de tailles différentes, ou à cause des dégradations entraînées par la présence de bruit. Voronoï++ détermine donc des seuils locaux plutôt que globaux. D'autre part, l'intégration des relations angulaires et de voisinage proposées par *Docstrum* permet d'augmenter la précision de la segmentation.

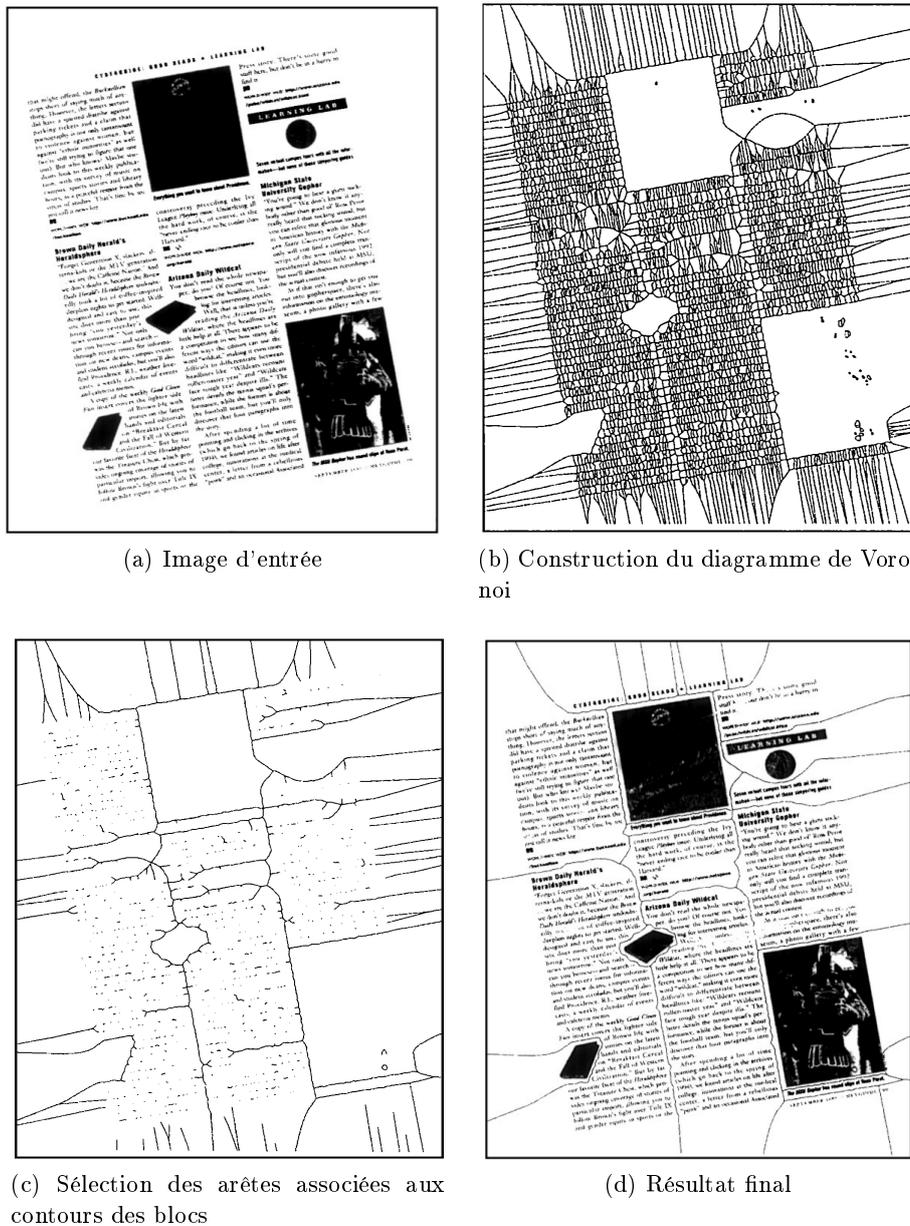


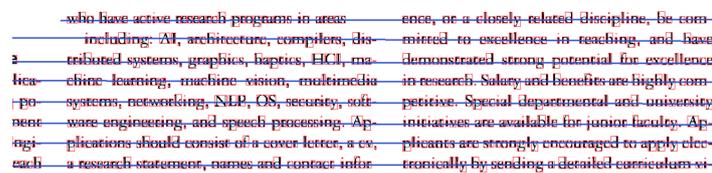
FIGURE 2.8 – Utilisation du diagramme de Voronoï pour la segmentation de documents. Crédits : [Kise 1998] (Figure 4)

Analyse du fond

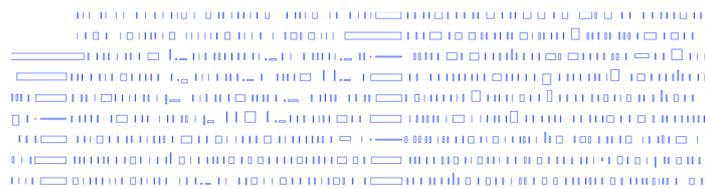
Les approches décrites jusqu'à maintenant sont basées principalement sur la segmentation des régions de premier plan. Lorsqu'un document est constitué de blocs qui ne se chevauchent pas, ces blocs peuvent être identifiés par les zones de fond (globalement rectangulaires) qui les séparent. Récemment, [Chen 2013] ont proposé une approche hybride basée sur ce principe pour l'extraction de blocs de texte.

Les régions de fond sont sélectionnées, puis filtrées et regroupées selon un ensemble d'heuristiques.

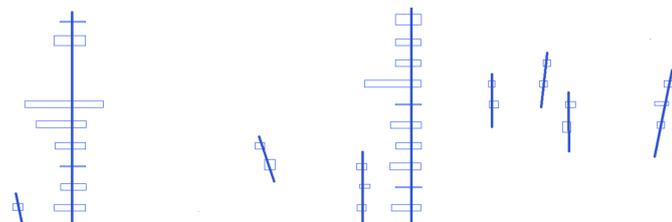
La figure 2.9 résume le processus de filtrage. Tout d'abord, les CCs de l'image sont regroupées sous forme de chaînes. L'hypothèse de base est que les éléments sont alignés horizontalement. Les zones de séparations entre CCs sont extraites comme des régions rectangulaires. Seules les plus larges sont conservées comme candidates. Parmi celles-ci, les rectangles isolés sont éliminés. Un filtrage basé sur des heuristiques permet de classer les rectangles comme étant des séparations probables inter- ou intra-colonnes. A ce niveau, des chaînes de rectangles alignés verticalement sont extraites. Un filtrage final en extrait alors les séparations inter-colonnes.



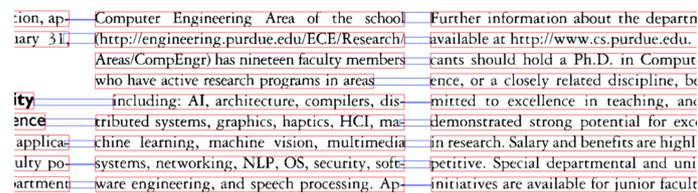
(a) Regroupement des CCs en chaînes



(b) Extraction des zones de séparations les plus larges (inter- et intra-colonnes)



(c) Chaînes de rectangles candidats (après filtrage)



(d) Lignes de texte après suppression des séparations intra-colonnes

FIGURE 2.9 – Segmentation de colonnes de texte basée sur la détection des zones de séparation inter-blocs. Crédits : [Chen 2013] (Figures 6 et 8)

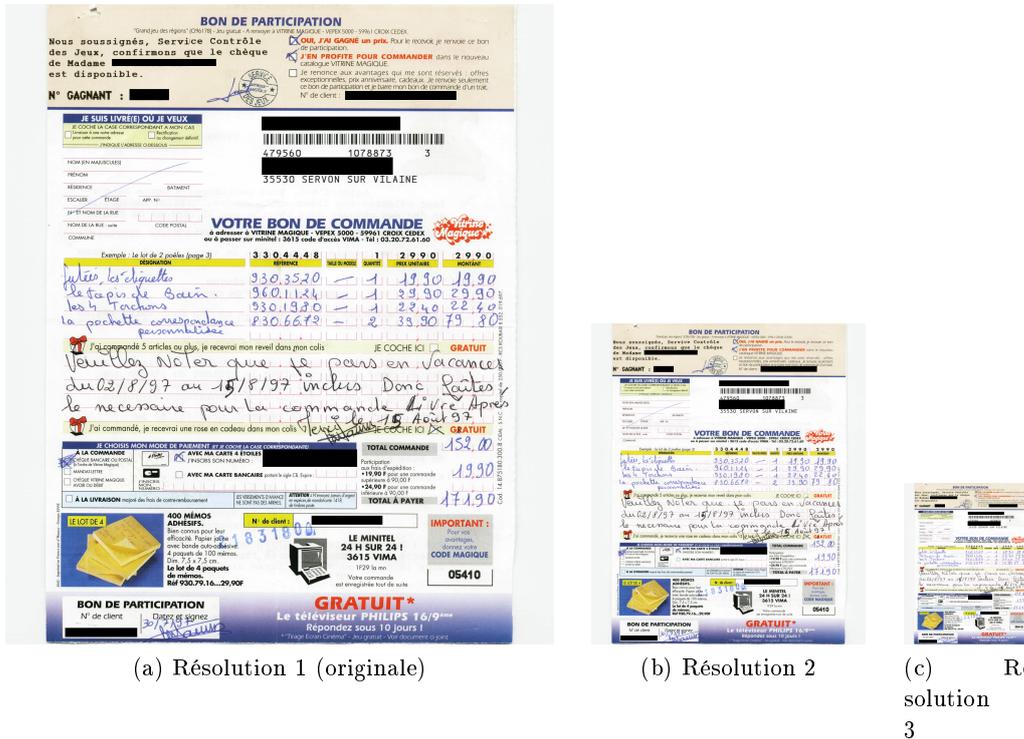


FIGURE 2.10 – Images en multi-résolution construites à partir d’une pyramide gaussienne. L’image 2.10a est divisée par 2 en résolution 2, et par 4 en résolution 3.

Approches multi-résolution

Certaines autres approches se basent sur une analyse multi-résolution de l’image. L’idée est d’observer la structure de l’image à différents niveaux de détails de façon à imiter la perception humaine. En effet, exploitant un phénomène qui est généralement qualifié de *vision périphérique*, un humain aura tendance à regarder tout d’abord une image dans son ensemble pour ne capter que sa structure globale. Dans un deuxième temps, en s’appuyant sur ce qu’on appelle généralement la *vision fovéale* il se focalisera sur une région d’intérêt pour l’étudier plus précisément.

Une analyse multi-résolution se fait sur un ensemble d’images obtenues à partir de simplifications successives d’une image originale (figure 2.10). L’image est alors lissée, puis sous-échantillonnée. Le processus crée ainsi une pyramide d’images dont la taille est divisée par deux à chaque changement de résolution (Figure 2.11). Une approche similaire, appelée multi-échelles, conserve au contraire la taille de l’image, et préfère réduire les détails en appliquant par exemple une technique de lissage. Une autre approche consiste à utiliser des processus de segmentation par fusion ou par division pour créer des pyramides de régions. Pour plus de détails, le lecteur est invité à se référer à l’état de l’art réalisé par [Dombre 2003].

Du fait notamment d'une certaine hiérarchie entre leurs éléments (ex : caractères, mots, lignes, etc), les images de documents se prêtent relativement bien à une analyse multi-résolution. Dans [Lemaitre 2008a] et [Lemaitre 2008b], les auteurs décrivent les bénéfices que peut apporter une telle approche à la reconnaissance de structures sur des images de documents complexes. Ils ont intégré ces concepts dans un système générique DMOS (Description et MODification de la segmentation) qui a été développé pour la reconnaissance de documents structurés. La généralité est introduite par l'utilisation d'un langage grammatical EPF (Enhanced Position Formalism) qui leur permet de décrire les éléments contenus dans une image de document, ainsi que leurs relations spatiales.

Une image en haute résolution contient énormément d'information, et éventuellement du bruit, qu'il n'est pas toujours aisé de traiter. En revanche certains éléments sont plus visibles à certaines résolutions. Par conséquent, l'intérêt majeur de ce type d'analyse est la possibilité d'adapter les traitements au mieux pour faciliter leur extraction. Par exemple, un ensemble de caractères en haute résolution, difficilement distinguables du bruit, apparaîtra comme une ligne, plus facile à extraire, en basse résolution. Il est aussi possible de réaliser une segmentation grossière en basse résolution afin de localiser des régions d'intérêt qui seront ensuite raffinées en haute résolution. D'autre part, la taille de l'image étant réduite, les traitements sont exécutés plus rapidement.

D'après [Lemaitre 2008a], un système d'analyse multi-résolution doit prendre en compte les outils suivants :

- un extracteur d'éléments caractéristiques (composantes connexes, lignes, etc) ;
- un mécanisme permettant de naviguer entre les différentes résolutions de l'image ;
- un mécanisme capable de mettre en correspondance les éléments extraits aux différentes résolutions.

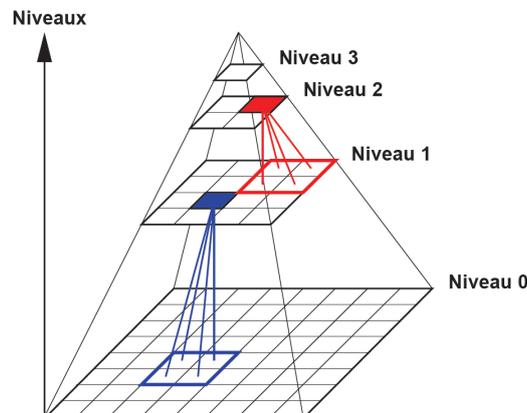


FIGURE 2.11 – Pyramide par réduction. Crédits : [Dombre 2003] (Figure 4.8)

2.3.3 Choix de la méthode

Nous avons présenté jusqu'ici les approches usuelles de segmentation de documents. [Shafait 2006] ont réalisé une étude comparative de plusieurs d'entre elles sur une base de documents contenant à la fois des éléments imprimés et manuscrits. Le choix de la méthode dépend fortement des caractéristiques structurelles des images de documents. Pour un document correctement orienté, et avec une structure de type Manhattan, une méthode basée sur l'analyse des profils de projections, ou une technique de lissage, peut être suffisante. En revanche, pour un document complexe dont le contenu et la structure sont hétérogènes, il faut privilégier une méthode plus générique telle que Docstrum ou un partitionnement de Voronoï qui seront plus flexibles.

Une approche basée sur un partitionnement de Voronoï est l'une des approches les plus performantes. En effet, ce type de méthodes est beaucoup moins sensible aux variations éventuelles de structures (taille, orientations, distances) que des méthodes plus simples. Le choix des paramètres doit être fait judicieusement. Plus le document contient d'éléments différents en termes de taille, ou de positionnement, et plus cette étape devra être adaptative. Les méthodes les plus flexibles réalisent une évaluation statistique (globales ou locales) des caractéristiques du document.

Beaucoup de travaux sont basées sur l'analyse des CCs de l'image. Celles-ci peuvent être pré-filtrées, regroupées, ou classifiées ce qui permet notamment de simplifier le processus de segmentation, et de réduire les coûts de traitements. D'autre part, elles permettent d'incorporer facilement les relations topologiques entre composantes. Les images de documents se prêtent relativement bien à ce type d'approches de par leur caractère structuré. En effet, les CCs associées aux éléments d'un même type ont généralement des caractéristiques similaires dont il est possible de profiter pour les extraire.

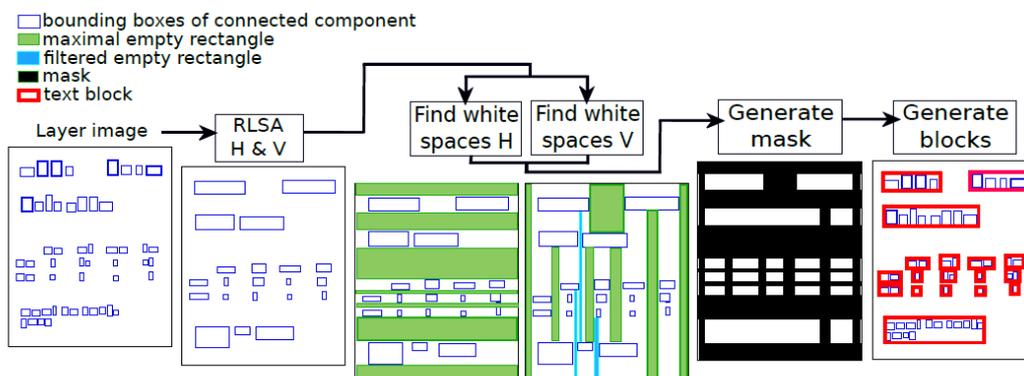


FIGURE 2.12 – Approche proposée par [Barlas 2014] pour la segmentation en blocs. Crédits : [Barlas 2014] (Figure 5)

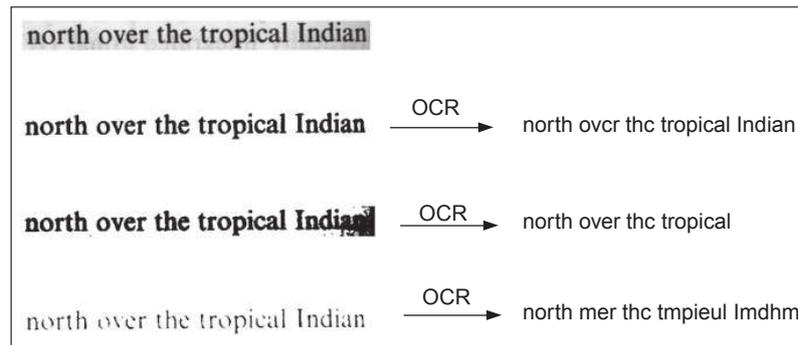


FIGURE 2.13 – Effets du choix de différentes valeurs de seuil de binarisation sur une image, et le résultat fourni par un outil de reconnaissance de texte de type ROC. Crédits : [Dengel 2014] (Figure 6.7)

Par exemple, [Barlas 2014] réalisent une classification des composantes connexes en deux étapes. Une approche basée sur l'utilisation d'un réseau de neurones permet de discriminer les composantes de texte des autres. Dans un deuxième temps, une séparation imprimé/manuscrit est réalisée grâce à une méthode basée sur un dictionnaire de formes. Les composantes connexes ainsi identifiées sont ensuite regroupées en blocs d'éléments homogènes en utilisant des techniques classiques d'analyse telles que le RLSA et une approche basée sur la détection des grands rectangles blancs. La chaîne complète de traitement est résumée sur la figure 2.12.

2.4 Problèmes liés à la binarisation

A travers les sections précédentes, nous avons vu qu'une image de document qualifiée de complexe avait généralement une structure et un contenu très hétérogènes. Le processus industriel historique simplifie ces images pour les traiter en noir et blanc et ainsi réduire les coûts de stockage, et de traitements. Au cours de cette étape, l'image est binarisée de façon à faire apparaître les éléments de premier-plan en noir sur un fond blanc.

Différentes opérations de segmentation peuvent être réalisées sur ces images. Cependant, l'objectif principal de l'analyse de documents est l'extraction de données métiers qui seront injectées dans des systèmes d'informations. Ces données se trouvent dans le contenu textuel des images. Une étape de segmentation va donc identifier et extraire les régions d'intérêt, qui seront ensuite transmises à un système de reconnaissance de texte.

La performance d'un système d'analyse de documents est dépendante du choix et du paramétrage de la méthode de binarisation (Figure 2.13). Une erreur à ce niveau (caractères incomplets ou effacés) engendre une perte d'information lors de la phase de reconnaissance du texte. L'hétérogénéité du contenu, et l'objectif de

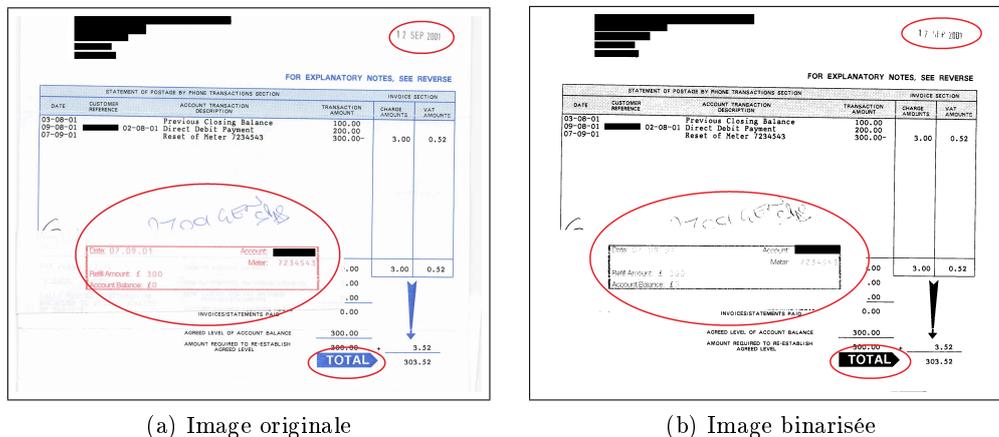


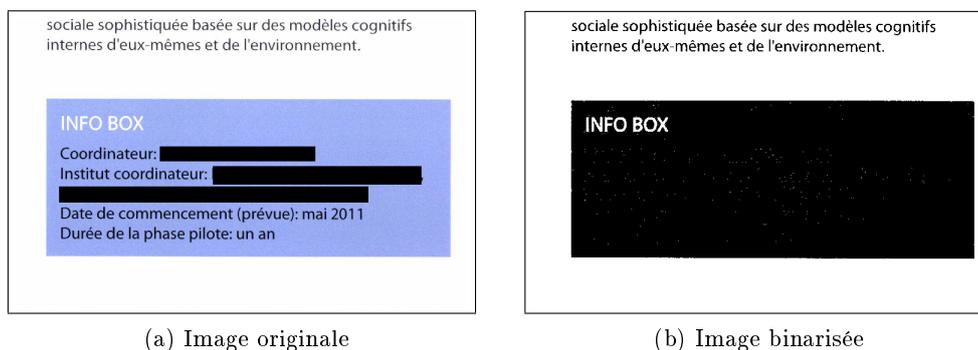
FIGURE 2.14 – Erreurs liées à l'étape de binarisation : difficulté de paramétrage de la méthode de binarisation dues aux variations colorimétriques des éléments, et variations au niveau de la couleur du contenu textuel (sombre sur un fond clair, et inversement, clair sur un fond sombre).

produire des traitements génériques, compliquent la tâche de paramétrage des algorithmes utilisés. En effet, un document couleur contient à la fois des éléments avec des couleurs plutôt contrastées et homogènes, et des éléments, pour lesquels des variations colorimétriques parfois importantes pourront être observées (tampons, surcharges manuscrites, etc)(Figure 2.14).

Par ailleurs, les approches de binarisation partent souvent de l'hypothèse que le texte est globalement plutôt sombre sur un fond clair. Or, ceci n'est pas toujours vérifié sur des documents couleur (Figure 2.14). De plus, la distribution globale des couleurs, ainsi que leur contraste, peuvent là aussi perturber le calcul du seuil. Par exemple, beaucoup de documents contiennent des cartouches avec un fond de couleur sombre et du texte de couleur foncée. La binarisation peut entraîner une perte d'information telle que celle observable sur la figure 2.15.

La segmentation peut aussi échouer lorsque des éléments se chevauchent, ce qui arrive fréquemment avec des objets graphiques et du contenu textuel (Figure 2.16). Des post-traitements peuvent éventuellement corriger certains problèmes. Cependant, ils nécessitent souvent une étape de paramétrage au cours de laquelle les éléments à segmenter doivent être décrits au moyen de connaissance *a priori*. Ce sont donc souvent des solutions *ad hoc*.

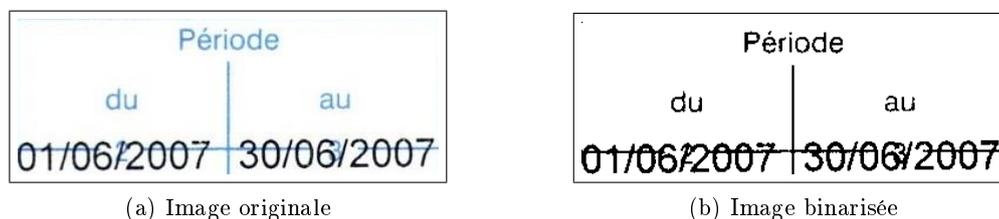
Pour terminer, sur ce type de documents, la couleur est généralement porteuse de sens : surlignement, corrections, etc. Il pourrait être avantageux de profiter de cette information. Or ici, l'étape de binarisation entraîne la perte de toute sémantique liée à la couleur (Figure 2.17). L'amélioration des techniques et des ressources



(a) Image originale

(b) Image binarisée

FIGURE 2.15 – Erreur de segmentation dues à la présence de plusieurs couleurs avec des contrastes différents.



(a) Image originale

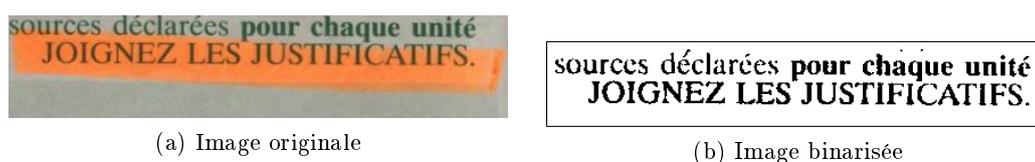
(b) Image binarisée

FIGURE 2.16 – Chevauchements texte/graphiques qui vont venir perturber la segmentation.

de stockage font qu'il est maintenant possible d'envisager des traitements couleur. Au cours de ces travaux, il nous a donc paru intéressant de proposer une alternative à cette séparation en deux classes (premier plan, fond), et d'étudier de quelle façon l'information colorimétrique pourrait améliorer l'étape de segmentation.

2.5 Conclusion

A travers cet état-de-l'art, nous avons synthétisé les principales caractéristiques structurelles des images de documents. Ce sont des images répondant à un ensemble de règles plus ou moins implicites sur lesquelles il est possible de se baser pour en extraire des données particulières. Cependant la densité d'informations contenues par ces images les rend complexes à traiter. Nous avons notamment observé des



(a) Image originale

(b) Image binarisée

FIGURE 2.17 – Perte de l'information sémantique portée par la couleur.

variations parfois importantes d'un document à un autre : surcharges, dégradations, disposition des éléments, etc. Ces variations peuvent aussi être notées pour les documents d'une même famille bien qu'à une plus petite échelle.

L'analyse d'un document passe par une étape de segmentation qui a pour but l'extraction de régions d'intérêt. Dans un contexte industriel, les données métiers se trouvent principalement dans les données textuelles. Pour réduire les coûts de traitement et de stockage, le processus traditionnel simplifie les images en proposant généralement une classification fond/premier-plan. Les documents sont donc convertis en noir en blanc grâce à une opération de binarisation. Par conséquent, les traitements historiques ont été implémentés spécifiquement pour ce type d'images.

Les méthodes de segmentation les plus simples se servent de la structure spécifique des documents qui est souvent de type Manhattan : régions rectangulaires, bien alignées et bien espacées. Ce type d'approches est très sensible aux variations d'orientation et de positionnement des éléments. Malgré tout, elles se sont révélées très efficaces pour des documents simples. Lorsque la structure du document est plus irrégulière (disposition, orientation, espacement), des méthodes plus complexes basées notamment sur les graphes permettent une segmentation plus précise. Cependant, elles nécessitent souvent un paramétrage particulier qui doit être fait avec attention. Ces approches sont fortement dépendantes des valeurs de seuils. La plupart sont donc dédiés à des tâches particulières, et à des catégories spécifiques d'images, partageant les mêmes caractéristiques structurelles.

La qualité de la binarisation impacte fortement l'ensemble des processus de segmentation et de reconnaissance. Or cette étape peut être source d'erreurs à plusieurs niveaux. D'une part, une partie parfois non négligeable de l'information peut être dégradée ou même perdue. Ceci peut se produire à cause des propriétés colorimétriques de certaines régions (couleurs peu contrastées, dégradations, etc). D'autre part, toute l'information attachée à la couleur disparaît. Cela peut être problématique lorsque des éléments de couleurs différentes se chevauchent. De plus, il n'est plus possible d'exploiter certaines informations sémantiques telles que par exemple les surlignements, les corrections, etc. Dans certains cas, il est possible de corriger ces erreurs. Mais ces solutions nécessitent souvent des connaissances *a priori*, et éventuellement, une phase supplémentaire d'interaction avec un utilisateur pour les injecter

En conclusion, les méthodes développées jusqu'à aujourd'hui fonctionnent relativement bien dans le cas de documents simples, ou lorsqu'elles peuvent être paramétrées pour une tâche particulière. Les verrous scientifiques qui ont été identifiés concernent la segmentation de documents complexes, et en particulier, la mise en place de méthodes génériques et donc adaptatives, ne faisant pas intervenir l'utilisateur. En effet, plus un système est spécialisé dans la réalisation d'un objectif donné, et plus il devient sensible aux variations apparaissant dans l'image. A cela vient

s'ajouter le problème d'interdépendance de la reconnaissance et de la segmentation, qui rend difficile l'identification et l'extraction de contenus hétérogènes tels que ceux observés sur les documents à notre disposition.

Malgré un intérêt croissant, il existe encore peu de travaux traitant de l'analyse de documents couleur. L'amélioration des techniques et des capacités de stockage permet maintenant d'utiliser la couleur pour améliorer le processus de segmentation. La problématique serait de proposer une alternative à l'étape de binarisation traditionnellement appliquée, et non pas de remplacer complètement les processus existants. Le chapitre 3 sera donc consacré au traitement de l'information colorimétrique, et à l'analyse de documents couleur. Tout l'enjeu de notre travail sera de pouvoir segmenter, de manière non supervisée, tout type de document, même rencontré pour la première fois, contenant tout type de contenu, et avec des caractéristiques structurelles très variables.

Traitement de l'information couleur

Sommaire

3.1	Introduction	37
3.2	Perception de la couleur	38
3.2.1	Phénomène physique	38
3.2.2	Vision humaine	39
3.2.3	Mécanismes de la vision	40
3.3	Espaces couleurs numériques	41
3.3.1	Systèmes de primaires	42
3.3.2	Systèmes luminance-chrominance	42
3.3.3	Systèmes perceptuels	43
3.3.4	Systèmes d'axes indépendants	43
3.3.5	Systèmes hybrides	43
3.3.6	Modèles d'apparence couleur	43
3.3.7	Choix d'un espace couleur	44
3.4	Traitement de l'information colorimétrique	44
3.4.1	Stratégies	44
3.4.2	Segmentation couleur	46
3.5	Quantification couleur	52
3.5.1	Approches classiques de réduction couleur	52
3.5.2	Approches de nommage de couleurs	53
3.6	Analyse de documents couleur	57
3.6.1	Caractéristiques spécifiques aux documents couleur	57
3.6.2	Segmentation de documents administratifs couleur	58
3.7	Conclusion	68

3.1 Introduction

Lors des travaux présentés dans ce mémoire, nous nous sommes intéressés à ce que pourrait apporter l'utilisation de la couleur au processus de segmentation et de reconnaissance de documents administratifs. L'enjeu n'est pas de remplacer le processus traditionnel, mais plutôt de proposer une alternative qui pourrait améliorer les systèmes actuels, et ouvrir la voie à de nouvelles applications.

Ce chapitre présentera tout d'abord les principaux mécanismes de la vision humaine et du traitement numérique de l'information colorimétrique. Certaines notions ne seront pas développées de manière exhaustive, l'objectif étant plutôt d'avoir une vue générale des différents phénomènes impliqués dans la définition de la couleur. Pour plus de détails, le lecteur est invité à se référer à [Vandenbroucke 2000], [Fairchild 2005], [Busin 2008], et [Fernandez-Maloigne 2013].

Nous décrirons ensuite les différentes stratégies spécifiques à la segmentation d'images couleur. Un des objectifs récurrents des approches existantes est la simplification des données couleur qui sont difficilement exploitables à l'état brut. Cette étape passe souvent par un processus de réduction du nombre de couleurs présentes sur l'image. Nous nous pencherons donc de manière plus détaillée sur les techniques de quantification, et d'identification des couleurs dominantes. Cette partie servira d'introduction à un inventaire des travaux liés à l'analyse de documents, et qui traitent l'information colorimétrique de manière directe. Il en existe encore peu et la plupart sont très récents.

3.2 Perception de la couleur

La colorimétrie est la science de la mesure de la couleur. Nous commencerons par évoquer les phénomènes physiques et les aspects anatomiques liés à la perception humaine. Les théories et modèles mis au point pour expliquer ces mécanismes sont à l'origine des méthodes de codage numérique de la couleur que nous listerons ensuite dans la section 3.3.

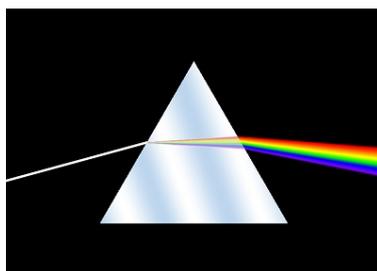


FIGURE 3.1 – Dispersion de la lumière à travers un prisme.

3.2.1 Phénomène physique

La couleur est indissociable de la lumière. Au 17^e siècle, c'est en étudiant la dispersion de faisceaux lumineux à travers un prisme qu'Isaac Newton démontra que la lumière blanche n'est rien d'autre qu'un mélange de lumières colorées (Figure 3.1). Les faisceaux obtenus à la sortie du prisme constituent des couleurs dites "pures" car ils ne peuvent être décomposés à nouveau. Cette expérience fut par la suite à

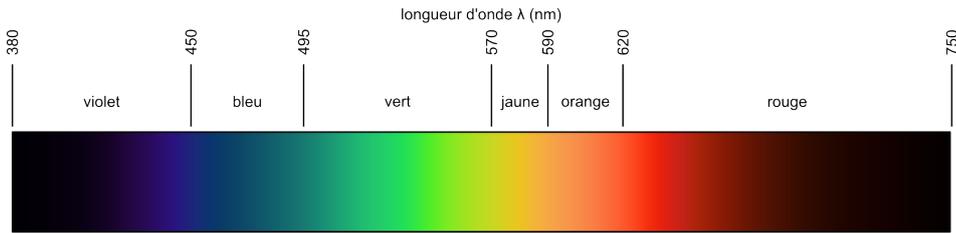


FIGURE 3.2 – Spectre du domaine visible.

l'origine de nombreux travaux qui permirent peu à peu d'expliquer le phénomène de la lumière. Ainsi, la lumière est une onde électromagnétique qui est caractérisée par sa longueur d'onde λ . Le spectre du domaine visible par l'oeil humain contient des ondes comprises entre 380 nm et 780 nm (Figure 3.2).

3.2.2 Vision humaine

L'environnement de perception constitue un triplet "lumière-objet-observateur" (Figure. 3.3). Il comprend :

- Une source lumineuse (appelée *illuminant*) ;
- Un matériau qui réfléchit et absorbe une partie des rayons lumineux ;
- Un observateur avec :
 - L'oeil qui capte la lumière,
 - Le cerveau qui fait une interprétation subjective de la couleur.

D'un point de vue anatomique, la vision des couleurs est rendue possible grâce à des cellules sensibles à la lumière, appelées *photorécepteurs*, qui tapissent la rétine de l'oeil : les bâtonnets et les cônes. Les bâtonnets perçoivent l'intensité de la lumière. Ils sont activés lorsque la luminosité environnante est faible. Les cônes sont associés à la perception des couleurs. Il en existe trois types qui sont plus ou moins sensibles à certaines longueurs d'ondes, et donc à certaines teintes : les teintes bleues (ondes courtes), les teintes vertes (ondes moyennes) et les teintes rouges (ondes longues).

La couleur est l'interprétation sensorielle d'un stimulus lumineux. Nous pouvons distinguer les couleurs *chromatiques* qui possèdent une teinte définie (rouge, vert, bleu, etc), des couleurs *achromatiques* (noir, blanc, et gris). L'environnement joue un rôle important dans la perception. La couleur d'un objet dépend fortement de la source de lumière. C'est pourquoi, les conditions d'observation les plus classiques (soleil, éclairages artificiels...) ont été normalisées par la CIE (Commission Internationale de l'Eclairage) afin de proposer des sources standards appelées *illuminants*.

Il existe un phénomène d'*adaptation visuelle* qui induit un changement de la perception en fonction de la couleur du fond environnant. Par exemple, une couleur apparaît claire sur un fond foncé, et inversement, foncée sur un fond clair. Le système visuel humain ne discrimine pas non plus les couleurs de façon linéaire : il est

plus sensible à certaines variations de teintes, ou de luminance qu'à d'autres. Cette conclusion découle des études réalisées par [Wright 1941] et [MacAdam 1942]. Nous ne rentrerons pas dans les détails des mécanismes impliqués dans ces phénomènes.

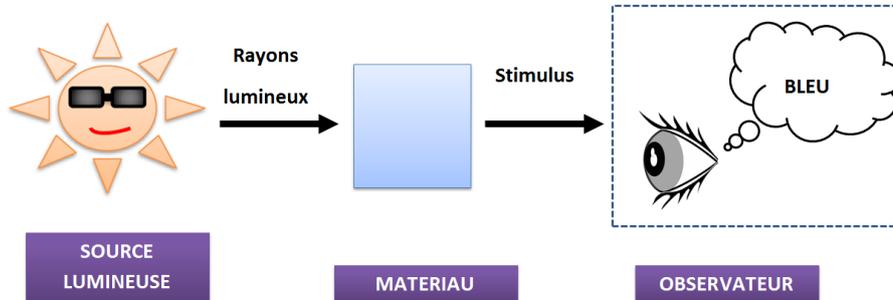


FIGURE 3.3 – Mécanisme de la vision humaine.

3.2.3 Mécanismes de la vision

Plusieurs théories ont été proposées pour expliquer les mécanismes de la vision. Nous allons présenter ici les deux modèles majeurs qui ont conduit à la définition des espaces couleurs numériques : la théorie trichromatique, et la théorie des couleurs opposées. Longtemps, ces théories sont restées incompatibles. Cependant, les recherches actuelles tendent à démontrer un comportement hybride combinant les deux modèles.

Théorie trichromatique

Au 19^e siècle, les travaux de [Young 1807], [Maxwell 1860] et [Helmholtz 1866] ont permis d'établir *la théorie trichromatique* ou *trichromie* basée sur les caractéristiques physiologiques de la vision humaine. Selon cette théorie, il est possible d'obtenir n'importe quelle nuance du domaine visible par un mélange de trois couleurs indépendantes dites *couleurs primaires* ou *couleurs fondamentales*. Il existe deux systèmes de mélanges mettant en jeu la trichromie : la synthèse additive (Figure 3.4a), et la synthèse soustractive (Figure 3.4b).

La synthèse additive est la superposition de trois faisceaux lumineux associés aux couleurs primaires rouge, vert et bleu. Le mélange de ces trois faisceaux dans des proportions égales forme de la lumière blanche. La synthèse additive est utilisée par exemple pour l'affichage sur des écrans. Au contraire, la synthèse soustractive va soustraire une certaine quantité de luminosité à chaque dépôt d'un colorant sur un support matériel. La synthèse soustractive est par exemple utilisée dans l'imprimerie. Les couleurs primaires sont le cyan, le magenta et le jaune. Leur mélange en proportions égales donne du noir. Cependant, en pratique, il est difficile d'obtenir une couleur noire parfaite. C'est pourquoi, une quatrième couleur est généralement ajoutée aux encres cyan, magenta et jaune pour pallier à ce problème.

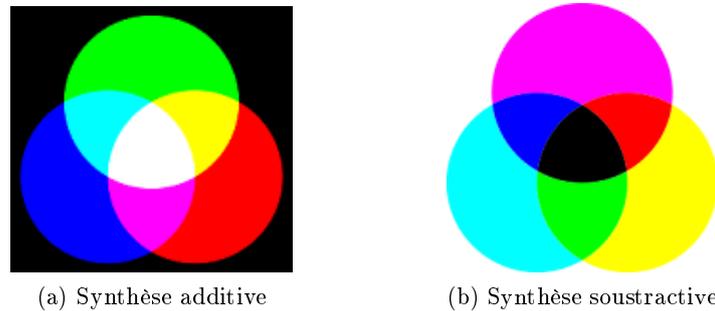


FIGURE 3.4 – Systèmes basées sur la théorie trichromatique.

En 1853, Grassmann énonça les trois lois qui définissent la théorie trichromatique :

- Pour caractériser l'égalisation visuelle d'une couleur, trois paramètres indépendants sont nécessaires et suffisants ;
- Dans un mélange additif de lumières colorées, ce sont les couleurs perçues qui sont significatives et non les répartitions spectrales ;
- Dans un mélange additif de lumières colorées, si une ou plusieurs lumières sont graduellement modifiées, la couleur résultante sera aussi graduellement modifiée.

De nombreux dispositifs d'acquisition d'images et d'affichage ont été conçus selon ces principes. La trichromie est ainsi à l'origine des premiers standards utilisés dans le codage de l'information couleur numérique.

Théorie des couleurs opposées

Dans la deuxième moitié du 19^e siècle, [Hering 1875] proposa un modèle concurrent à la trichromie : la théorie des couleurs opposées. Selon ce modèle, il existe 4 couleurs fondamentales : le rouge, le jaune, le vert, et le bleu. Certaines de ces teintes sont perçues comme antagonistes, ce qui a conduit à la définition d'un modèle impliquant trois types de récepteurs : un pour l'expression de la luminance (l'opposition blanc-noir), et deux pour l'expression de la chrominance (l'opposition rouge-vert, et l'opposition jaune-bleu).

3.3 Espaces couleurs numériques

Au cours de la dématérialisation, un support physique est numérisé lors d'une phase d'acquisition. C'est à dire qu'il va être discrétisé spatialement et colorimétriquement pour donner une image numérique qui sera exploitable par une machine. Dans la section 3.2, nous avons décrit les mécanismes physiques et psychovisuels impliqués dans la vision de la couleur. Nous allons maintenant voir comment ils ont influencé la définition des espaces couleurs numériques.

Depuis longtemps, les hommes ont cherché à organiser les couleurs dans des catalogues ou nuanciers dans lesquels toute couleur considérée trouverait une correspondance. Ces méthodes de classification ont été à l'origine de la définition d'*espaces couleurs*. Les espaces couleurs sont des systèmes de représentation multidimensionnels (souvent tridimensionnels) qui ont pour base des composantes couleurs. Ils permettent d'associer à chaque pixel d'une image un point dans l'espace qui est identifié par ses coordonnées.

Les principaux systèmes de représentation sont :

- les systèmes de primaires ;
- les systèmes luminance-chrominance ;
- les systèmes perceptuels ;
- les systèmes d'axes indépendants ;
- les systèmes hybrides ;
- les modèles d'apparence couleur.

3.3.1 Systèmes de primaires

Les systèmes de primaires reposent sur la théorie trichromatique selon laquelle n'importe quelle nuance peut être reconstituée par un mélange de trois couleurs primaires. Le système le plus communément utilisé est le RVB (Rouge Vert Bleu), qui est historiquement celui mis en place sur de nombreux matériels d'acquisition et d'affichage. Il existe d'autres systèmes de primaires définis par des bases différentes. La seule contrainte pour le choix de ces variables est qu'elles doivent être indépendantes. C'est à dire qu'aucune d'entre elles ne doit pouvoir être reproduite par un mélange des deux autres.

Les systèmes de primaires imitent le fonctionnement anatomique de l'oeil humain. Cependant, en termes de représentation, ils ne sont pas très intuitifs. Leur inconvénient majeur est la forte corrélation entre les composantes couleurs. Par ailleurs, ce sont des systèmes non uniformes qui sont mal adaptés à la perception humaine.

3.3.2 Systèmes luminance-chrominance

Ces systèmes séparent l'information de luminance, de l'information de chrominance. Parmi ceux-ci, nous trouvons les systèmes antagonistes, basés sur la théorie des couleurs opposées, les systèmes de télévision, qui doivent prendre en compte des normes et des contraintes liées au matériel, et les systèmes perceptuellement uniformes. Ces derniers ont été établis de telle manière que les différences calculées dans ces espaces géométriques correspondent aux différences perçues par l'homme. En effet, le système visuel humain ne discrimine pas les couleurs de façon linéaire. Il lui est ainsi plus facile de distinguer certaines variations de teintes ou de luminance, que d'autres.

3.3.3 Systèmes perceptuels

Les systèmes perceptuels ont pour but de caractériser la couleur de manière plus intuitive en utilisant un vocabulaire humain :

- la *teinte* : le nom de la couleur (rouge, vert, bleu, etc) ;
- la *saturation* : un indice de pureté (couleur pure, ou délavée) ;
- la *luminance* : aspect clair ou foncé de la couleur (vive, pâle, terne).

Ils sont basés sur l'interprétation visuelle des couleurs. Ils ont pour objectif de reproduire fidèlement les écarts observables entre deux couleurs. Lors de leur construction, les couleurs sont donc classées selon des critères psychovisuels. Ils vont différer par leur représentation géométrique : systèmes de coordonnées perceptuelles, et systèmes de coordonnées polaires.

3.3.4 Systèmes d'axes indépendants

Les variables utilisées comme bases dans les systèmes précédents peuvent laisser apparaître une certaine corrélation. Cette propriété peut poser problème pour réaliser certaines opérations de traitement d'images et notamment pour les calculs de distance couleur. Les systèmes d'axes indépendants utilisent donc des méthodes statistiques pour étudier la distribution des couleurs sur un ensemble d'images, afin d'en extraire des variables non corrélées. Ces variables sont ensuite utilisées comme base de l'espace. Le principal inconvénient de ces systèmes est qu'ils sont fortement dépendants du jeu de données.

3.3.5 Systèmes hybrides

Certains espaces dits *hybrides* sélectionnent K composantes couleurs parmi plusieurs espaces couleurs en ne conservant que les plus discriminantes pour un cas donné ([Vandenbroucke 2000]). Ce choix est fait de manière supervisée et implique donc l'utilisation de connaissances *a priori*. Il faut alors avoir suffisamment de données annotées pour pouvoir déterminer un bon espace hybride. Ils sont aussi dépendants des images traitées.

3.3.6 Modèles d'apparence couleur

Nous avons vu précédemment que la perception d'une couleur est fortement influencée par l'environnement. Des modèles d'apparence couleur ont été développés pour prendre en compte cette variable : CIECAM97s, puis CIECAM02 qui est plus évolué. Ces modèles cherchent à modéliser le comportement du système visuel humain en incorporant directement des paramètres environnementaux tels que les conditions d'éclairage, les couleurs voisines, etc. L'intérêt majeur est de pouvoir manipuler une image (scanner, afficher, imprimer) indépendamment du support matériel. Les paramètres du modèle seront ajustés de façon à adapter l'apparence visuelle de l'image aux conditions environnementales.

3.3.7 Choix d'un espace couleur

Il existe de nombreux systèmes de représentation qui diffèrent par le choix des primaires, par leur représentation géométrique, et par la métrique qui leur est associée. Une synthèse des principales familles d'espaces couleurs est présentée sur la figure 3.5. Un même espace peut bien sûr appartenir à plusieurs catégories en fonction de ses caractéristiques. Chacun à des avantages et des inconvénients.

Le système RVB est souvent utilisé du fait de son implémentation par de nombreux matériels. Il s'avère suffisant pour de nombreuses tâches et évite une étape de conversion supplémentaire. Cependant, la forte corrélation de ces variables peut rendre le calcul des distances difficile. Les systèmes perceptuels, et en particulier les systèmes perceptuellement uniformes, reflètent mieux les différences réellement perçues par un humain. Les systèmes luminance-chrominance favorisent l'analyse de caractéristiques telles que la teinte, la saturation ou la luminance. Les systèmes d'axes indépendants, hybrides, et les modèles d'apparence couleur sont des systèmes plus complexes à mettre en place, du fait de leur forte dépendance au jeu de données, ou aux paramètres environnementaux. Mais cela les rend aussi plus robustes et plus flexibles que les systèmes classiques. Ils sont donc recommandés pour des applications plus spécifiques pour lesquelles les autres espaces ne sont pas efficaces.

En conclusion, il peut être possible de discriminer plus facilement certains éléments d'une image dans l'un ou dans l'autre de ces espaces. Mais il est communément admis qu'il n'existe pas de systèmes de représentation générique. Par conséquent, le choix d'un espace est fortement lié aux images disponibles et doit être déterminé expérimentalement en fonction de l'application souhaitée.

3.4 Traitement de l'information colorimétrique

Nous avons décrit les différents systèmes numériques de codage de la couleur dans la section 3.3. Nous allons maintenant nous pencher sur les différentes stratégies pouvant être mises en place pour traiter cette information. Nous décrirons ensuite les principales approches de segmentation couleur.

3.4.1 Stratégies

Une image est une matrice de pixels qui peut être traitée en noir et blanc, en niveaux de gris, ou en couleur. Les traitements en noir et blanc ne manipulent que deux valeurs et sont donc relativement simples. En niveau de gris, plus de nuances sont disponibles. Mais il est possible de profiter du caractère ordonné des données (du noir au blanc). La couleur au contraire s'exprime dans un espace géométrique donné, en fonction d'un ensemble de variables indépendantes, généralement au nombre de trois. Les traitements usuellement appliqués à l'analyse d'image en noir et blanc et au niveau de gris doivent alors être adaptés.

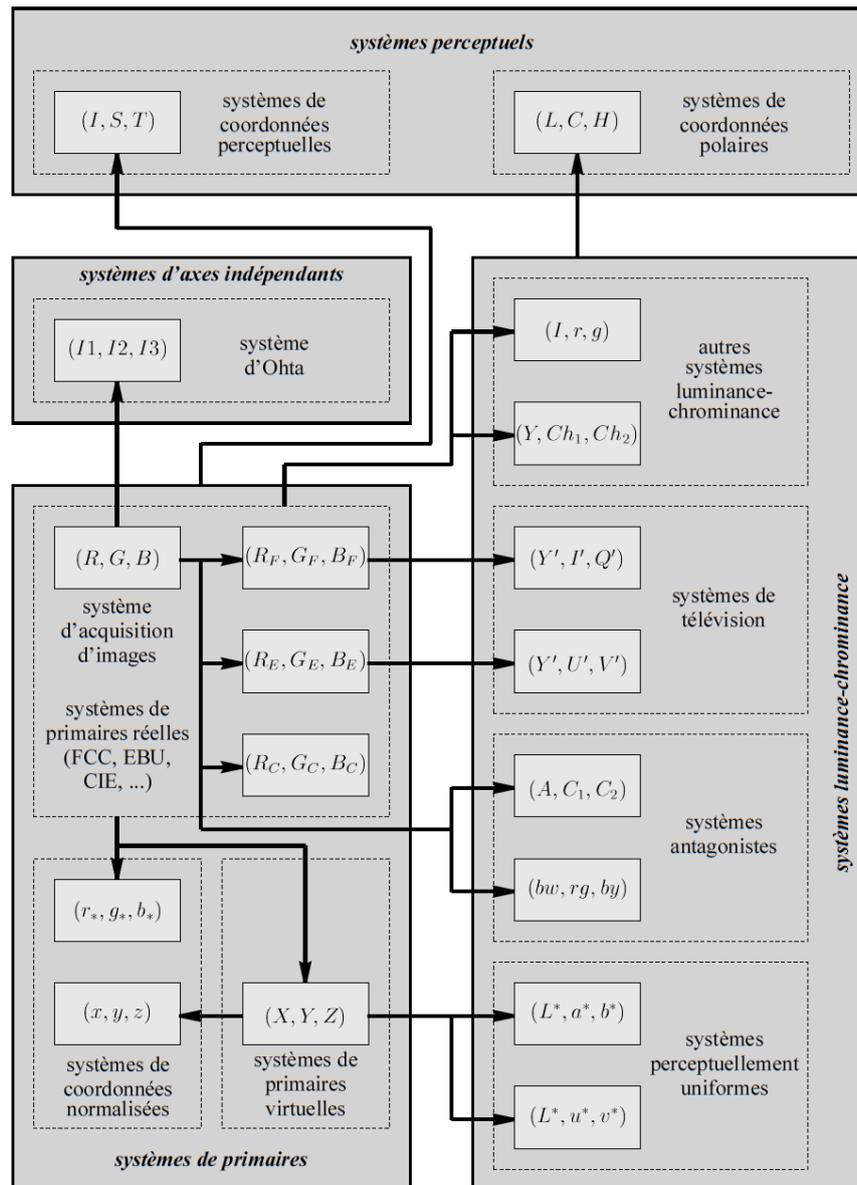


FIGURE 3.5 – Les principales familles d'espaces couleurs. Crédits : [Vandenbroucke 2000] (Figure 2.12).

Plusieurs stratégies de traitement peuvent être appliquées :

- la stratégie marginale ;
- la stratégie vectorielle.

Une approche *marginale* traite chaque composante couleur de manière indépendante, puis fusionne les résultats. Le principal avantage est de faciliter les processus d'analyse en appliquant sans adaptation des approches originellement développées pour

les images en niveaux de gris. Cependant, la corrélation entre les canaux est perdue, au contraire, d'une approche *vectorielle* qui manipule l'information couleur de manière globale. Le traitement d'un vecteur entraîne alors la mise en place de traitements plus complexes. De plus certaines propriétés telles que les relations d'ordres ne sont plus vérifiées.

Nous allons maintenant décrire les principales techniques de segmentation en nous focalisant sur celles manipulant l'information couleur de manière directe.

3.4.2 Segmentation couleur

Nous nous limiterons ici aux principales approches qui nous permettront ensuite d'introduire plus en détails, l'analyse de documents couleur dans la section 3.6.

Les approches de segmentation sont historiquement classées en trois catégories ([Fu 1981]) suivant les éléments qu'elles cherchent à extraire :

- des régions ;
- des contours ;
- des caractéristiques particulières (couleur, texture, etc) de l'image par des techniques de seuillage ou de partitionnement.

Elles sont souvent complémentaires et peuvent dans certains cas être appliquées de manière coopérative. La segmentation peut être réalisée *globalement*, en considérant l'image dans son ensemble, ou bien *localement* en étudiant certaines parties de l'image, et en adaptant les paramètres de l'algorithme aux caractéristiques de la zone étudiée. Il existe aussi des approches *hybrides* qui prennent à la fois en compte le contexte global et local.

Plus récemment, [Vantaram 2012] ont présenté une taxonomie des algorithmes de segmentation bas-niveaux (Annexe A), en les classant en deux catégories selon le type des informations manipulées :

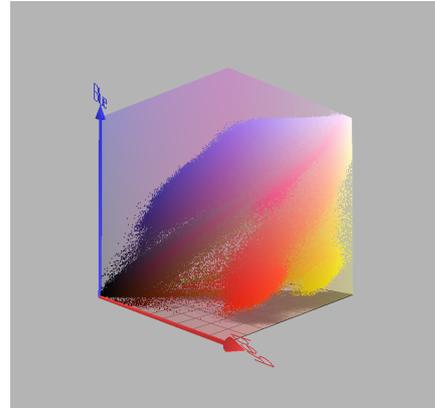
- les approches colorimétriques ;
- les approches spatio-colorimétriques.

Approches colorimétriques

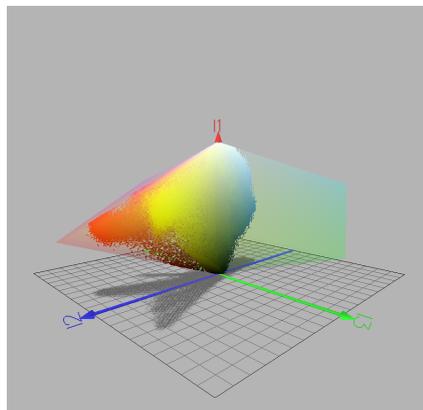
Les approches colorimétriques étudient la distribution des couleurs dans un espace couleur sans tenir compte des relations spatiales entre les pixels de l'image. Elles sont basées sur l'hypothèse qu'une région colorimétriquement homogène dans l'image peut être associée à un sous-ensemble particulier de points ayant des caractéristiques proches dans l'espace couleur. On parlera de *classification pixellaire*. Ces méthodes comprennent les techniques basées sur l'étude d'histogrammes, et les techniques de classification de données.



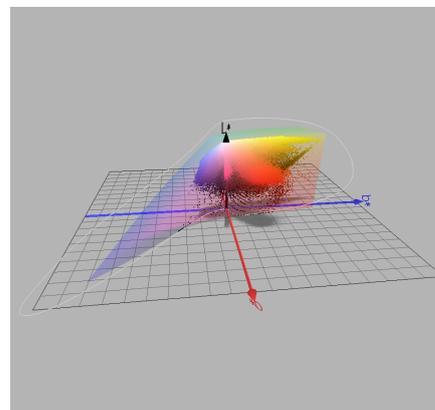
(a) Image originale



(b) Espace RVB



(c) Espace I1I2I3



(d) Espace Lab

FIGURE 3.6 – Affichage des nuages des points d'un document réel dans différents espaces couleurs.

Commençons par les approches basées sur les histogrammes. Un histogramme représente la distribution de la couleur dans une image. D'après l'hypothèse, il est possible d'extraire les principales régions d'une image grâce à l'étude de ses modes (maximaux locaux). La segmentation consiste alors à délimiter les pics, et les vallées (histogramme 1D), ou les nuages de points significatifs (histogrammes 2D et 3D). Le lecteur est invité à se référer à [Busin 2008] pour une étude plus détaillée. La figure 3.6 montre la visualisation des points d'une image de document réel dans plusieurs espaces couleurs.

Les approches basées sur l'étude des histogrammes sont des techniques relativement simples à mettre en oeuvre, qui ne nécessitent aucune connaissance *a priori*, ce qui en fait toute leur force. Cependant, une bonne segmentation nécessite des modes, ou des nuages de points, bien marqués, ce qui n'est pas toujours le cas. La visualisation d'une image dans un espace couleur met en effet en évidence une distribution des

points sous forme de trainées plutôt que sous forme de nuages de points bien séparés. De plus, le nombre de couleurs différentes potentiellement observables sur une image peut poser des problèmes de traitement s'il est trop important, et notamment pour l'analyse d'histogramme 3D ([Fernandez-Maloigne 2013]). Des couleurs faiblement représentées, ou peu contrastées, ne seront pas suffisamment représentées au sein de la distribution globale pour pouvoir être extraites correctement. L'information recherchée se retrouvera noyée par la quantité énorme de données disponibles.

Finalement, l'utilisation de la couleur complexifie le problème. Le choix de la stratégie d'analyse est important. Les modes potentiellement détectés par l'analyse d'un histogramme 1D (stratégie marginale) ne sont pas toujours ceux extraits par l'analyse d'un histogramme 2D ou 3D (stratégie vectorielle). Pour toutes ces raisons, le fait que ces approches ne prennent pas en compte les informations spatiales les rend très sensibles au bruit. En conclusion, il n'est pas toujours possible de trouver une correspondance entre les régions de l'image et les couleurs détectées par une analyse d'histogramme.

Les techniques de *classification* de données quant à elles cherchent à identifier des regroupements, ou classes d'éléments similaires, dans un espace de caractéristiques. En français, ces approches sont aussi appelées méthodes de regroupement, ou encore *clustering* d'après le terme anglais. Les classes qui correspondent ici aux régions segmentées sont aussi couramment appelées par le terme anglais *cluster*. Dans la suite de ce mémoire, nous emploierons indifféremment l'un ou l'autre de ces termes.

[Jain 2010] a réalisé une étude détaillée des différentes techniques de clustering existantes. Il en existe de nombreuses. Les classes sont constituées de manière à minimiser les différences entre les éléments d'un même cluster, selon un critère de similarité donné, tout en maximisant les différences entre les éléments appartenant à deux clusters différents. Lorsque les classes ne sont pas identifiées *a priori*, l'approche est dite *non supervisée*. Dans le cas contraire, elle est *supervisée*. Ces dernières techniques sortent du cadre de notre application et nous ne les évoquerons pas ici.

Les approches de clustering peuvent être classées en trois catégories :

- les approches de partitionnement ;
- les approches hiérarchiques ;
- les approches basées sur la densité.

Les approches de partitionnement divisent un espace en K classes. L'objectif est d'organiser un ensemble de données de façon à identifier des sous-ensembles homogènes, partageant les mêmes propriétés caractéristiques. Les techniques les plus communes sont l'algorithme des *K-moyennes* (*K-means* en anglais), et son extension, les *C-moyennes floues* (*Fuzzy C-means* en anglais). La procédure des *K-moyennes* fait partie des regroupements dits *durs*, qui répartissent les données en K classes disjointes ; chaque élément appartient à une seule classe. Par opposition, les méthodes

de regroupement dites *floues* (*soft clustering* en anglais), telles que les C-moyennes floues, affectent les données à plusieurs classes avec un degré d'appartenance. Cela permet de prendre en compte le chevauchement éventuel des classes, ce qui rend ce type d'approche plus flexible et plus précis.

Les approches de partitionnement sont généralement efficaces et relativement simples à mettre en place. Elles peuvent manipuler de grands volumes de données. Cependant, la plupart du temps, ces méthodes nécessitent de connaître le nombre de classes. L'initialisation des centres des clusters peut poser problème. En effet, différentes initialisations entraîneront des résultats finaux différents. Ces centres doivent donc être choisis avec attention. Pour finir, elles imposent une certaine forme aux clusters : circulaire, sphérique voire hyper-sphérique (si la distance euclidienne est utilisée). Comme nous pouvons le voir sur la figure 3.6, il est difficile d'extraire des nuages de points avec une forme bien définie à partir des données à notre disposition.

Les approches hiérarchiques quant à elles construisent des arbres de clusters organisés sous forme de *dendogrammes*. L'idée générale est de procéder soit par fusion (approche ascendante), soit par division (approche descendante) des groupes d'éléments en se basant sur leur similarité et jusqu'à la validation d'un critère d'arrêt. Ces procédures produisent donc un nombre variable de classes qui dépendra directement de l'image traitée. De plus, dans le cadre de la segmentation, il peut être intéressant de visualiser les relations hiérarchiques entre différentes partitions de l'image car cela permet l'adoption de différents points de vue. Cependant, ces méthodes sont coûteuses en temps de calcul, et la définition d'un bon critère d'arrêt est un problème difficile.

Pour terminer, les approches basées sur la densité ont pour but d'identifier les zones de plus haute densité de points dans l'espace couleur (clusters) qui sont, par hypothèse, séparées les unes des autres par des régions où la densité est plus faible (bruit). Parmi ces méthodes, l'algorithme du *MeanShift* est réputé pour ses bonnes performances en segmentation. MeanShift est une approche itérative non paramétrique qui recherche les maxima locaux dans l'espace couleur grâce à une estimation de la densité basée sur l'utilisation d'un noyau. Les points de l'espace sont ensuite distribués parmi les clusters ainsi identifiés. MeanShift a été proposé pour la première fois par [Fukunaga 1975], et a été développé pour la segmentation couleur par [Comaniciu 2002]. Ce type d'approche est beaucoup plus lent qu'une approche de type K-moyennes. Cependant, aucune hypothèse n'est faite sur le nombre et la forme des clusters. D'une part, cela la rend beaucoup plus robuste au bruit. Et d'autre part, elle peut être appliquée dans des applications pour lesquelles aucune connaissance n'est disponible sur les classes à segmenter.

En conclusion, les approches colorimétriques cherchent à partitionner un espace couleur en groupes d'éléments homogènes. Ces techniques sont relativement simples à mettre en place et peuvent traiter de grands volumes de données rapidement. Elles

permettent aussi une vision globale de l'information couleur des images. Cependant, elles sont très dépendantes des images à traiter et de l'application. En effet, la plupart des méthodes décrites précédemment nécessitent la connaissance du nombre de classes à extraire. Le choix des paramètres d'initialisation est très important, et influence fortement la qualité des résultats de la segmentation. D'autre part, ces approches, aussi bien celles basées sur l'analyse d'histogrammes, que les méthodes de clustering, requièrent des classes bien séparées. Or les images de documents à notre disposition (Figure 3.6) laissent apparaître des données très bruitées, avec une distribution des points sous forme de traînées, plutôt que sous forme de nuages de points bien marqués. Le nombre important de couleurs, les faibles contrastes, et le bruit noient l'information ce qui rend les classes difficiles à extraire.

Approches spatio-colorimétriques

D'après l'état-de-l'art, les approches de segmentation qui utilisent uniquement l'information couleur sont très sensibles au bruit. C'est pourquoi, certains travaux ont cherché à intégrer des données supplémentaires pour décrire l'organisation spatiale des pixels sur l'image et ainsi améliorer la précision de la segmentation. En effet, malgré les dégradations introduites par la présence de bruit, un pixel partage un certain nombre de caractéristiques proches avec son voisinage. Par conséquent, les approches spatio-colorimétriques vont utiliser ces relations spatiales et les combiner avec l'information couleur pour produire des régions compactes et cohérentes par rapport aux régions logiques observables sur l'image. Nous ne ferons pas ici d'étude exhaustive des différentes approches existantes, mais nous évoquerons quelques grands principes de fonctionnement.

Les principales techniques spatio-colorimétriques sont basées sur des approches de croissance, de division, de fusion de régions, ou sur une combinaison de ces trois étapes. Le processus de croissance se fait à partir de *germes*, et s'étend de proche en proche en fonction d'un prédicat d'homogénéité (ex : distance couleur). La croissance stoppe lorsque ce critère n'est plus respecté. Les techniques de division-fusion partent d'une région non homogène qui est récursivement divisée en plus petites régions jusqu'à la satisfaction d'un critère d'homogénéité. Une fois l'étape de division terminée, les petites régions voisines très proches colorimétriquement sont souvent regroupées pour affiner les résultats.

Les approches basées sur les techniques de croissance de régions peuvent être mises en échec lorsque des régions présentent des variations lentes des couleurs (ex : dégradés de couleurs). L'algorithme du CSC (Color Structure Code en anglais) proposé par [Rehrmann 1998] tente de résoudre ce problème. CSC est une méthode de croissance de régions hiérarchiques qui a été appliquée à la segmentation de scènes réelles, et qui ne dépend ni du choix des germes, ni de l'ordre de traitement des éléments. Il se déroule en deux étapes : une étape de fusion qui se déroule de façon hiérarchique en utilisant des relations topologiques de type hexagonal (Figure 3.7), et une

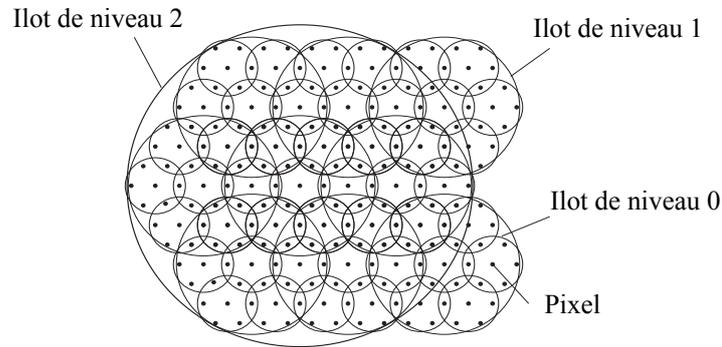


FIGURE 3.7 – Structure hiérarchique de type hexagonal de l'algorithme du CSC. Crédits : [Rehrmann 1998] (Figure 1)

étape de vérification qui redivise les régions quand cela est nécessaire. Le premier inconvénient de cette méthode est sa complexité. D'autre part, la structure topologie choisie peut entraîner une imprécision au niveau de la segmentation : segments de forme rectangulaire. Cependant, le processus hiérarchique intègre des informations à la fois locales et globales, et rend ainsi la segmentation plus robuste même dans des zones faiblement contrastées localement.

La principale difficulté de ces approches est généralement le paramétrage des algorithmes. Les germes et les prédicats d'homogénéité doivent être choisis soigneusement. En effet, l'ordre de fusion peut influencer les résultats et entraîner une sur-segmentation ou une sous-segmentation de l'image. Ces approches fonctionnent généralement bien lorsque les couleurs sont contrastées. Certaines erreurs peuvent apparaître dans des zones particulières comme les régions de couleurs dégradées pour lesquelles la couleur varie très peu localement, mais où la distance entre les extrema est importante. Ces approches peuvent aussi être relativement coûteuses en ressources de calcul.

Cependant, elles peuvent aussi être utilisées comme un pré-traitement appliqué avant la segmentation finale. Certaines approches sur-segmentent ainsi volontairement l'image en un ensemble de régions locales qui tendent à adhérer aux contours réels de l'image. Elles permettent l'application de traitements plus lourds directement sur les régions, appelées *superpixels* (SP), qui deviennent alors l'unité élémentaire du processus de segmentation. En théorie, n'importe quel algorithme de segmentation peut être appliqué. [Achanta 2012] ont comparé plusieurs méthodes basées sur l'utilisation de graphes, ou sur l'étude du gradient. Ces approches sont généralement lourdes en temps de calcul et en utilisation des ressources mémoire. Ils ont donc proposé un nouvel algorithme, *Simple Linear Iterative Clustering* (SLIC),



FIGURE 3.8 – Images segmentées par la méthode SLIC en superpixels de tailles 64, 256 et 1 024 (approximativement). Crédits : [Achanta 2012](Figure 1)

basée sur l'application locale de l'algorithme des K-moyennes. L'approche profite ainsi de l'efficacité et de la rapidité de cette méthode de partitionnement. La combinaison d'une distance couleur et d'une distance spatiale permet de produire des régions qui respectent les caractéristiques géométriques de l'image. L'impact du bruit s'en trouve réduit. Quelques résultats sont visibles sur la figure 3.8. L'algorithme SLIC a été appliqué à la segmentation d'images de documents au cours de ces travaux. Le processus est décrit plus en détails dans le chapitre 6.

3.5 Quantification couleur

Le nombre de couleurs différentes sur une image est potentiellement très important. Par exemple, dans l'espace RVB, la valeur de chaque composante couleur appartient à un intervalle $[0, 255]$, soient 256 valeurs possibles. Par conséquent, cet espace permet de coder $256^3 = 16\,777\,216$ couleurs différentes. Or, le système de vision humain ne permet pas d'en distinguer un si grand nombre. Beaucoup d'approches d'analyse d'images couleur commencent donc par réaliser une étape de *réduction des couleurs*, ou *quantification* dans le but de réduire la complexité de l'information colorimétrique. Ce processus consiste à remplacer la valeur colorimétrique réelle par une autre issue d'une palette réduite. La problématique est alors : comment réduire le nombre potentiel de couleurs à une valeur raisonnable tout en préservant la perception originale de l'image ? Nous allons maintenant présenter quelques approches classiques utilisées en réduction couleur. Pour un état-de-l'art plus détaillé, le lecteur est invité à se référer à [Brun 2002].

3.5.1 Approches classiques de réduction couleur

Les approches classiques de réduction couleur peuvent être classées en deux catégories :

- les approches basées sur des divisions successives ;
- les approches de classification.

Les approches basées sur des divisions successives : Ces méthodes ont pour objectif de diviser un espace couleur de manière itérative. A chaque itération, un ensemble de points de l'espace est choisi pour être divisé en deux nouveaux sous-ensembles. Le processus s'arrête à l'obtention du nombre souhaité de couleurs. Parmi les techniques communément utilisées, nous pouvons citer entre autres la méthode de la coupe médiane (*Median-Cut*, en anglais) proposée par ([Heckbert 1982]), et l'*Octree* ([Gervautz 1990]). La figure 3.9 montre le processus de division du *Median-Cut*. L'*Octree* est une variante qui organise les différents ensembles sous forme d'arbre, et qui permet le regroupement ou l'abandon de certaines branches. L'avantage de ce type d'approche est qu'elles sont rapides et simples à mettre en oeuvre. Cependant, elles ne tiennent pas compte de la distribution des couleurs lors du processus de division.

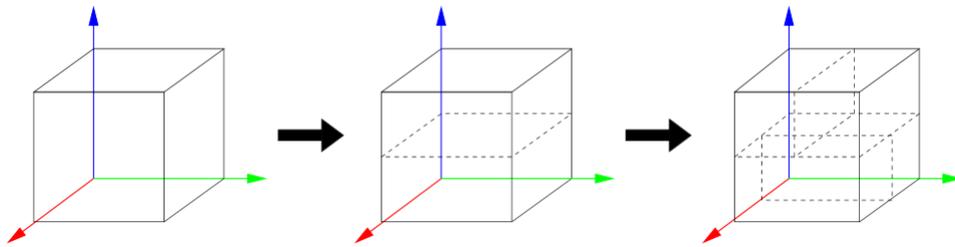


FIGURE 3.9 – Découpage du cube RVB par la méthode *Median-Cut*. Crédits : [Montagne 2005] (Figure 6.21)

Les approches de classification : Ces techniques consistent à rechercher les groupes partageant des caractéristiques similaires dans l'espace couleur. Les techniques de partitionnement ont été présentées dans la section 3.4.2. Parmi elles, nous trouvons des approches basées sur l'algorithme *MeanShift*, qui a été appliqué en analyse de documents par [Nikolaou 2009] et [Lebourgeois 2013] notamment. Nous présenterons ces travaux plus en détails dans la section 3.6.2. Des travaux plus récents, comme par exemple l'approche proposée par [Zagoris 2007], sont basées sur l'utilisation de réseaux de neurones et en particulier, sur les *cartes auto-organisatrices de Kohonen* (*Self-Organizing Map*, en anglais). Ces méthodes donnent souvent de bons résultats. Le temps de calcul nécessaire est cependant important, ce qui peut être un frein à leur utilisation pour de grands volumes de données. Ce critère a d'autant plus d'impact que l'étape de quantification est la plupart du temps utilisée comme un pré-traitement.

3.5.2 Approches de nommage de couleurs

Les approches dites de *nommage de couleurs* ont pour but d'associer à une valeur numérique un label aisément compréhensible par un humain. Même si ce n'est pas forcément leur objectif premier, elles peuvent être considérées comme des méthodes

de quantification dans le sens où elles fournissent en sortie un nombre réduit de couleurs possibles. De plus, contrairement aux approches classiques, les modèles et processus utilisés sont souvent basés sur des caractéristiques perceptuelles du système de vision humain.

La difficulté principale provient de la modélisation : Quels termes utiliser ? Combien ? Quelles sont les frontières dans l'espace couleur ? Beaucoup de travaux se basent sur l'étude réalisée par [Berlin 1969] sur les différents termes linguistiques associés à la couleur. Selon cette étude, les langages évolués partageraient les mêmes catégories basiques, au nombre de 11 : bleu, marron, vert, orange, rose, violet, rouge, jaune, noir, gris et blanc. Ces couleurs peuvent être réparties en deux groupes : les couleurs chromatiques pour lesquelles une teinte spécifique est définie (ex : rouge, vert, bleu, etc), et les couleurs achromatiques (ex : noir, gris et blanc). Certaines études ont cherché à confirmer ce nombre. Nous ne rentrerons pas dans les détails ici.

Cet état-de-l'art reprend globalement celui, très détaillé, réalisé par [Heer 2012]. Les approches les plus simples consistent à partitionner un espace couleur, ou à créer un dictionnaire associant un label à une valeur numérique. Cependant, elles ne prennent pas en compte les caractéristiques psychovisuelles humaines. C'est pourquoi, certains travaux de recherche ont mis en place une phase d'acquisition de données impliquant une interaction humaine. Des modèles statistiques, paramétriques ou non, ont ensuite été construits à partir de ces données. Les modèles paramétriques comme par exemple celui proposé par [Benavente 2002] sont moins sensibles au bruit. Cependant, au contraire des modèles non paramétriques, ils font des hypothèses sur la forme des régions des couleurs principales qui ne sont pas toujours vérifiées en réalité.

La méthode d'acquisition la plus classique consiste à placer un sujet dans un environnement très contrôlé (Figure 3.10) et de lui demander de classer des patches de couleurs ([Benavente 2006] entre autres). Plusieurs expériences peuvent alors être proposées. Les termes colorimétriques utilisés peuvent être imposés : par exemple les 11 cités précédemment. L'utilisateur doit alors trouver la couleur la plus représentative pour une étiquette donnée. Un autre protocole appliqué consiste à nommer un patch de couleur avec le terme le plus approprié. L'objectif est ici le partitionnement d'un espace couleur (Figure 3.11). La consigne peut aussi être, étant donné un nom de couleur, de classer un ensemble de patches par pertinence. Ce type d'approche permet d'introduire de la logique floue dans la construction du modèle ([Benavente 2006]), et ainsi de classer les couleurs avec un degré d'incertitude.

L'environnement contrôlé permet d'assurer la qualité des données. Cependant, il est assez contraignant à mettre en place, et le nombre de participants est généralement réduit pour des raisons organisationnelles évidentes. C'est pourquoi, de plus en plus d'approches se basent sur des données récoltées en masse sur les réseaux. Par exemple, [Heer 2012] utilisent la base publique proposée par [Munroe 2010]. Celle-ci

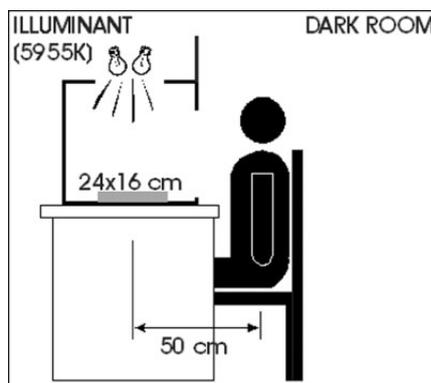


FIGURE 3.10 – Schéma des conditions d'expérimentation de nommage de couleurs. Les expériences ont été réalisées dans des conditions contrôlées, dans une chambre noire, pour s'assurer que les échantillons ne seraient pas illuminés par d'autres sources que celle placée en haut du dispositif. Crédits : [Benavente 2006]

a été constituée en laissant l'utilisateur proposer lui-même un nom de couleur grâce à un outil d'annotation en ligne. La question de la pertinence des données se pose alors. En effet, cette façon de faire implique obligatoirement une phase de nettoyage. Cependant, les réseaux permettent d'une part de récolter une quantité énorme de données : plusieurs millions d'entrées pour [Munroe 2010]. Et d'autre part, ils autorisent la participation de personnes venant de tous horizons. [Moroney 2011] ont validé la pertinence de ce type de ressources : la baisse de la qualité est compensée par la masse de données disponibles. Les résultats sont proches de ceux obtenus dans un environnement contrôlé.

L'utilisation des réseaux permet d'aller encore plus loin dans la construction des modèles en supprimant la phase d'acquisition nécessitant des expériences psychovisuelles qui sont difficiles à mettre en place. En effet, il existe aujourd'hui de très grandes bases d'images publiques déjà annotées dont il est possible de profiter. [van de Weijer 2009] ont utilisé des images réelles provenant de Google ou de Ebay pour alimenter une base d'apprentissage pour construire un modèle grâce à la méthode d'Analyse Sémantique Latente Probabiliste (*Probabilistic Latent Semantic Analysis*, PLSA en anglais). Cette approche permet de réaliser un apprentissage des noms de couleurs en utilisant l'histogramme L^*a^*b normalisé des images. La procédure classique utilise des patches de couleurs qui sont créés dans des conditions connues. Cependant, ces patches ne sont pas représentatifs du bruit qui peut être introduit au cours du cycle de vie d'une image numérique. Cette procédure d'apprentissage au contraire manipule des images réelles qui sont soumises à des variations d'illuminations, de couleurs, etc. Par conséquent, elles sont plus susceptibles de correspondre aux images réellement utilisées dans des applications de nommage de couleurs. Des travaux récents étendent ces modèles en proposant d'associer n'im-

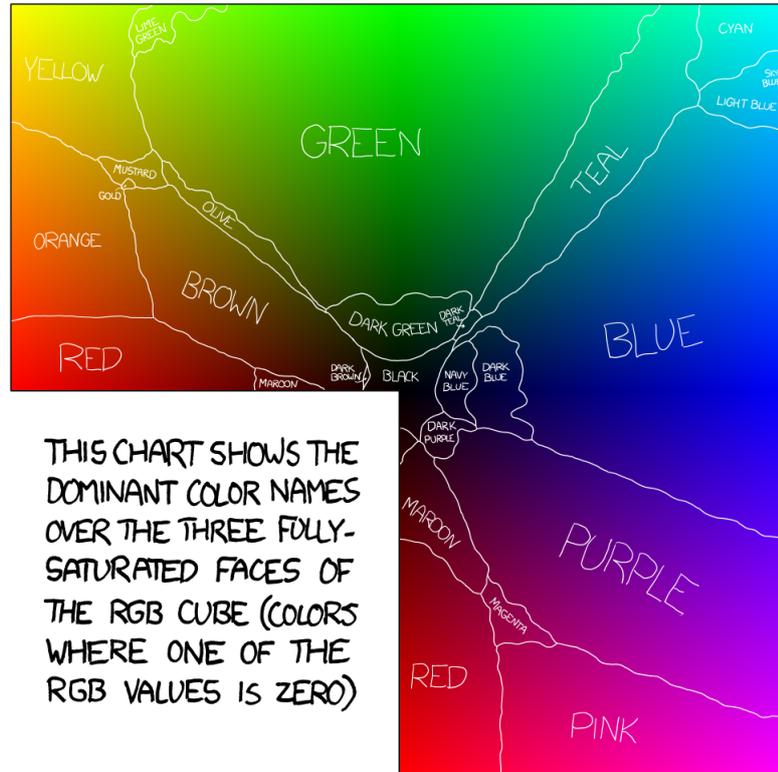


FIGURE 3.11 – Partition de l'espace couleur RGB par une approche nommage de couleur. Ce schéma a été obtenu grâce à des données obtenues à partir d'un outil d'annotation mis en ligne sur l'Internet. Crédits : [Munroe 2010]

porte quelle expression sémantique à un terme de couleur ([Lindner 2012]). Cette approche utilise le concept de *couleurs mémorielles*. L'être humain décrit son environnement en se basant sur son expérience. Une couleur mémorielle correspond ainsi à un objet familier qui s'il n'est pas correctement reproduit choquera l'observateur humain : l'herbe verte, le ciel bleu, etc.

En conclusion, les approches de nommage de couleurs proposent une manière très humaine de caractériser l'image. La plupart des modèles décrits dans cet état-de-l'art met en ligne leurs jeux de données. Les applications sont très variées : descripteurs d'images, segmentation, réduction couleur, etc. Les principales problématiques liées à l'application de ces modèles sont donc la définition d'un nombre pertinent de termes de couleur, et la définition consistante des frontières entre ces couleurs. Dans un contexte de segmentation qui est celui de notre projet, une mauvaise définition de ces paramètres peut rendre les applications sensibles au bruit. En effet, les images de documents sont très hétérogènes en termes de couleurs. Certaines des images fournies par notre partenaire privé telles que les bons de commandes contiennent beaucoup de couleurs. D'autres au contraire, comme les formulaires en contiennent peu, mais

présentent souvent des nuances d'une même teinte. Le risque de sur-segmentation ou au contraire de sous-segmentation est par conséquent assez élevé. Il est possible de réduire cet effet en mettant en place des méthodes de logique floue qui intègrent des degrés d'incertitude et rendent l'application plus adaptative.

Nous avons présenté jusqu'ici les différentes approches et stratégies pour coder la couleur numérique et l'exploiter. Nous allons maintenant inventorier les travaux concernant l'analyse de documents administratifs couleur en nous focalisant plus particulièrement sur les processus qui leur sont dédiés.

3.6 Analyse de documents couleur

Dans la section 2.4, nous avons montré que l'approche classique génère un certain nombre de problèmes du fait des erreurs de segmentation liées à l'étape de binarisation. De par l'amélioration des techniques, il est maintenant envisageable de manipuler directement l'information colorimétrique pour améliorer ce processus. Récemment quelques approches ont été proposées pour apporter des solutions, mais malgré un intérêt croissant, il existe encore peu de travaux traitant de la segmentation de documents administratifs couleur.

Dans un premier temps, nous synthétiserons les caractéristiques spécifiques aux documents couleur, ainsi que les perturbations pouvant affecter leurs propriétés colorimétriques. Puis nous réaliserons un état-de-l'art dédié à l'analyse de cette catégorie de documents.

3.6.1 Caractéristiques spécifiques aux documents couleur

Une image de document administratif est créée spécifiquement pour transmettre des données. Celles-ci doivent donc être facilement localisables et identifiables par un humain. Par conséquent, les couleurs sont généralement choisies pour mettre en relief ces informations. Il est donc possible de faire quelques hypothèses sur les caractéristiques colorimétriques des images de documents.

Globalement, nous pouvons supposer qu'un document contiendra peu de couleurs (souvent moins de 10), et qu'elles seront plutôt contrastées. Cependant, il peut arriver qu'un document contienne plusieurs nuances d'une même teinte principale. Il existe une certaine structuration dans l'utilisation de la couleur : titres, cartouches de couleur, etc. Les surcharges telles que les tampons ont de préférence des couleurs très contrastées par rapport au fond de façon à attirer l'oeil. Les zones textuelles sont souvent de couleurs sombres sur un fond clair mais pas systématiquement.

Les documents administratifs sont très hétérogènes en termes de couleurs comme nous l'avons décrit dans la section 2.2. Nous trouverons donc aussi bien des régions homogènes, que des régions texturées ou des dégradés de couleurs qui peuvent être

problématiques. A cela s'ajoutent des effets de mélange de couleurs lorsque deux éléments se chevauchent (figure 3.12), et l'introduction de bruit, qui viennent là aussi perturber le processus de segmentation. Même si une région apparaît comme globalement homogène, visuellement, des variations et la présence de fausses couleurs pourront être observées à un niveau pixel.



FIGURE 3.12 – Lorsque des éléments se chevauchent, des effets de mélange créent de nouvelles couleurs.

[Ye 2013] ont répertorié les grandes étapes du cycle de vie des documents qui sont susceptibles de dégrader la qualité de ces images :

1. la création (support, impression, éléments manuscrits, etc).
2. des dégradations externes (vieillessement, pliures, annotations, etc)
3. la numérisation (matériel, apparition de fausses couleurs, etc)
4. les traitements (compression, binarisation, etc)

[Ouji 2012] distingue trois types de bruits dégradant la couleur :

1. Le bruit de saturation : Des pixels chromatiques (colorés) sont introduits au voisinage de traits noirs ou dans des zones originellement achromatiques (en niveaux de gris). Ce type de bruit est plus particulièrement engendré par les caméras linéaires (Figure 3.13).
2. Le bruit de teinte : il s'agit de pixels chromatiques dont la couleur a été altérée. Ce genre de distorsion est généralement dû à une inadéquation entre la résolution de numérisation et le tramage (Figure 3.14).
3. Le bruit de luminance généralement causé par les basses résolutions. Ce bruit consiste en des traits originellement noirs qui paraissent gris dans l'image scannée.

3.6.2 Segmentation de documents administratifs couleur

L'étape de binarisation qui est historiquement appliquée consiste globalement à faire une séparation fond/premier-plan, souvent par la mise en place d'une approche de seuillage. Reportée dans le contexte de la segmentation couleur, cette opération s'apparente à une approche de *partitionnement* de données. L'idée est alors d'identifier des regroupements, ou classes d'éléments similaires, dans un espace de caractéristiques. Pour plus d'informations, le lecteur est invité à se référer à la section 3.4.2 qui présentent les techniques de segmentation couleur.

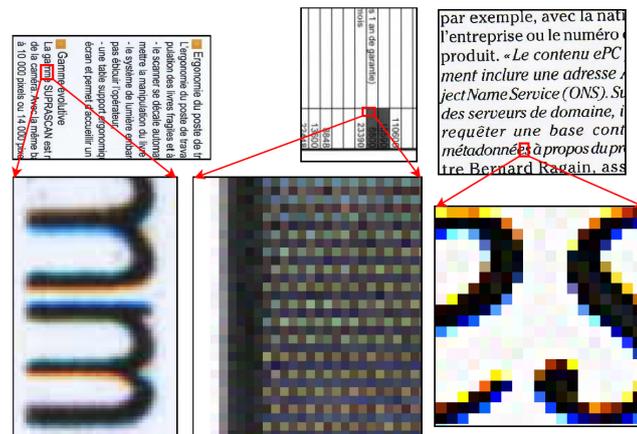


FIGURE 3.13 – Bruit de saturation. Crédits : [Ouji 2012](Figure I.1)

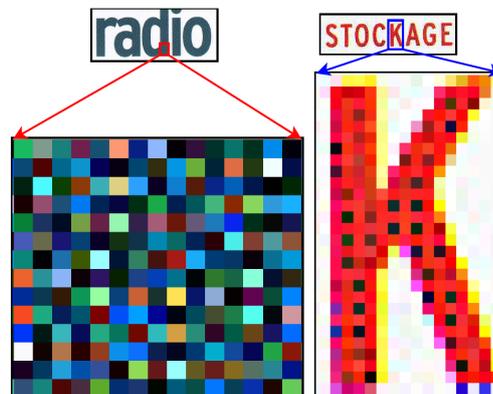


FIGURE 3.14 – Bruit de teinte. Crédits : [Ouji 2012] (Figure I.2)

Comme nous l'avons vu dans la section 3.6.1, du bruit est introduit par la numérisation et la compression sous forme de fausses couleurs et d'artéfacts. Les dégradations qui apparaissent alors peuvent aussi affaiblir le contraste entre les différentes couleurs observables sur l'image. Par conséquent, une des difficultés de la segmentation est la détection du nombre de couleurs observables sur l'image, et la définition de leurs frontières. De plus, une image de documents contient des éléments de différentes couleurs aussi bien de premier-plan que de fond, ce qui complique la tâche de segmentation. Une solution pour résoudre ce problème serait de combiner des données provenant de plusieurs sources. Il pourrait être intéressant d'utiliser notamment la structure des objets en plus de leur couleur.

Il existe encore peu de travaux concernant l'analyse de documents administratifs couleur. Cependant, l'amélioration des techniques et les applications industrielles potentielles ont entraîné un intérêt croissant pour ce type d'approches. Cette sec-

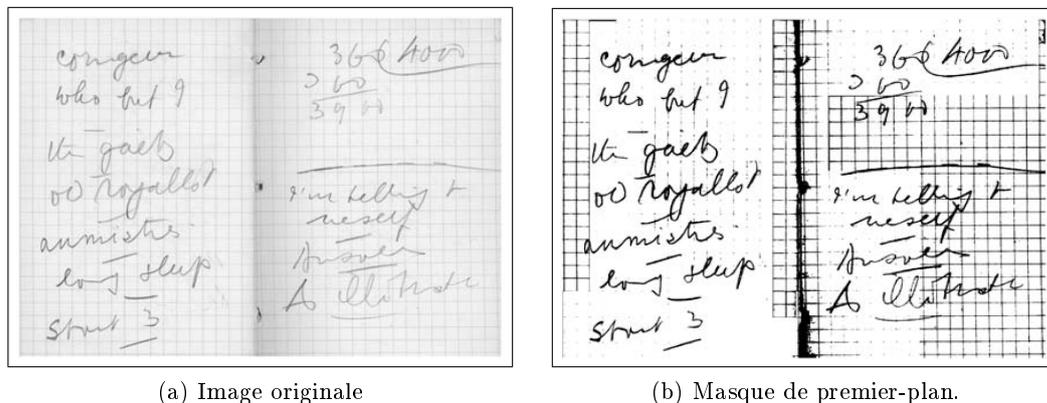


FIGURE 3.15 – Erreurs possibles observables après la séparation fond/premier-plan proposée par [Garain 2005]. Crédits : [Garain 2005](Figure 13)

tion leur sera consacrée. Nous les avons classés arbitrairement selon leur objectif :

- séparation fond/premier-plan prenant en compte la couleur ;
- segmentation en couleurs principales ;
- extraction des zones textuelles ;
- identification de blocs d'éléments ;
- localisation d'éléments.

Séparation fond/premier-plan prenant en compte la couleur

Certains travaux conservent le principe de la séparation fond/premier-plan, toujours dans un souci de simplification, mais proposent des processus de binarisation intelligents, prenant en compte la couleur.

L'un des travaux les plus anciens traitant les documents couleur est le format de fichier DjVu qui a été proposé par [Bottou 1998] pour la compression de ce type d'images. L'idée derrière cette approche est d'adapter la technique de compression au contenu des documents. En effet, certaines zones telles que les régions textuelles nécessitent une image de bonne qualité, en haute résolution, pour pouvoir être exploitées. Les zones de fond en revanche qui contiennent moins de détails peuvent supporter une compression plus élevée. La séparation fond/premier-plan est réalisée grâce à une classification multi-échelles en deux couleurs. Cette approche s'est révélée très efficace lorsque les régions de l'image sont suffisamment contrastées. En revanche, des erreurs peuvent se produire lorsque celles-ci sont bruitées ou lorsque le contraste est faible comme par exemple dans le cas de texte manuscrit.

[Garain 2005] ont proposé une séparation fond/premier-plan pour des images de documents couleur de basse qualité. La méthode consiste tout d'abord à extraire les composantes connexes. Les grands zones de fond sont alors localisées, et organisées hiérarchiquement en fonction de leur taille et de leurs caractéristiques géométriques.

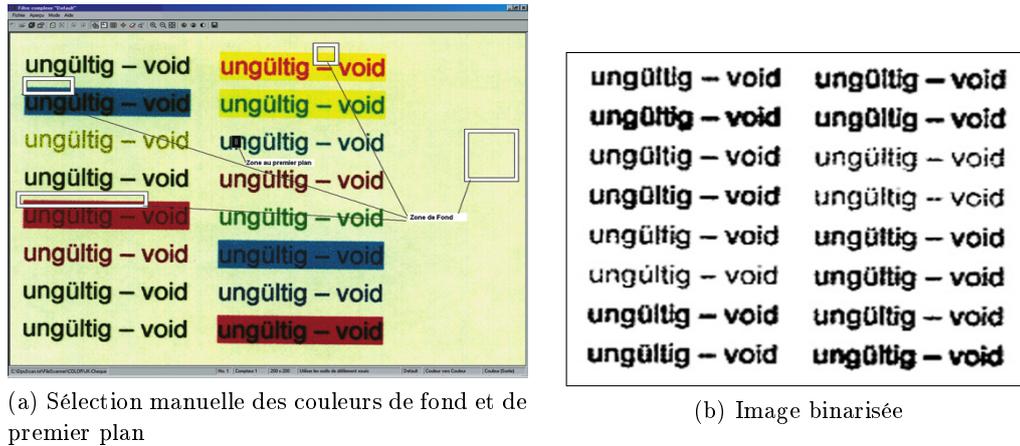


FIGURE 3.16 – Filtrage des couleurs de fond par l'outil ACDP de DpuScan, et binarisation impliquant un seuillage intelligent. Crédits : DpuScan ¹.

Une classification en deux couleurs est appliquée localement pour binariser l'image. Ils proposent ensuite une phase supplémentaire, qui peut être appliquée sur chacun des masques, pour détecter les éléments dont la couleur varie beaucoup. Ceux-ci sont alors étiquetés avec des labels différents, ce qui permet de conserver cette information au contraire des approches classiques. Ces approches semblent donner de bons résultats sur des documents contenant des régions textuelles sur du fond. Cependant, la séparation fond/premier-plan est relativement restrictive et peut provoquer des erreurs sur des documents complexes contenant différents types éléments (Figure 3.15).

D'autres approches consistent à guider le processus de segmentation grâce à des connaissances *a priori* injectées par un utilisateur. Ainsi, le logiciel DpuScan ¹ a été développé par l'entreprise Janich&Klass Computertechnik GmbH pour l'extraction de texte en couleur dans des documents industriels. L'idée est de proposer à l'utilisateur de sélectionner manuellement les couleurs de fond à éliminer, et les couleurs de premier-plan à conserver (Figure 3.16a). Leur module *Advanced Color Document Processing* (ACDP) convertit ensuite l'image en niveaux de gris tout en filtrant les couleurs correspondantes. Finalement, une approche de seuillage binarise l'image en adaptant sa stratégie à la qualité de la zone locale considérée. La figure 3.16b montre les résultats obtenus. L'intérêt principal de ce type d'approche est de faciliter la séparation des zones textuelles des zones de fond en particulier lorsque ces régions sont faiblement contrastées (surlignement bleu d'un texte noir par exemple). Cependant, même si cette interaction peut être considérée comme minimaliste pour du traitement par lots, elle est beaucoup plus contraignante pour un flux de documents hétérogènes.

1. <http://dpuscan.de>

Segmentation en couleurs principales

La couleur est une donnée complexe à traiter en particulier lorsque les images sont soumises à des dégradations causées par l'introduction de bruit. Pour faciliter les traitements, beaucoup de travaux appliquent une étape de quantification afin d'identifier un nombre raisonnable de couleurs principales. Nous avons évoqué les principales techniques de réduction couleur dans la section 3.5. Nous allons maintenant décrire plus en détails celles dédiées à l'analyse de documents couleur.

Les méthodes de partitionnement, dont l'objectif est de regrouper des éléments en fonction d'une mesure de similarité, donnent généralement de bons résultats pour une tâche de quantification. Parmi celles-ci, le MeanShift a été utilisée par plusieurs approches en analyse de documents. Son but est de localiser les maxima d'une fonction de densité. L'un de ces avantages est sa capacité à pouvoir trouver des clusters non sphériques. En effet, sur des images de documents, la couleur se présente rarement sous forme de nuages de points bien marqués. Cependant, le processus est coûteux en temps de calcul et beaucoup de travaux cherchent à réduire la quantité de données à traiter. Par exemple [Nikolaou 2009] commencent par appliquer une phase d'échantillonnage suivie par une première étape de regroupement, ce qui a pour résultat de réduire significativement le nombre de couleurs candidates. Les temps de calcul s'en trouvent fortement réduits d'une part ; et d'autre part, en ne se limitant qu'aux informations essentielles, elle permet d'avoir une meilleure vision de la distribution des couleurs. [Lebourgeois 2013] ont proposé une version optimisée et plus rapide du MeanShift pour traiter des documents de grandes tailles. Cette approche se base notamment sur l'utilisation de la technique des *Volumes Intégrales*, qui permet de calculer la somme de tous les points contenus dans une boîte rectangulaire en temps constant, réduisant ainsi la complexité du MeanShift. Cette version a été ensuite utilisée par [Kessi 2015]. La figure 3.17 montre les résultats de cette quantification.

Dans sa thèse, [Ouji 2012] a proposé une segmentation des images de documents en différentes couches en fonction de leurs propriétés colorimétriques. L'objectif est d'une part de distinguer les vraies couleurs de celles causées par la présence de bruit (voir section 3.6.1). D'autre part, l'enjeu est de pouvoir adapter les processus appliqués en fonction de propriétés particulières à chacune des couches. La figure 3.18 résume l'ensemble des couches ainsi extraites. La première étape consiste en une *séparation chromatique-achromatique*. En effet, les régions de l'image qui sont en noir et blanc, ou en niveaux de gris, (régions achromatiques) ne doivent pas forcément être traitées de la même manière que les régions en couleur (régions chromatiques) ; c'est à dire qui ont une teinte définie (rouge, vert, bleu, jaune, etc). Une zone en noir et blanc peut être binarisée par des méthodes classiques sans risque de perte d'informations. De la même façon, une zone mono-chromatique peut être quantifiée sans danger. Cependant certaines régions telles que les zones multi-chromatiques, ou en niveaux de gris sont plus compliquées à segmenter. Dans un second temps, l'algo-



(a) Image originale contenant 73 727 couleurs différentes.

(b) Image contenant 22 couleurs différentes après l'étape de réduction couleur.

FIGURE 3.17 – Application du Fast Integral MeanShift proposé par [Lebourgeois 2013]. Crédits : [Kessi 2015] (Figure 4)

rithme va donc les séparer du reste. Dans une application de réduction couleur, elles seront laissées telles quelles afin de ne perdre aucune information. Cette approche repose sur le seuillage des histogrammes de teinte, pour la séparation chromatique, et de luminosité pour la séparation achromatique. Pour améliorer la précision de la segmentation, le processus combine des informations globales et locales.

A titre d'exemple, la figure 3.19 compare différentes approches de réduction du nombre de couleurs.

Extraction des zones textuelles

Une partie non négligeable des applications d'analyse de documents est consacrée à l'extraction de données métiers contenues par les régions textuelles. Ces dernières ont des caractéristiques de taille, d'orientation et de couleurs très variables, en particulier pour l'écriture manuscrite. En fonction de la qualité de l'image, il peut être difficile de distinguer les régions réelles de l'image, des régions de bruit car elles ont des caractéristiques très proches. Elles peuvent aussi se chevaucher avec d'autres structures, ce qui complique encore la segmentation. En termes de couleurs, nous pouvons supposer que le texte présent sur une image de document sera de préférence de couleur foncée sur un fond clair, mais il est aussi possible de trouver l'inverse. A la fin de la segmentation, il est souvent nécessaire de reconstituer les caractères ou les mots dont une partie peut avoir été fragmentée et perdue.

En analyse de documents, l'extraction du texte se fait généralement par un filtrage des composantes connexes, initialement obtenues par segmentation, selon leur caractéristiques géométriques (taille, largeur, alignement, etc). En effet, ces méthodes sont plus rapides que d'autres basées sur l'étude des textures ([Badekas 2006]). [Loo 2004] ont utilisé une approche de segmentation basée sur une croissance de régions. Comme nous l'avons vu dans la section 3.4.2, ces approches combinent des données spatiales et colorimétriques, ce qui les rend plus robustes. Cependant, en ne considérant que

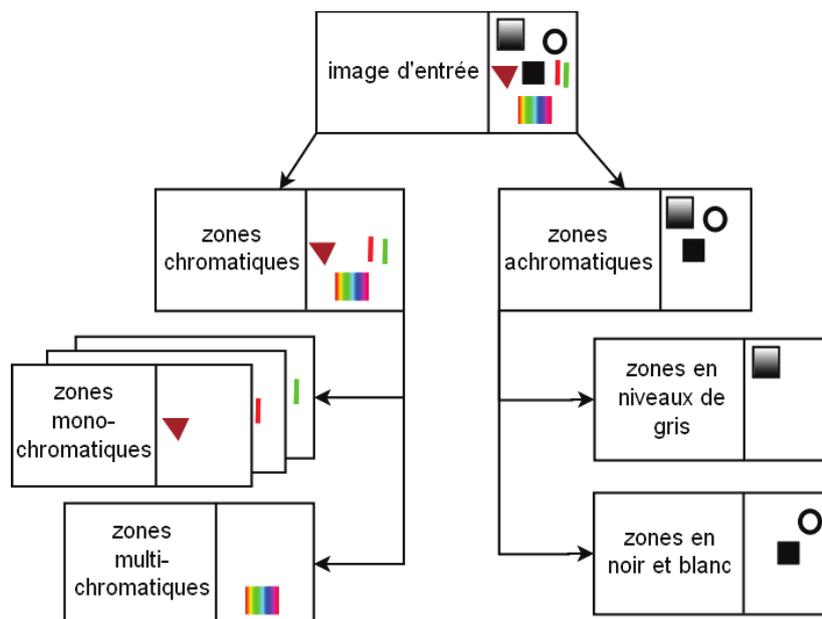


FIGURE 3.18 – Couches issues de la segmentation colorimétrique proposée par [Ouji 2012]. Crédits : [Ouji 2012] (Figure I.4)

des caractéristiques locales, elles peuvent échouer à segmenter des régions faiblement contrastées. Les étapes d'extraction des germes et de fusion des régions sont généralement faites selon un ensemble de règles heuristiques dont dépend grandement la qualité des résultats. Le risque de sur-segmentation est élevé dans le cas d'une base de documents hétérogènes car ces règles ne sont pas forcément généralisables à toutes les situations possibles.

La segmentation en couches couleur proposée par [Ouji 2012] (section 3.6.2) fournit en sortie des images binaires sur lesquelles sont utilisées des techniques classiques d'extraction de texte : groupement de connexités pour les couches mono-chromatiques, technique des gradients cumulés pour les couches multi-chromatiques et grises. Cette approche est intéressante. Mais elle est dédiée aux images de magazines de qualité raisonnable, et qui sont beaucoup moins complexes en termes de contenus que les images à notre disposition. Elles ne contiennent pas d'éléments manuscrits, et globalement pas de surcharges comme sur les images de documents administratifs. La présence de ces éléments pourrait être sources d'erreurs du fait notamment des dégradations colorimétriques, mais aussi de par la richesse du contenu dont les variations en termes de textures, et de caractéristiques géométriques sont généralement importantes.

Récemment, [Kessi 2015] ont proposé une approche basée sur la morphologie mathématique couleur pour segmenter du texte quelle que soit sa couleur et celle du fond. Cependant, leur méthode nécessite plusieurs post-traitements pour séparer le texte

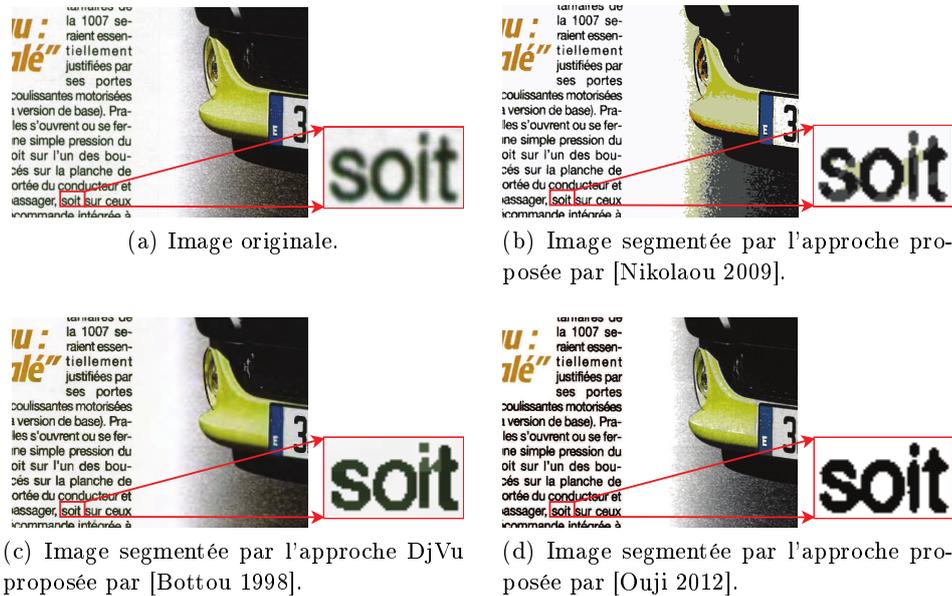


FIGURE 3.19 – Comparaisons de différentes méthodes de quantification dédiées aux images de documents. Crédits : [Ouji 2012] (Figure I.27)

du bruit et des structures graphiques. La taille des éléments structurants et des fenêtres utilisées peuvent poser problème pour une base de documents génériques. De plus, l'image étant traitée au niveau pixel, il faut ajouter à cela une étape de fusion pour éliminer les zones de transition qui auraient été détectées entre les caractères et le fond.

Identification de blocs d'éléments

Beaucoup de travaux sont dédiés à la segmentation de blocs d'éléments particuliers. Cela permet d'une part de réduire les temps de calculs en ne se focalisant que sur une partie des informations contenues par l'image. D'autre part, les processus utilisés peuvent être orientés pour les rendre plus spécialisés, ce qui améliore la performance de la segmentation. Pour finir, la détection des principales structures de l'image telles que les graphiques, qui sont très présents sur les documents administratifs, permettent ensuite de localiser plus facilement d'autres informations.

[Cohen 2013] ont utilisé des caractéristiques mesurées sur des superpixels pour extraire des zones de dessins sur des documents historiques grâce à un classifieur. Il y a encore peu de travaux en analyse de documents basés sur l'utilisation de superpixels. Ces méthodes semblent pourtant avoir un grand potentiel pour traiter ce type d'images. En effet, leur objectif est de produire un ensemble de régions locales respectant les caractéristiques géométriques de l'image. Le caractère structuré des documents pourrait se prêter au développement de ces approches.

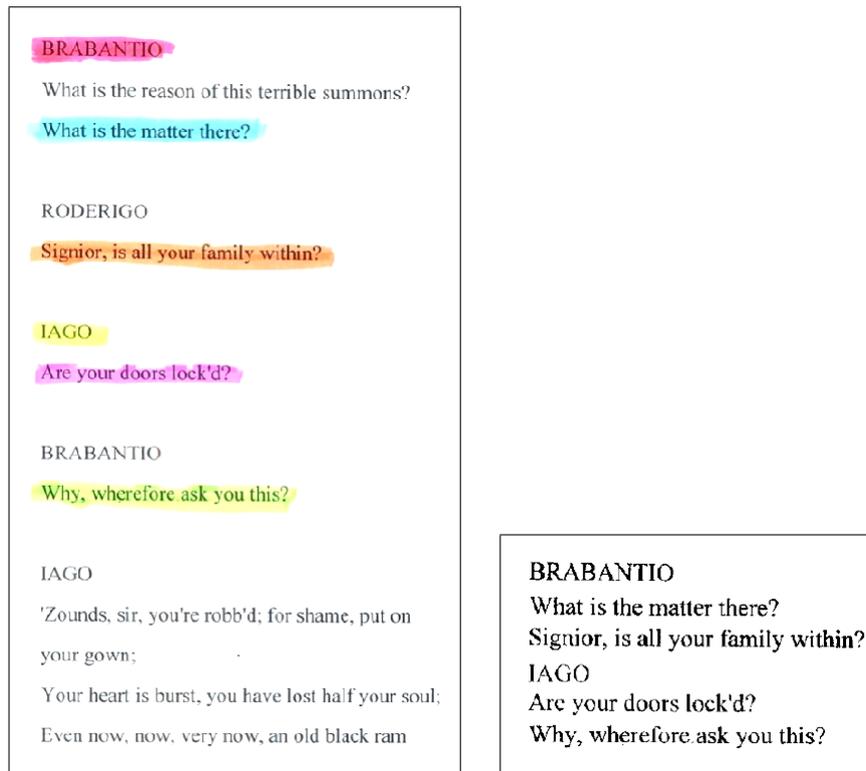
[Smigiel 2004] proposent une méthode de segmentation en quatre classes : fond, texte, texte coloré, texte du verso de l'image apparaissant par transparence. Pour cela, leur approche est basée sur la mise en place d'un réseau de neurones, et plus particulièrement sur l'utilisation des cartes auto-organisatrices de Kohonen. Une étape d'apprentissage est réalisée sur une portion représentative de la page à segmenter. Les caractéristiques utilisées sont les vecteurs constitués des trois composantes couleurs RVB des pixels. Les couleurs principales de l'image sont ainsi extraites. Elles sont ensuite fournies à un réseau de type perceptron qui se chargera de l'étape de classification. Cette approche donne de bons résultats, cependant elle est très spécialisée. La classification est supervisée et il n'est pas possible de mettre en place une telle phase d'apprentissage dans le contexte de nos travaux. En effet, les documents traités par cette approche ont tous des caractéristiques similaires, et le nombre de classes à extraire est bien défini, ce qui n'est pas le cas de notre base.

Les deux approches que nous venons de présenter concernent les images de documents historiques. Même si les documents administratifs ne présentent pas tout à fait les mêmes dégradations et caractéristiques visuelles en termes de couleurs, ces méthodes sont applicables aux documents administratifs sur lesquels nous retrouvons des éléments similaires. Une classification supervisée donne la plupart du temps de bons résultats. Cependant, en spécialisant les traitements, elle les rend aussi sensibles à des variations apparaissant dans la structure des éléments à segmenter.

Localisation d'éléments

Sur une image de document, la couleur sert souvent à guider la lecture, et à mettre en évidence certains éléments par rapport à d'autres. Il paraît donc intéressant d'utiliser les propriétés colorimétriques caractéristiques de certaines régions d'intérêt pour localiser des informations particulières.

Par exemple, il est fréquent de surligner les informations pour les mettre en évidence. Ce sont des données que le lecteur a jugé pertinentes et qu'il souhaite repérer facilement. [da Silva Barboza 2011] ont proposé une approche pour les détecter, et les utiliser comme base pour créer un résumé d'un bloc de texte. Ces travaux traitent des images originellement en noir et blanc, sur lesquelles apparaissent des zones surlignées, et qui ont été scannées en couleur. Les processus de segmentation sont appliqués dans l'espace RVB. Les auteurs considèrent que le surlignement affecte les pixels de l'image en diminuant la valeur de leur intensité. La détection des régions de surlignement consiste à mesurer pour chaque pixel les distances entre ses composantes couleurs deux à deux : rouge-vert, rouge-bleu, vert-bleu. Si l'une de ces distances est inférieure à un seuil d , le pixel correspondant est associé à une zone de surlignement. Celle-ci peut être considérée comme du bruit et corrigée en reconstituant le fond original. Pour cela, il suffit d'affecter aux trois canaux rouge, vert, et bleu, la valeur de la composante couleur ayant une intensité maximale (comparée



(a) Image originale.

(b) Zones de texte associées aux régions surlignées.

FIGURE 3.20 – Détection des zones de surlignement par l’approche proposée par [da Silva Barboza 2011], dans le but de créer un résumé d’un bloc de texte. Paramètres : seuil premier-plan/fond $s = 170$, distance $d = 35$ utilisée pour détecter un surlignement. Crédits : [da Silva Barboza 2011] (Figure 1)

aux deux autres). La segmentation du texte suit la même logique. Lorsqu’un pixel est détecté comme étant localisé dans une région surlignée, la valeur moyenne de ses composantes est seuillée par un paramètre s : inférieure à ce seuil, le pixel appartient au premier-plan, sinon il est affecté au fond. La figure 3.20 présente les résultats. Cette approche donne des résultats intéressants, cependant, elle ne peut être utilisée que dans le cas de documents noir et blanc, sur lesquelles ont été ajoutées des régions de couleurs, dont il faut être sûr qu’elles correspondent à des zones surlignées.

Récemment, [Hammami 2014] ont proposé une approche d’extraction de zones informatives dans des images de formulaires en couleur. L’idée est de détecter les zones rectangulaires de couleur homogène qui sont très fréquentes sur ce type de documents, et qui pourraient ensuite servir de points de départ pour localiser les informations textuelles recherchées (Figure 3.21). Le bruit est tout d’abord filtré. Puis, une étape de quantification permet de réduire le nombre de couleurs différentes à K . En se basant sur les caractéristiques particulières des formulaires, les auteurs

ont supposé que $K = 3$: une couleur pour l'arrière-plan, une pour les zones informatives, et une dernière pour le texte. Pour terminer, les composantes connexes sont extraites sur les K couches, puis filtrées selon un critère de rectangularité. Cette approche est intéressante et semble donner de bons résultats. Cependant, elle est dédiée à une famille d'images spécifique, pour laquelle le nombre de couleurs est connu.

FIGURE 3.21 – Détection des rectangles. Crédits : [Hammami 2014] (Figure 4)

3.7 Conclusion

A travers cet état-de-l'art nous avons décrit les principaux mécanismes de traitement de la couleur en analyse d'images. Les systèmes de codage de l'information couleur numérique imitent le fonctionnement de la vision humaine. Des espaces couleurs ont ainsi été créés en se basant notamment sur la théorie trichromatique et sur la théorie des couleurs opposées. De nombreux systèmes ont été proposés, mais il est admis qu'il n'existe pas de modèle de représentation générique. Le choix d'un espace est généralement réalisé expérimentalement en fonction de l'application et des données disponibles. De plus, la couleur peut être manipulée selon une approche marginale ou vectorielle, orientant ainsi le choix des traitements.

La stratégie d'analyse est elle aussi fortement dépendante des images à traiter. Les techniques de segmentation couleur sont réparties en deux classes selon qu'elles combinent ou non l'information spatiale avec l'information colorimétrique. Une approche utilisant uniquement l'information colorimétrique, telle que celles basées sur

les histogrammes ou les techniques de regroupements, requiert des classes bien séparées. Ce type d'approche peut donc ne pas être suffisamment robuste au bruit dans un contexte de segmentation de documents administratifs du fait de leur grande variabilité informationnelle. L'intégration de données spatiales permet d'améliorer les résultats. La combinaison de données globales et locales peut aussi affiner la segmentation, en rendant l'approche plus adaptative aux variations locales. Cependant, quelle que soit la technique choisie, il faut la plupart du temps prêter attention au choix des paramètres, ce qui entraîne un manque de généralité de la méthode.

Le traitement de la couleur passe la plupart du temps par une étape de quantification qui a pour objectif de réduire la quantité de couleurs observables sur l'image à un nombre raisonnable. Pour cela, ces techniques tentent de partitionner l'espace couleur. Les approches basées sur des divisions successives sont les plus simples et les plus rapides. Leur inconvénient majeur est qu'elles ne tiennent pas compte de la distribution des points dans l'espace. Or, celle-ci est rarement uniforme, ce qui peut conduire à des erreurs de segmentation. Des approches basées sur des techniques de classification peuvent pallier à ce problème. Les temps de calculs plus importants peuvent cependant être un frein à leur utilisation. Enfin, des approches de nommage de couleurs permettent d'associer un terme de couleur à une valeur numérique, et peuvent être considérées en cela comme des techniques de quantification. Le bénéfice d'utiliser des étiquettes aisément compréhensibles est certain en termes de communication. La difficulté majeure est la définition d'un modèle pertinent et adapté aux images à traiter. Le choix du nombre de couleurs, et la définition des frontières est important, et doit être fait judicieusement. Cela peut rendre ces approches peu flexibles dans un contexte de segmentation. Cependant, dans un contexte d'extraction des couleurs dominantes, il peut être intéressant de les utiliser à un plus haut niveau, pour la création de descripteurs afin d'étiqueter les régions obtenues par une technique de segmentation.

Nous avons ensuite réalisé un inventaire des travaux traitant les images de documents couleur. Au contraire d'une image de scène naturelle, une image de document a un caractère structuré sur lequel il est possible de s'appuyer. La communauté s'intéresse de plus en plus à ce type d'approches, et quelques solutions ont été proposées récemment. Il n'existe cependant aucune méthode complètement générique, et non supervisée. La difficulté majeure provient de la complexité de l'information couleur, et de la richesse des images de documents en termes de contenus. Un certain nombre de problèmes concernant la segmentation a été identifié. Tout d'abord, une séparation fond/premier-plan telle qu'elle est traditionnellement appliquée est très restrictive. En effet, les documents contiennent des éléments de différentes couleurs aussi bien dans les régions de fond, que de premier-plan. D'autre part, les zones de texte apparaissant sur des fonds texturés sont particulièrement difficiles à identifier et à extraire. En effet, le texte a des caractéristiques de taille, et de couleurs très proches de celle du bruit. De nombreuses approches confrontées à ce problème doivent appliquer des post-traitements pour restaurer les caractères.

Néanmoins, les approches existantes ont montré que l'information couleur est une donnée importante qui peut apporter une plus-value à l'étape de segmentation. Parmi toutes les applications possibles, il nous a paru intéressant de nous pencher sur l'extraction des couches couleur principales qui peuvent être observées sur une image de documents. L'idée est de regrouper tous les éléments ayant des teintes très proches visuellement afin d'identifier un nombre réduit de couleurs, permettant ainsi de simplifier le document tout en conservant l'information couleur. Les possibilités sont ensuite multiples : quantification, compression, description, etc.

Les verrous scientifiques qui ont été identifiés portent sur plusieurs points. Le plus important concerne l'extraction de manière adaptative et non supervisée d'un petit nombre de couleurs principales. La généralité qui nous est imposée par le contexte industriel complexifie la tâche. Le bruit présent sur les images est une donnée importante à prendre en compte. Un autre enjeu important porte sur la qualité de la segmentation elle-même. Pour pouvoir être exploitées, les couches extraites doivent conserver la segmentation logique des régions réelles de l'image (ex : caractères ; mots, graphiques). Là aussi, ne pas utiliser de connaissance *a priori* permettant de reconnaître ces régions est un véritable challenge.

Au cours de ces travaux, nous avons implémenté différentes approches de manière linéaire. Celles-ci ont peu à peu formé une chaîne d'extraction des couches couleur que nous allons décrire dans la suite de ce mémoire. Pour terminer, il n'existe pas à notre connaissance de bases de référence officielles pour comparer les résultats de la segmentation d'images de documents en couches couleur. Nous avons donc mis en place un protocole d'évaluation basé sur l'utilisation d'images synthétiques simulant les caractéristiques propres aux images de documents.

Deuxième partie

Segmentation d'images de
documents en couches couleur

Jeu de données pour validations expérimentales

Sommaire

4.1	Introduction	73
4.2	Evaluation de la segmentation en analyse de documents	74
4.3	Base de documents réels	76
4.3.1	Données	76
4.3.2	Evaluation sur des données réelles	77
4.4	Constitution d'une base de vérité-terrain de documents synthétiques	78
4.4.1	Données	78
4.4.2	Simuler les dégradations causées par le bruit	80
4.4.3	Mesures d'évaluation	82
4.5	Vers un modèle de dégradation des caractères applicable aux images de documents couleur	83
4.5.1	Introduction	83
4.5.2	Modèle de dégradation des caractères	84
4.5.3	Résultats et discussion	91
4.5.4	Conclusion sur le modèle de bruit	91
4.6	Conclusion sur le jeu de données	94

4.1 Introduction

L'objectif de nos travaux est la segmentation de documents administratifs en couches couleur, proposée comme une alternative à l'étape de binarisation. Nous définissons une couche couleur comme un ensemble de pixels ayant à la fois une apparence colorimétrique très proche et présentant une unité logique ou sémantique. Pour respecter le système de représentation traditionnel, les méthodes implémentées fournissent en sortie un ensemble d'images binaires qui font apparaître les pixels de la couche en noir sur un fond blanc.

Au cours de ces travaux de recherche, nous avons proposé différentes approches qui ont été développées de manière linéaire. Leurs performances ont été évaluées en suivant le même protocole. C'est pourquoi nous avons choisi d'explicitier avant

toutes choses le contexte général d'évaluation (méthodes et données). Les expérimentations propres à chacune des méthodes seront décrites dans les chapitres 5, 6, et 7.

Il est important de pouvoir mesurer quantitativement les performances de différentes techniques pour pouvoir les comparer. Cette mesure peut être réalisée soit sur une base de documents réels annotés, soit sur un ensemble d'images synthétiques. Or, la génération de grands ensembles de vérité-terrain annotés est un challenge comme nous avons pu le mentionner dans les conclusions de nos états-de-l'art. Nous commencerons donc par présenter brièvement les principaux concepts de création de bases d'images qui sont utilisées en analyse de documents dans un but d'évaluation.

La seconde partie de ce chapitre introduira les images de documents réels qui sont à notre disposition et qui proviennent de différentes sources (section 4.3). Les travaux portant sur l'analyse de documents couleurs sont encore très récents. Il n'existe pas à notre connaissance d'ensembles d'images annotées permettant d'évaluer une segmentation en couches couleur. La création de bases d'images pouvant servir de référence à la communauté est une problématique essentielle que nous proposons d'adresser ici. C'est pourquoi nous avons produit une base d'images synthétiques (section 4.4). La section 4.4.3 présentera les mesures d'évaluation.

Pour finir, une évaluation sur des images synthétiques nécessite que ces dernières soient proches de la réalité en termes de structures, mais aussi en termes de dégradation. Des modèles de bruit existent et permettent de simuler ces dégradations. Au cours de cette thèse, nous avons participé à des travaux collaboratifs dont l'objectif était la proposition d'un modèle de dégradation des caractères qui serait applicable aux images de documents couleur. La présentation de ce modèle conclura ce chapitre.

4.2 Evaluation de la segmentation en analyse de documents

Il est important de pouvoir évaluer la performance des algorithmes de segmentation sur des données réelles. Cette étape est généralement faite sur des images annotées. La difficulté majeure concerne la mise en place de protocoles d'évaluation pertinents. Il faut déterminer préalablement quelles sont les images concernées, les limites de cette évaluation, et quelle forme doit prendre la vérité-terrain : boîtes englobantes, régions réelles, niveau pixel, etc. La réponse à ces questions aura un impact important sur le choix de la mesure d'évaluation, et elle sera bien entendu déterminée par l'application et les données disponibles.

Le processus de création d'une vérité-terrain étant difficile, la mise à disposition, lorsque cela est possible, de bases publiques qui puissent servir de référence pour la communauté est un apport non négligeable. Cela est d'autant plus essentiel que ces

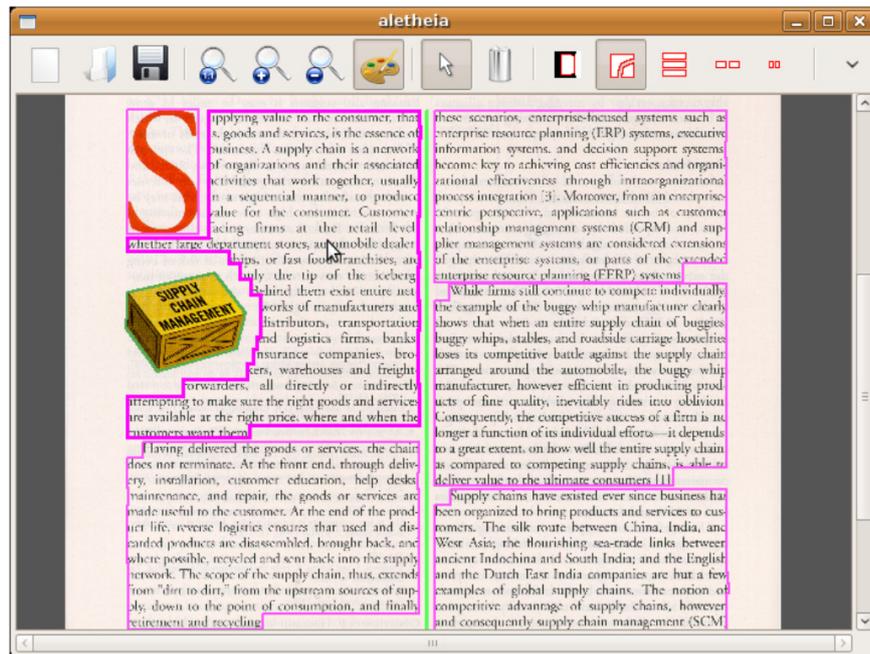


FIGURE 4.1 – Vérité-terrain constituée de régions de forme complexe qui ont été extraites grâce au logiciel *Aletheia*. Crédits : [Antonacopoulos 2009a] (Figure 4)

ensembles sont utilisés pour évaluer et comparer différentes approches sur le même jeu de données. Plusieurs bases d'images ont ainsi été proposées :

- UWASH (Université de Washington,[Philips 1993]), contient des documents simples principalement de type journaux en noir et blanc, dont la plupart ont été générés synthétiquement ; les régions sont décrites par leurs boîtes englobantes ;
- ISRI ([Nartker 2005]), contient principalement des images en noir et blanc de journaux ; les régions sont décrites par leurs boîtes englobantes ;
- MARG (Medical Article Records Groundtruth en anglais, [The National Library of Medicine 2010]), contient principalement des images en noir et blanc de journaux ; les régions sont décrites par leurs boîtes englobantes ;
- MediaTeam Document Database ([Sauvola 1998]), contient divers documents couleur avec une structure relativement simple ; les régions sont décrites par leurs boîtes englobantes ;
- Uva Document Dataset ([Todoran 2005]), contient principalement des pages de publicités ou de magazines, orientée pour la segmentation couleur.

Récemment, le groupe PRImA¹ a mis à disposition plusieurs outils d'annotations permettant de générer des ensembles de vérité-terrain sur des documents divers.

1. Pattern Recognition & Image Analysis en anglais : <http://www.prima.cse.salford.ac.uk>

Nous pouvons citer entre autres le logiciel *Aletheia* (figure 4.1), qui permet l'extraction de zones ayant des formes complexes. L'objectif est d'annoter ces régions avec des étiquettes telles que texte, images, graphiques, etc. Les ensembles ainsi construits ont été rendus public. Par exemple, la base présentée par [Antonacopoulos 2009a] regroupe des documents ayant des structures diverses avec un intérêt plus particulier pour les magazines, et les journaux techniques. Ces images proviennent de différentes sources.

Pour avoir un intérêt, un ensemble de vérité-terrain doit aider à résoudre les verrous scientifiques. Le choix des images de documents est particulièrement important. Actuellement, les bases se constituent donc plutôt autour de documents complexes, qui ont des structures hétérogènes, car ces images posent encore problème. La vérité-terrain a pour objectif de décrire les résultats qui devraient être obtenus dans un cas de segmentation idéale. L'objectif est bien sûr l'évaluation et la comparaison de différentes approches. La base de PRImA a ainsi été utilisée lors d'une compétition d'ICDAR (2009, [Antonacopoulos 2009b]) pour évaluer des méthodes de segmentation de pages.

Le processus d'annotation de documents réels est long et contraignant. L'utilisation de bases d'images synthétiques est une autre solution pour évaluer les performances d'une segmentation. L'idée est de générer des images qui vont imiter la structure de vrais documents. L'application d'un modèle de bruit simule ensuite les dégradations subies par les images. Il en existe plusieurs. Nous pouvons citer entre autres le modèle de Kanungo ([Kanungo 1993]) qui simule le bruit introduit au cours de la numérisation de documents en noir et blanc, et l'approche proposée par [Kieu 2012] imitant des dégradations (de différents types, et en quantité variable) qui sont observables sur des images de documents anciens en niveaux de gris. L'avantage principal de ce type d'évaluation est qu'il est possible de produire des images synthétiques en très grandes quantités et dans des conditions contrôlées. Le choix du modèle de bruit doit cependant être adapté à l'application souhaitée. Les images de documents subissant des dégradations de plusieurs types, l'étape de paramétrage peut être complexe.

La section suivante sera consacrée à la description des documents réels de notre base de test.

4.3 Base de documents réels

4.3.1 Données

Nous disposons d'un ensemble de documents administratifs réels scannés à une résolution de 200 ou 300 dpi, fournis par notre partenaire industriel. Nous n'avons eu aucun contrôle sur le processus d'acquisition. De plus, certains documents provenant d'un fournisseur donné ne nous ont été transmis qu'en un seul exemplaire. Les images de documents ont été caractérisées dans le chapitre 2, dans la section

2.2. Pour illustrer la diversité des images à notre disposition, quelques exemples sont présentés dans l'annexe B.

Ces documents qui peuvent être qualifiés de complexes proviennent de différentes sources. Ils sont très hétérogènes en termes de contenus : texte manuscrit et imprimé, différentes fontes, des éléments de différentes tailles, et qui peuvent éventuellement se chevaucher. Leur structure est la plupart du temps de type non-Manhattan. Le nombre de couleurs apparaissant sur ces images est très variable d'un document à un autre. Les régions de fond peuvent être homogènes mais aussi plus complexes avec des zones texturées ou des dégradés de couleurs. L'enjeu de nos travaux est de pouvoir traiter un document, même rencontré pour la première fois, sans connaissance *a priori*. Nous ne disposons donc pas de modèles de documents.

Pour information, le tableau 4.1 en présente une liste non-exhaustive avec les quantités approximatives à notre disposition. Les factures et les bons de commandes ont été comptés dans une seule catégorie car ils ont des caractéristiques structurelles très proches : données apparaissant dans un tableau, blocs d'adresses, etc. Comme nous pouvons le voir, notre base est très largement constituée de factures qui représentent une part non négligeable des documents traités par les industriels. Les formulaires sont aussi très représentés. On les retrouve en particulier dans les administrations publiques. Ces images proviennent de nombreuses sources différentes : approximativement plus de 250.

Catégories de documents	Quantités approximatives
factures et bons de commandes	7300
formulaires (sécurité sociale, impôts, etc)	1200
chèques	3

TABLEAU 4.1 – Ensemble de documents fournis par notre partenaire industriel.

4.3.2 Evaluation sur des données réelles

Il n'existe pas à notre connaissance de bases publiques de documents administratifs couleurs annotés. Les contraintes de confidentialité liées au contexte industriel limitent en effet la divulgation de ce type de données. De plus, les travaux traitant de la segmentation d'images de documents couleurs sont encore relativement récents. Nous avons donc sélectionné un échantillon de 150 images dans la base de données qui a été présentée précédemment dans la section 4.3.1. Les images ont été choisies de façon à reproduire les contraintes et les difficultés rencontrées par les industriels : éléments de fond et de premier-plan de différentes couleurs, textures, dégradés de couleurs, surcharges, chevauchements, etc.

Cet ensemble a été utilisé pour réaliser une évaluation visuelle de la segmentation. Nous nous sommes notamment attachés à la détection d'un nombre de couches couleur raisonnable, tout en nous assurant que l'image ne soit ni trop sur-segmentée, ni trop sous-segmentée. Nous avons prêté une attention particulière à la segmentation des éléments fins en veillant surtout à ce que les régions textuelles restent lisibles, et puissent être exploitées par un outil de reconnaissance.

Des mesures plus précises requièrent des documents annotés. Cependant, la phase d'annotation est un processus long et coûteux à mettre en place : création du protocole, mise en place des outils, annotations faites par plusieurs personnes, questions de la précision et de la validation des annotations, etc. Par conséquent, l'annotation manuelle ne permet pas la création de grands jeux de données. De plus, notre contexte introduit des problèmes supplémentaires liés à l'interprétation obligatoirement subjective d'une couche couleur. En effet, les documents doivent être annotés en termes de couleurs tout prenant en compte les différentes structures logiques.

N'ayant pas à notre disposition de vérité-terrain constituée d'images de documents réels annotées, nous avons choisi de générer automatiquement des images synthétiques afin d'évaluer quantitativement les performances des approches de segmentation. Ces images synthétiques ont l'avantage de pouvoir être produites en très grand nombre, et grâce à un protocole contrôlé. Le processus de génération sera décrit dans la section 4.4. Cette évaluation complètera l'évaluation visuelle réalisée sur la base de documents réels.

4.4 Constitution d'une base de vérité-terrain de documents synthétiques

4.4.1 Données

L'utilisation d'images synthétiques autorise la génération automatique de bases de test de grande taille dans des conditions contrôlées. Elles résolvent ainsi les problèmes, notamment de subjectivité, entraînés par l'annotation manuelle de documents réels. Cependant, la création de bases synthétiques soulève des questions importantes. La principale difficulté consiste à générer des images reflétant la réalité, que ce soit en termes de structures, ou de dégradations.

Pour ces travaux, nous avons généré 2000 images synthétiques pour lesquelles l'information couleur est disponible à un niveau pixel. Cette base est à disposition de la communauté² depuis février 2015. La vérité-terrain associée à chaque document synthétique est un ensemble d'images binaires correspondant aux différentes couches couleur présentes sur cette image. Pour une couche donnée, tous les pixels ayant la couleur considérée apparaissent noirs sur un fond blanc (figure 4.2). La représenta-

2. <http://navidomass.univ-lr.fr/ColorSegmentationGT/>



FIGURE 4.2 – Images synthétiques et couches associées aux couleurs rose, vert et marron.

tion historique est ainsi préservée. L'information couleur portée par chaque couche est la moyenne des valeurs des composantes couleurs de chaque pixel appartenant à cette couche.

Deux modèles ont été créés pour obtenir des images semi-structurées ressemblant à des documents réels : une image simulant un courrier (figures 4.3a, 4.3b, 4.3d et 4.3f), et une autre contenant uniquement du texte sur un fond (figures 4.3c et 4.3e). Ces modèles ont été utilisés en proportions égales dans la base de test. Les éléments qui sont initialisés aléatoirement sont : l'affichage ou non de certaines zones, les couleurs, le texte, ainsi que la taille et la forme des éléments. Un échantillon d'images synthétiques est présenté sur la figure 4.3.

Certaines images de documents ont été générées plusieurs fois afin de simuler des documents provenant d'une même source. Nous simulons également la variabilité que l'on trouve dans les documents réels. Ces images contiennent ainsi une partie fixe, propre à la famille du document, et une partie variable spécifique à l'image considérée. En pratique, la partie fixe est générée aléatoirement une seule fois pour une famille de documents donnée. Puis, la partie variable est ajoutée sur chaque image en faisant varier le contenu toujours aléatoirement (voir figures 4.3a et 4.3b). En effet, les entreprises reçoivent fréquemment des documents provenant d'un même fournisseur. La vérité-terrain reflète ainsi la réalité. Cette fonctionnalité n'a pas été directement exploitée lors de ces travaux. Cependant, une des perspectives futures possibles est le développement d'un système qui serait capable de reconnaître un document déjà rencontré, et d'utiliser les résultats obtenus pour guider le processus de segmentation de la nouvelle image.

Pour résumer, nous avons un ensemble constitué de 2000 images synthétiques formé de 1000 images de lettres, et 1000 images de texte sur un fond coloré. Les paramètres ont été initialisés de telle façon que les images contiennent au minimum deux couleurs différentes (une de fond, et une de premier-plan), et jusqu'à N couleurs. N est un maximum que nous avons fait varier pour chaque famille de documents. Il permet de créer des listes de couleurs qui sont ensuite affectées de manière aléatoire

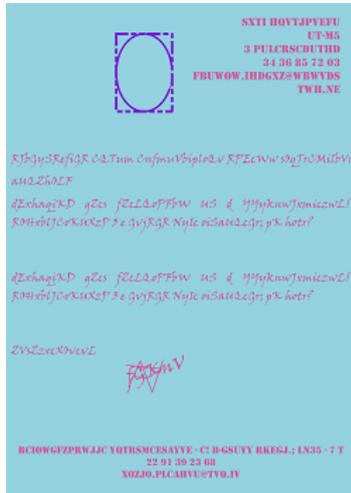
aux éléments. En pratique, le nombre de couleurs total est souvent inférieur à la valeur choisie. La plupart de nos images contiennent moins de 6 couleurs. Cependant, tout l'intérêt d'utiliser un logiciel de génération automatique d'images synthétiques est qu'il peut être facilement adapté pour générer des modèles différents. De plus, il est possible de constituer de nouvelles bases en faisant varier les paramètres, et en générant un nombre indéfini d'images. Des travaux futurs ont été identifiés pour simuler des images de documents plus complexes : formulaires, factures, etc.

4.4.2 Simuler les dégradations causées par le bruit

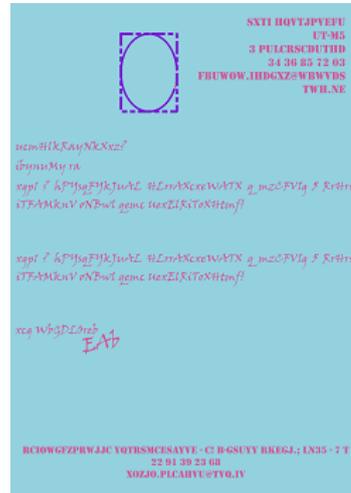
Telles quelles, les images synthétiques ne permettent pas de simuler les dégradations subies par la couleur qui sont observables sur une image de document réel. Pour conduire les expérimentations, les images de la base ont été bruitées. Le modèle de génération de bruit le plus reconnu en analyse de documents est celui proposé par [Kanungo 1993] qui est applicable sur des images binaires. L'idée est d'ajouter du bruit de type *poivre et sel* sur les caractères de l'image, puis de les lisser, afin de simuler leur dégradation. Il n'existe actuellement pas de modèle simulant le bruit apparaissant sur les images de documents couleurs.

Pour introduire du bruit sur nos images synthétiques, nous avons utilisé un logiciel de traitement d'images (Photoshop) et le processus a été appliqué de manière identique sur la totalité des images de notre ensemble. Nous avons choisi ici de sauvegarder les images avec une compression jpeg élevée afin de respecter les contraintes auxquelles est soumis notre partenaire industriel dans son flux de dématérialisation (figure 4.4). Cette opération a alors simulé les artéfacts qui apparaissent sur les images réelles sous forme de blocs. Cependant, dans le cas des images synthétiques qui ne contiennent pas de bruit à l'origine, ces zones de bruit n'apparaissent que près des contours. C'est pourquoi nous avons choisi de faire précéder cette étape de compression par l'ajout d'un bruit Gaussien appliqué sur la totalité de l'image. Ainsi, toutes les régions de l'image sont soumises à cette dégradation, qui permet de simuler les fluctuations aléatoires subis par les pixels de l'image lors de l'impression et de la numérisation.

Un bruit gaussien n'est pas forcément représentatif de la réalité. Néanmoins, la combinaison de ce filtre relativement simple, et de la compression, a produit des résultats visuellement très proches du bruit que nous pouvons observer sur les images de notre base, et que nous avons jugé satisfaisants pour une première évaluation de la segmentation. En parallèle de ces travaux de recherche, nous avons, dans le cadre du stage de Master de Mlle Do Thy Luyen, proposé un modèle de dégradation des caractères applicable aux images de documents couleur. Ce modèle étend l'approche proposée par [Kieu 2012] pour simuler le bruit observable sur des images de documents anciens en niveaux de gris. Cette méthode est elle-même basée sur le modèle de Kanungo ([Kanungo 1993]), qui est applicable sur des images binaires. L'approche proposée sera décrite plus en détails dans la section 4.5.



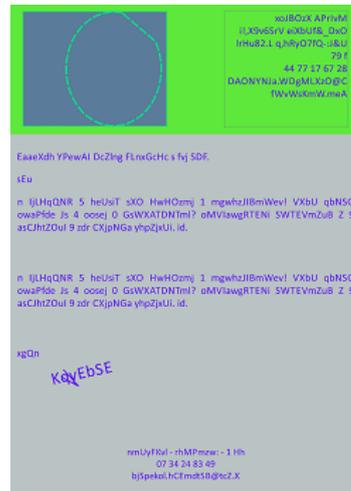
(a)



(b)



(c)



(d)



(e)



(f)

FIGURE 4.3 – Images synthétiques générées automatiquement.

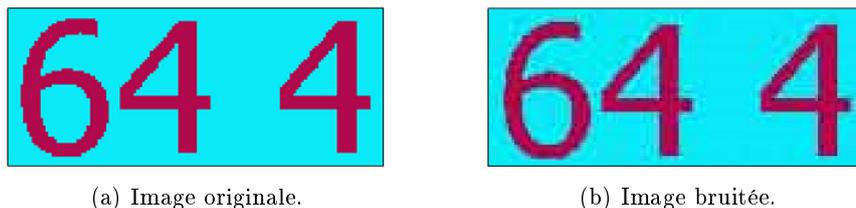


FIGURE 4.4 – Ajout de bruit sur les images synthétiques.

4.4.3 Mesures d'évaluation

Les mesures d'évaluation sont liées aux caractéristiques de la base de vérité-terrain, et au contexte d'application. Lorsqu'il est mis à disposition du public, le jeu de données est par ailleurs généralement fourni avec un protocole d'évaluation détaillé, ceci afin de réaliser des comparaisons objectives des approches testées. Nous allons introduire ici les principales approches d'évaluation appliquée en segmentation d'images de documents.

Par exemple, dans sa thèse, [Ouji 2012] a proposé une segmentation des images de documents en différentes couches en fonction de leurs propriétés colorimétriques. Des images réelles ont tout d'abord été étiquetées manuellement, ce qui a permis d'évaluer quantitativement l'approche grâce aux mesures de rappel et de précision. Devant la difficulté d'annoter ainsi manuellement une base conséquente de documents, l'auteur a décidé de réaliser une évaluation indirecte en mesurant les performances d'applications utilisant son approche. Dans le cas d'une application de lecture de texte, il est souvent intéressant d'étudier les taux de reconnaissance obtenus par un outil de reconnaissance (ROC, RIC). Par conséquent, la procédure d'évaluation a été conduite sur les résultats obtenus par une approche de détection des lignes de texte, comparés à ceux obtenus à partir du logiciel de reconnaissance *FineReader*, et évalués en termes de précision et de rappel.

Lors de la compétition concernant la segmentation de page organisée en 2007 pour ICDAR, la méthode d'évaluation était elle aussi basée sur une procédure classique consistant à mesurer la précision et le rappel au niveau pixel ([Antonacopoulos 2007]). En 2009, pour cette même compétition, les organisateurs ont opté pour une évaluation plus précise prenant en compte le contexte à travers la définition de scénarios ([Antonacopoulos 2009b]). L'objectif était l'extraction des classes suivantes : texte, graphiques, lignes, séparateurs et bruit. Pour évaluer les résultats, les correspondances entre les régions segmentées par la méthode testée, et les régions associées à la vérité-terrain sont tout d'abord extraites. Puis, les erreurs sont détectées et classifiées en plusieurs catégories. L'idée est que la qualité d'une segmentation dépend aussi bien du nombre d'erreurs, que du type de ces erreurs. En effet, en fonction du but recherché, celles-ci peuvent être jugées acceptables ou non. Par exemple, manquer un numéro de page est pénalisant dans un système d'indexation. En revanche,

la fusion de deux paragraphes localisés dans une même colonne n'est pas bloquant pour leur lecture. Finalement, les différents scénarios sont résumés ainsi :

- une région de la vérité-terrain est correctement détectée ;
- une région segmentée ne correspond à aucune région de la vérité-terrain ;
- une région de la vérité-terrain a été complètement ou partiellement détectée par plusieurs régions au cours de la segmentation (région divisée) ;
- plusieurs régions de la vérité-terrain ont été complètement ou partiellement détectées par une unique régions au cours de la segmentation (régions fusionnées) ;
- une région de la vérité-terrain n'a pas été complètement voire pas du tout détectée lors de la segmentation.

Nous avons choisi d'évaluer les résultats de notre segmentation au niveau pixel car notre objectif est une segmentation bas-niveau. Nous n'avons en effet à notre disposition aucune connaissance *a priori* permettant de caractériser les éléments sémantiques d'une image de document réel (tableau, graphiques, etc). En revanche, la vérité-terrain associée aux images synthétiques nous permet de connaître la couleur de chaque pixel. Notre évaluation est donc basée sur les mesures de rappel et précision.

Soient R , les pixels détectés sur l'image résultat, et G , ceux détectés sur la vérité-terrain. Pour chaque couche couleur de la vérité-terrain, nous recherchons l'image la plus probable parmi les résultats de la segmentation, c'est à dire celle ayant le plus de pixels en commun avec la couche annotée. Nous mesurons ensuite la précision et le rappel de la manière suivante sur les 2000 images synthétiques de notre base de vérité-terrain :

$$Précision = \frac{card(R \cap G)}{card(R)} \quad (4.1)$$

$$Rappel = \frac{card(R \cap G)}{card(G)} \quad (4.2)$$

Ces métriques ont été utilisées pour évaluer les contributions scientifiques de ce manuscrit : deux approches colorimétriques, une approche spatio-colorimétriques et pour terminer, une approche basée sur une analyse multi-résolution. Elles seront présentées successivement dans les chapitres 5, 6 et 7.

4.5 Vers un modèle de dégradation des caractères applicable aux images de documents couleur

4.5.1 Introduction

La question de l'évaluation d'une approche de segmentation est un problème complexe. Il n'est pas toujours possible de mettre en place un processus d'annotation manuelle pour des raisons organisationnelles évidentes. L'utilisation d'images

synthétiques est une solution qui autorise la création de grandes bases de tests dans des conditions contrôlées. Les processus d'analyse de documents réels sont confrontés à la présence de bruit. Des dégradations parfois importantes peuvent être observées et en particulier près des zones textuelles qui sont des régions de hautes variations (gradient fort). Il est donc important que la vérité-terrain permettant l'évaluation des traitements reflète ces contraintes.

A la fin de cette thèse, des travaux préliminaires ont été réalisés en collaboration avec Mlle Do Thi Luyen dans le cadre de son stage de master, afin de proposer un modèle de dégradation des caractères applicable aux images de documents couleur. Ce modèle étend au traitement des images couleur, celui proposé par [Kieu 2012] et [Kieu 2013], pour simuler, sur des images en niveaux de gris, des dégradations observables sur des documents anciens (encre qui bave, tâches sombres/clairées proches de caractères, etc). Notre choix s'est porté sur cette approche car les effets de ces dégradations sont similaires au bruit présent sur nos images de documents administratifs.

Ces travaux concernent une autre partie du projet "*analyse de documents couleur*" qui est plus focalisée sur l'évaluation de la qualité des images. Ceux-ci sont actuellement en cours, et nécessitent notamment une étape d'évaluation et de validation plus poussée. Nous ne les avons pas mis en application dans le processus d'évaluation des approches de segmentation qui sont présentées dans ce mémoire. Cependant, les premiers résultats ouvrent des perspectives intéressantes pour la génération d'outils de vérité-terrain plus performants dans des travaux futurs. Pour plus de détails, le lecteur est invité à se référer à l'article original ([Luyen 2015]). Nous présenterons tout d'abord notre modèle, puis les résultats obtenus, et pour terminer les perspectives de ces travaux.

4.5.2 Modèle de dégradation des caractères

Proposition

Comme nous l'avons vu dans la section 3.6.1 de ce mémoire, différents processus introduisent du bruit sur les images de documents, et dégradent ainsi la qualité de l'image. Ce bruit est particulièrement gênant lorsqu'il apparaît au niveau des régions textuelles (ex : caractères partiellement ou pas segmentés), car il peut impacter la performance des outils de reconnaissance. Plusieurs modèles ont donc été proposés pour simuler la dégradation des caractères.

L'une des approches de dégradation des caractères les plus reconnues en analyse de documents a été proposée par [Kanungo 1993]. Ce modèle statistique permet de simuler le bruit *poivre-et-sel* que l'on peut observer sur les images de documents imprimés, scannés et numérisés. Il est appliqué sur des images binaires. L'idée de base est de changer aléatoirement la valeur des pixels : un pixel blanc devient noir, et

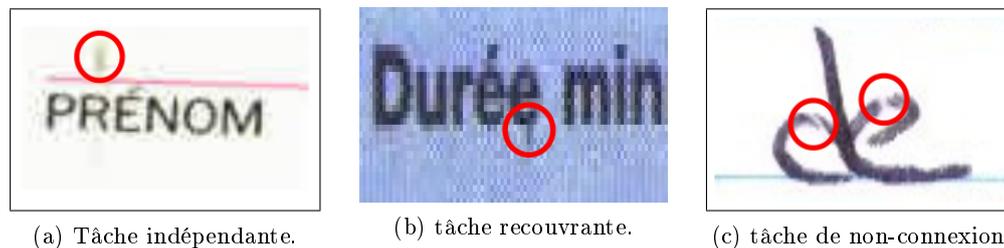


FIGURE 4.5 – Exemples de différents types de bruit observables sur les images de documents couleur. Crédits : Do Thi Luyen

inversement. Le modèle repose sur l’hypothèse que la probabilité que cette inversion se produise pour un pixel donné ne dépend que de la distance entre ce pixel et le contour du caractère le plus proche.

L’approche de Kanungo nécessite l’initialisation de 8 paramètres. Récemment, dans [Kieu 2012] et [Kieu 2013], les auteurs ont proposé d’étendre ce modèle aux images en niveaux de gris, afin de simuler les dégradations observables sur des documents anciens. Leur méthode automatise l’étape de paramétrage en laissant la possibilité à l’utilisateur de choisir la quantité et la catégorie de bruit qui doit être généré. En effet, les documents sont soumis à des dégradations de différents types. L’objectif est de produire des images qui soient le plus proches possible de la réalité, et qui soient adaptées à l’application de vérité-terrain souhaitée.

Les dégradations sont classées en trois classes :

- les tâches indépendantes correspondent au bruit qui apparaît à l’intérieur ou à l’extérieur d’un caractère (Figure 4.5a) ;
- les tâches recouvrantes sont les régions de bruit qui chevauchent un caractère et modifient ainsi son contour (Figure 4.5b) ;
- les tâches de non-connexion cassent la connexité d’un caractère (Figure 4.5c).

Une zone de dégradation correspond à une région elliptique dont le centre est appelé *germe* ou *point de dégradation*. Ces germes sont extraits selon l’algorithme pseudo-aléatoire proposé par [Kanungo 1993] dont les paramètres sont déterminés automatiquement. La taille et la direction de l’ellipse dépendent de la valeur du gradient en son centre. Concrètement, dans la modélisation proposée par [Kieu 2013], l’utilisateur a la main sur la quantité de bruit (nombre de germes générés), et sur le type de bruit (pourcentage de chaque catégorie de bruit). Des règles heuristiques sont utilisées dans le processus de classification de chaque point de dégradation parmi les catégories de bruit possible. L’attribution d’un type de bruit à chaque germe dépend de son contenu et de son voisinage. Finalement, les pixels localisés à l’intérieur de l’ellipse sont modifiés de façon à simuler l’effet de la dégradation.

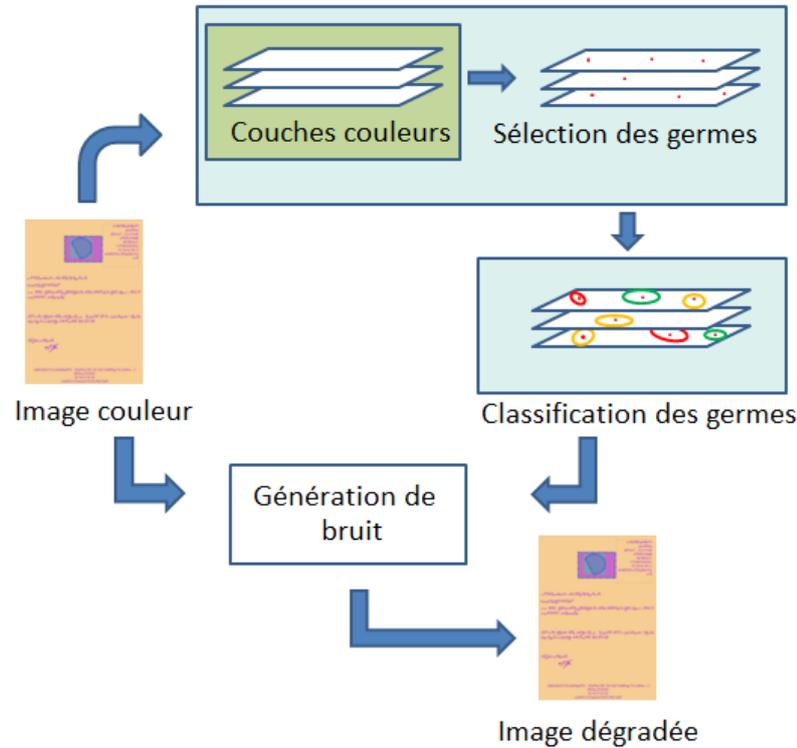


FIGURE 4.6 – Processus complet de dégradation des caractères appliqué sur des images de documents couleur. Crédits : Do Thi Luyen

Les étapes du processus de dégradation peuvent être résumées ainsi :

- extraction des points de dégradation (les germes) ;
- détermination du type de dégradation associé à chaque germe ;
- modification des pixels localisés à l'intérieur de la zone de dégradation.

L'approche que nous proposons ici a pour objectif d'étendre ce modèle au traitement des images de documents couleur. Pour intégrer l'information colorimétrique dans le processus, nous avons proposé d'utiliser les travaux présentés dans ce mémoire, dont le but est la segmentation de documents administratifs en couches couleur. L'idée est d'extraire les couleurs principales observables sur un document sous forme d'un ensemble d'images binaires. Ces images binaires peuvent donc être fournies en entrée des modèles de bruit existants. L'originalité du modèle de dégradation que nous proposons est que les germes sont extraits sur chacune des couches couleur ainsi détectées. Ces germes sont ensuite classifiés et le bruit est généré en tenant compte de l'information couleur. La figure 4.6 résume l'ensemble du processus.

Les différentes étapes du processus de dégradation des caractères vont maintenant être décrites plus en détails dans les sections suivantes.

Sélection des points de dégradation

Cette section sera dédiée à la sélection des germes qui sont les centres des zones de dégradation. L'idée est d'utiliser les couches couleur extraites à partir d'une image de documents couleur par un processus de segmentation. Chaque couche est associée à une image binaire, et correspond à une des couleurs principales qui sont observables sur l'image. L'originalité du modèle est l'étape d'extraction des germes qui est appliquée sur chaque couche (figure 4.7). Les points de dégradation étant probablement situés près des caractères, la couche correspondant au fond n'est pas traitée.

Ce modèle de bruit a été développé en parallèle de l'approche de segmentation décrite dans le chapitre 6. Celle-ci combine des informations spatiales et colorimétriques pour générer un ensemble de régions locales appelées superpixels. Ces régions sont ensuite fusionnées pour obtenir des couches couleur globales. Pour plus de détails sur l'extraction des couches couleur, le lecteur est invité à se référer aux chapitres 5, 6, et 7. Même si le choix de la méthode de segmentation à une certaine influence sur la qualité des résultats obtenus, ce modèle a été développé pour pouvoir être appliqué de manière indépendante. Les dernières évolutions concernant l'approche de segmentation pourront être aisément intégrées lors de travaux futurs.

Les points de dégradation sont extraits sur chaque couche de manière indépendante grâce à une adaptation du modèle de Kanungo ([Kanungo 1993]) proposée par [Kieu 2012] et [Kieu 2013]. Le principe est d'automatiser le paramétrage du modèle de Kanungo. Pour cela, l'utilisateur peut fixer la quantité de bruit souhaité en initialisant le nombre de germes (N_{sp}). Le modèle statistique produit ensuite un ensemble de germes qui sont localisés près des caractères. L'hypothèse est que la probabilité qu'un pixel change (du blanc au noir, et inversement) ne dépend que de la distance entre ce pixel et le caractère le plus proche. Cette probabilité est seuillée par un nombre généré aléatoirement. Finalement, le processus fournit deux ensembles de points pour chaque couche : P_{fb} et P_{bf} . P_{fb} (respectivement P_{bf}) contient tous les points associés aux centres des futures dégradations de fond, (respectivement de premier-plan).

La figure 4.7 montre un ensemble de points de dégradation extraits à partir d'une image de documents synthétiques. Ces points seront ensuite affectés à une catégorie de bruit particulière par un processus de classification que nous expliciterons dans la section 4.5.2.

Classification des régions de bruit

Dans la section précédente, nous avons explicité l'étape d'extraction des points de dégradation qui sont localisés à proximité des caractères. Les images de documents étant soumis à plusieurs types de dégradations qui peuvent avoir différents effets sur les caractères, il est important pour un modèle de proposer la simulation de plusieurs catégories de bruit. Cette section sera dédiée à l'association grâce à une

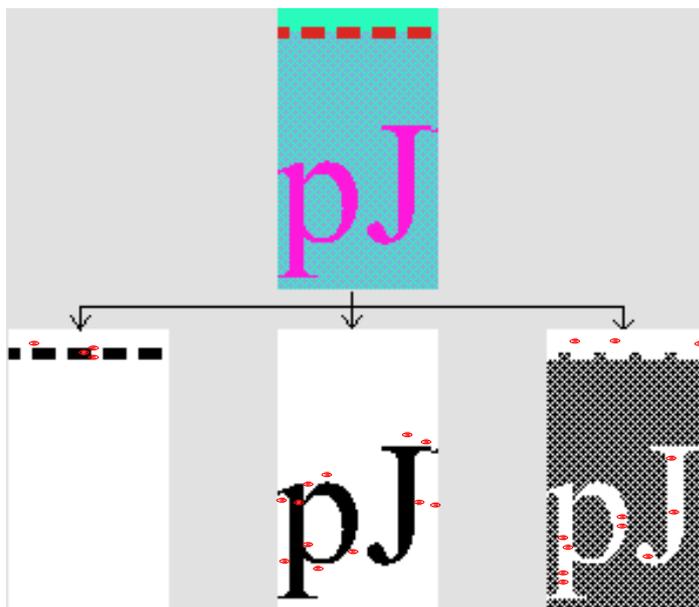


FIGURE 4.7 – Exemple de sélection de germes (points rouges), sur trois couches couleur extraites à partir d'une image synthétique. Crédits : Do Thi Luyen

étape de classification d'une catégorie de bruit aux germes préalablement extraits.

Tout comme pour l'étape de sélection des germes, le processus de classification est appliqué sur toutes les couches de manière indépendante. Les points sont répartis en trois classes : tâche indépendante, tâche recouvrante, et tâche de non-connexion. L'approche appliquée est celle proposée par [Kieu 2013]. La classification est faite selon des règles heuristiques, et selon les quantités de chaque catégorie définies par l'utilisateur. Le choix de ces heuristiques permet d'obtenir du bruit qui soit le plus proche possible de celui observable sur des images réelles. Elles dépendent du contenu de la zone de dégradation et de son voisinage.

La zone de dégradation correspond à une ellipse dont le centre est l'un des germes sélectionnés. Le type de dégradation est lié à la forme de cette ellipse. Ses dimensions sont donc calculées en fonction de la position de ce point par rapport aux contours du caractère le plus proche. Sa direction est déterminée en fonction du gradient local dans l'image au point de dégradation. Nous avons modifié le processus original de façon à profiter de l'information couleur. Toutes les mesures ont été réalisées à partir du gradient proposé par [Di Zenzo 1986] qui permet de traiter la couleur comme un vecteur.

La classification des germes est importante car une mauvaise définition du bruit peut entraîner des résultats qui ne seront pas réalistes. Le modèle proposé par [Kieu 2013] intègre les choix de l'utilisateur quant aux quantités souhaitées pour les différentes

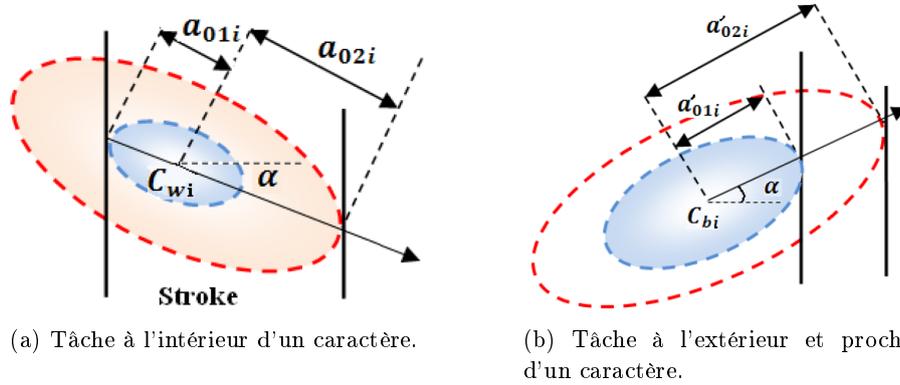


FIGURE 4.8 – Evolution d'une région de bruit lorsque la taille a_i de son axe majeur augmente. Crédits : [Kieu 2013]

catégories de bruit. Soient $p1$, $p2$, $p3$ la répartition souhaitée pour les trois catégories de dégradation. L'avantage est de laisser la possibilité d'adapter facilement le modèle à l'application souhaitée.

La classification est faite selon une règle heuristique de la manière suivante :

- les points de dégradation sont calculés et classés par ordre croissant de distance à son plus proche caractère ;
- les $p1$ premiers germes seront utilisés pour générer des tâches de non-connexion ;
- les $p2$ points suivants seront utilisés pour générer des tâches recouvrantes ;
- les derniers points permettront de générer des tâches indépendantes.

Génération de bruit

La figure 4.8 présente l'évolution d'une région de bruit lorsque la taille de son axe majeur augmente. Soient a_i , la taille de l'axe majeur de l'ellipse, C_i son centre, et a_{01i} (respectivement a_{02i}), la distance entre le centre de la tâche et le point de contour le plus proche (respectivement le point localisé sur le côté opposé de ce contour).

Une fois la catégorie de bruit définie, les pixels localisés à l'intérieur de l'ellipse sont modifiés pour simuler cette dégradation. Une règle heuristique fixe aléatoirement la taille de l'axe majeur de chaque ellipse en fonction du type de dégradation choisi. Soient a_j , a_k , et a_l , les tailles des ellipses respectivement associées à des tâches indépendantes, recouvrantes, et de non-connexion. Nous avons le système suivant :

$$\begin{cases} a_j = a_{01j} \times \mu(j), & \text{si } 0 < \mu_j < 1 \\ a_k = a_{01k} + \mu(k) \times (a_{02k} - a_{01k}), & \text{si } 0 < \mu_k < 1 \\ a_l = a_{02l} + \mu(l) \times \delta, & \text{si } 0 < \mu_l < 1 \end{cases} \quad (4.3)$$

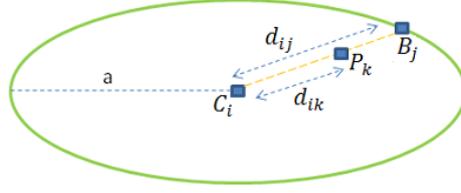


FIGURE 4.9 – Zone de dégradation elliptique. Crédits : [Kieu 2012]

où μ_j , μ_k , et μ_l sont générés aléatoirement avec $0 < \mu_j, \mu_k, \mu_l < 1$. Le facteur d'aplatissement g de l'ellipse est déterminé aléatoirement ($g \in [0, 1]$). Le paramètre δ correspond à la largeur moyenne de toutes les composantes connexes, ce qui permet d'éviter de générer des tâches de non-connexion trop larges. Nous allons maintenant expliciter le processus de génération du bruit à l'intérieur d'une ellipse telle que celle présentée sur la figure 4.9.

Soit $\bar{c}_i(l, a, b)$ la couleur du centre c_i de l'ellipse. Le germe associé à c_i a été extrait à partir d'une couche couleur l . Selon le modèle de Kanungo, ce point a été classé soit dans l'ensemble P_{fb} s'il doit passer du premier-plan au fond, soit dans l'ensemble P_{bf} , si au contraire il devient un pixel de premier-plan. Ainsi, \bar{c}_i correspond à la moyenne des pixels de fond (respectivement de premier-plan) de la couche l , si c_i appartient à l'ensemble P_{bf} (respectivement P_{fb}).

Soit B_j , un point situé sur le contours de l'ellipse. \bar{b}_i est la nouvelle couleur affectée à ce point. Pour la calculer, nous mesurons les distances couleurs entre B_j et les 8 points situés dans son voisinage. Nous utilisons la valeur euclidienne. Puis ces distances sont triées et nous affectons à \bar{b}_i la valeur $[L^*a^*b]$ du point associé à la médiane de l'ensemble de ces valeurs. La variable \bar{b}_i est utilisée pour calculer les nouvelles valeurs des points situés sur la ligne $C_i B_j$. Soit p_k , la couleur associée au pixel P_k situé sur cette ligne. p_k est calculé en utilisant une distribution Gaussienne aléatoire qui permet d'obtenir un résultat final plus réaliste :

$$p_k = N(\mu, \sigma^2) \quad (4.4)$$

où l'écart type σ est un paramètre d'entrée du modèle, et où μ est calculée de la manière suivante :

$$\mu = \bar{c}_i + (\bar{b}_j - \bar{c}_i) \times \left(\frac{d_{ik}}{d_{ij}}\right) \quad (4.5)$$

avec d_{ik} , la distance entre le pixel P_k et le centre de l'ellipse, C_i ; et d_{ij} qui est la distance entre C_i et B_j .

Pour terminer, un flou Gaussien est appliqué sur la zone de dégradation dans le but de l'uniformiser et de générer un bruit plus réaliste.

4.5.3 Résultats et discussion

Le modèle de bruit a été appliqué sur des images de documents synthétiques. Le processus de création de ces images est décrit dans la section 4.4. Les résultats ont été évalués visuellement par deux séries d'expériences :

- variation du nombre de germes ;
- variation du type de bruit.

Variation du nombre de germes : Les pourcentages associés à chaque catégorie de bruit ont été fixés expérimentalement de façon à être proches de ce qui peut être observé sur une image réelle (à titre indicatif, 15% de tâches indépendantes, 60% de tâches recouvrantes, et 25% de tâches de non-connexion). L'objectif ici était de faire varier le nombre de germes. La figure 4.10 présente les résultats obtenus avec 112, 192 et 1616 points de dégradation. Les figures 4.10b, et 4.10c correspondent à nos attentes. Cependant, le choix du nombre de germes est important. En effet, plus il y a de points, et plus l'image sera dégradée jusqu'à ne plus paraître suffisamment réaliste comme sur la figure 4.10d. L'avantage de ce modèle est de pouvoir fixer la quantité de bruit en fonction de l'application souhaitée.

Variation du type de bruit : L'objectif était de visualiser les effets de chaque catégorie de bruit prise séparément. Le modèle a été paramétré de façon à ne générer qu'une seule classe de dégradation. Les résultats sont présentés sur la figure 4.11. Nous pouvons noter que certaines des dégradations ne correspondent pas à la classe recherchée : par exemple, la figure 4.11b, pour laquelle des tâches indépendantes ont été générées, laisse apparaître des tâches recouvrantes. Ceci s'explique par l'utilisation des couches couleur lors de l'étape d'extraction des germes. En effet, ces points de dégradation sont déterminés sur chacune des couches de manière indépendante, et sont ensuite classés en fonction de celle-ci. Puis l'ensemble des germes est combiné lors de la phase de dégradation. Les éléments de fond et de premier-plan pouvant être de couleurs différentes, des effets non désirés sont susceptibles d'apparaître. D'autre part, ce modèle est dépendant de la qualité de la segmentation. Néanmoins ces erreurs ont une faible influence sur la qualité générale de la dégradation et n'impacteront que le pourcentage des différentes catégories de bruit. Les couches couleur nous permettent de prendre en compte l'information couleur. Les résultats visuels nous semble suffisamment réalistes visuellement pour pouvoir être utilisés dans des applications de vérité-terrain.

4.5.4 Conclusion sur le modèle de bruit

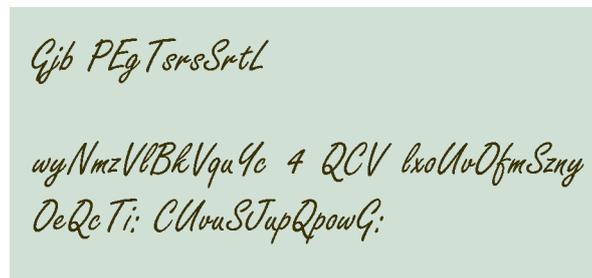
A travers cette section, nous avons présenté un modèle de dégradation de caractères qui est applicable aux images de documents couleur. L'enjeu est de pouvoir générer de grands ensembles d'images synthétiques pouvant être utilisées comme vérité-terrain pour diverses applications d'évaluation : segmentation, débruitage, etc. Ce modèle propose différents types de dégradation qu'il est possible d'adapter



(a) Image originale.

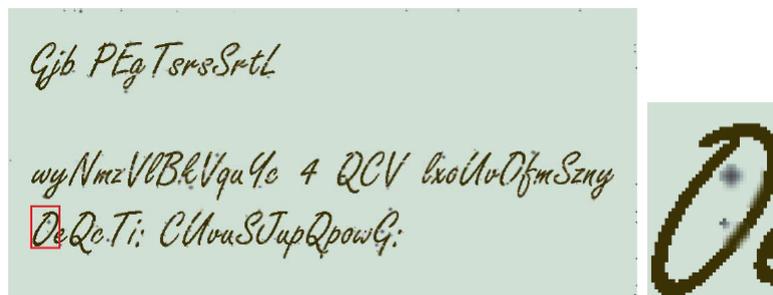
(b) $N_{sp} = 112$ (c) $N_{sp} = 192$ (d) $N_{sp} = 1616$

FIGURE 4.10 – Génération de bruit avec différentes quantités (N_{sp}) de germes. Crédits : Do Thi Luyen



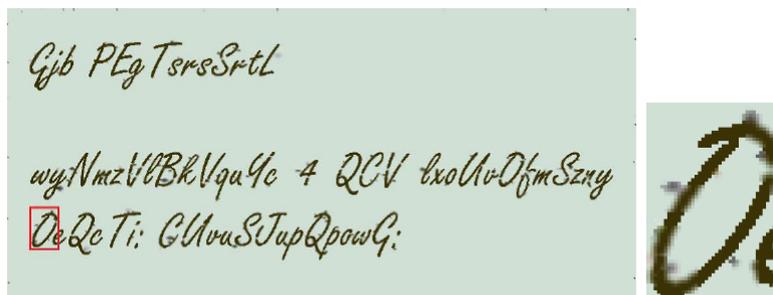
Gjb PEgTsrSrtL
wyNmzVlBkVquYe 4 QCV lxoUvDfmsZny
DeQcTi: CUvaSTupQpowG:

(a) Image originale.



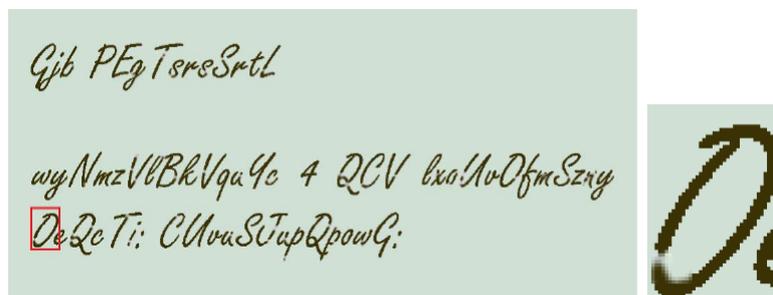
Gjb PEgTsrSrtL
wyNmzVlBkVquYe 4 QCV lxoUvDfmsZny
DeQcTi: CUvaSTupQpowG:

(b) Tâche indépendante



Gjb PEgTsrSrtL
wyNmzVlBkVquYe 4 QCV lxoUvDfmsZny
DeQcTi: CUvaSTupQpowG:

(c) Tâche recouvrante



Gjb PEgTsrSrtL
wyNmzVlBkVquYe 4 QCV lxoUvDfmsZny
DeQcTi: CUvaSTupQpowG:

(d) Tâche de non-connexion

FIGURE 4.11 – Dégradation des caractères avec différentes catégories de bruit prises séparément. Le pourcentage associé à la classe étudiée est initialisé à 100%, celui des autres catégories de bruit est alors égal à 0%. Crédits : Do Thi Luyen

grâce à quatre paramètres : le nombre de germes (niveau de dégradation), et le pourcentage de chaque catégorie de bruit (tâches indépendantes, tâches recouvrantes, et tâches de non-connexion).

Les résultats obtenus sont visuellement satisfaisants pour notre objectif. Les travaux futurs qui ont été identifiés sont l'intégration des derniers développements concernant la segmentation des documents administratifs en couches couleur dans le but d'améliorer la qualité de la segmentation. Pour limiter les erreurs liées à l'utilisation des couches couleur, il serait intéressant d'identifier certaines couleurs dominantes, ou bien les couleurs de fond et de premier-plan pour une région d'intérêt donnée. Ainsi, seules certaines couches jugées "pertinentes" dans un contexte spécifique seraient traitées. D'autre part, les enjeux les plus importants concernent la validation du modèle qui doit être plus poussée, et complétée notamment par une comparaison avec des images réelles. De plus, pour que les images soient plus réalistes, il serait intéressant de rendre le paramétrage complètement automatique, et dépendant du contenu des images de documents.

4.6 Conclusion sur le jeu de données

A travers ce chapitre, nous avons présenté les données utilisées pour valider expérimentalement nos approches de segmentation de documents en couches couleur. Notre partenaire industriel a fourni un ensemble d'images de documents réels relativement important, et constitué essentiellement de factures et de formulaires. Ceux-ci proviennent de nombreuses sources et sont très hétérogènes ce qui rend leur traitement complexe. A partir de cette base, nous avons sélectionné un échantillon de 150 images qui présentaient des caractéristiques posant problème aux industriels : textures, dégradés de couleurs, etc. En effet, même s'il existe quelques ensembles de données de référence en analyse de documents, ceux-ci ne prennent pas en compte les documents administratifs couleur.

Nous avons ensuite présenté les difficultés liées à l'annotation d'une base conséquente d'images de documents réels, et l'impossibilité d'en extraire manuellement des couches couleur. Nous nous sommes alors tournés vers les images synthétiques qui peuvent être générées à volonté, en très grand nombre, et en faisant varier les paramètres en fonction de l'application souhaitée. La couleur est alors connue au niveau pixel. La validation expérimentale est réalisée grâce à l'étude des valeurs de précision et de rappel qui sont des mesures classiques d'évaluation de méthodes de segmentation.

Pour terminer, nous avons introduit les travaux préliminaires qui ont été conduits dans le but de proposer un modèle de dégradation de caractères qui serait applicable aux images couleur. Tout l'enjeu est de pouvoir simuler le bruit apparaissant sur ces images au cours du processus de numérisation. Dans cet objectif, notre ap-

proche profite des travaux réalisés dans ce mémoire pour étendre à la couleur un modèle originellement proposé pour des images de documents anciens en niveaux de gris. Ce modèle propose la génération de différentes catégories de bruit similaires aux dégradations qui sont observables sur les images de documents administratifs. L'utilisateur a la possibilité de choisir la quantité et le type de bruit qu'il souhaite, ce qui permet de produire des images les plus réalistes possibles pour une tâche donnée.

En conclusion, la validation expérimentale se déroule en deux étapes : une évaluation visuelle sur les documents réels qui sont représentatifs des conditions réelles, et une évaluation sur un ensemble de 2000 images synthétiques dont l'avantage est de pouvoir comparer quantitative deux approches. Cette procédure a été mise en place pour évaluer les approches présentées dans les chapitres 5, 6, et 7 de ce mémoire.

Segmentation colorimétrique

Sommaire

5.1	Introduction	97
5.2	Contributions	98
5.2.1	Pré-traitements	98
5.2.2	Approche de quantification basée sur des mesures psychovisuelles	99
5.2.3	Approche basée sur une classification hiérarchique descendante	100
5.3	Expérimentations	105
5.3.1	Approche de quantification basée sur des mesures psychovisuelles	105
5.3.2	Approche basée sur une classification hiérarchique descendante	105
5.4	Résultats et discussion	106
5.4.1	Documents réels	109
5.4.2	Images synthétiques	110
5.5	Conclusion	114

5.1 Introduction

La segmentation colorimétrique est basée sur l'analyse de la distribution des couleurs dans un espace couleur, l'objectif étant de partitionner cet espace. Ces méthodes se basent sur l'hypothèse qu'il est possible d'associer un sous-ensemble de points ayant des caractéristiques proches dans l'espace couleur à une région colorimétriquement homogène dans l'image.

D'un point de vue colorimétrique, nous partons de l'hypothèse qu'une image de document administratif contient généralement peu de couleurs (souvent moins de 10). De plus, un document étant fait spécifiquement pour transmettre de l'information, il faut qu'il puisse être facilement lu. Nous considérons donc que les couleurs seront relativement bien contrastées. Plusieurs nuances d'une même teinte, avec des niveaux de saturation différents, peuvent néanmoins apparaître sur une image de document. Nous avons évoqué plus en détails les caractéristiques des documents couleur dans la section 3.6.

Nous avons envisagé deux approches colorimétriques qui seront présentées dans ce chapitre. Tout d'abord, nous décrirons une approche de quantification basée sur des mesures psychovisuelles. En effet, il nous a paru intéressant de tester ce modèle qui

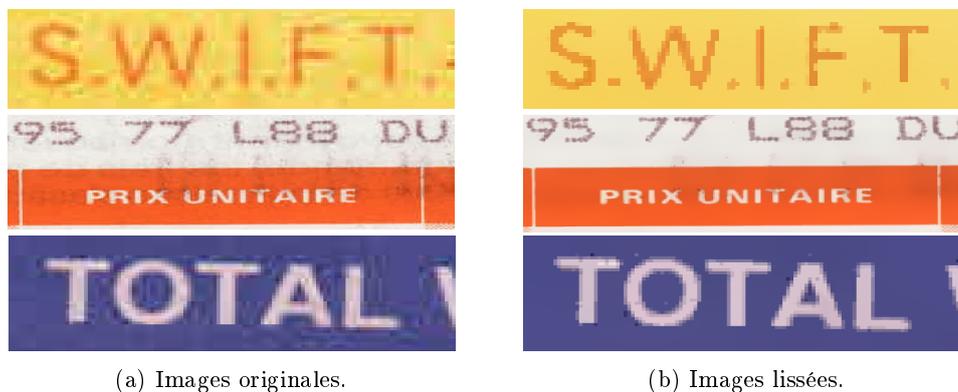


FIGURE 5.1 – Résultats du lissage des images de documents par un filtre bilatéral.

a été construit à partir de caractéristiques perceptuelles du système visuel humain. Dans un deuxième temps, nous nous sommes tournés vers une classification hiérarchique descendante qui permet de détecter le nombre de couleurs, et leurs frontières de manière adaptative.

5.2 Contributions

5.2.1 Pré-traitements

Le processus de numérisation et la compression introduisent du bruit dans les images sous forme d'artéfacts et de fausses couleurs, qui apparaissent notamment le long des contours (Figure 5.1a). Afin de réduire les perturbations liées à ce bruit, nous avons lissé notre ensemble d'images. Pour cela, nous avons choisi le filtre bilatéral proposé par [Tomasi 1998]. Ce filtre préserve les contours, ce qui nous permet de ne pas perdre trop d'informations au niveau des zones sensibles que sont les régions textuelles. L'idée est de calculer la valeur d'un pixel par rapport à celles de ses voisins, en tenant compte à la fois de leur proximité spatiale et colorimétrique.

L'hétérogénéité des images en termes de structure et de qualité, et le manque de connaissance sur le processus d'acquisition, rendent très difficile le paramétrage automatique de la méthode de lissage. Il pourrait être intéressant de détecter le type de document à traiter, ou la qualité de l'image, de façon à adapter la méthode en fonction de ces données. Dans ce mémoire cependant, nous nous sommes focalisés sur l'étape de segmentation. Les paramètres du filtre bilatéral ont été fixés expérimentalement de façon à corriger au mieux le bruit sur la majorité de nos images sans pour autant perturber l'extraction d'information par un lissage trop important. La figure 5.1b présente les résultats.

Nous avons appliqué ce filtre plusieurs fois de manière successive. Les images synthétiques bruitées de notre vérité-terrain ont été traitées de la même façon. Le protocole

expérimental a montré que les valeurs ci-dessous pour le filtre bilatéral permettaient d'obtenir le meilleur lissage sur notre jeu de données :

- taille de la fenêtre autour du pixel courant = 15 ;
- $\sigma_{couleur} = 15$ (plus la valeur est élevée, et plus les couleurs distantes en termes colorimétriques, localisées dans le voisinage du pixel seront mélangées) ;
- $\sigma_{spatial} = 60$ (plus la valeur est élevée, et plus les pixels éloignés auront d'influence les uns sur les autres à partir du moment où leur couleur est suffisamment proche) ;
- nombre d'itérations = 5.

5.2.2 Approche de quantification basée sur des mesures psychovisuelles

Dans un premier temps, nous avons testé une approche de *nommage de couleurs*. L'idée est d'associer un label à une valeur colorimétrique donnée. Le but est de pouvoir décrire une image de documents en utilisant uniquement un nombre réduit de couleurs définies par avance. Pour plus de détails, le lecteur est invité à se référer à l'état-de-l'art réalisé dans la section 3.5.2. Nous avons mis en application le modèle paramétrique proposé par [Párraga 2009], qui est lui-même basé sur les travaux réalisés par [Benavente 2008]. Nous n'avons pas contribué au développement de cette approche. Notre objectif était d'étudier ce qu'un modèle basé sur des mesures psychovisuelles pouvait apporter à la segmentation d'images de documents. Cette section décrira son fonctionnement.

Cette approche propose d'utiliser les 11 termes de couleurs suivants : bleu, marron, vert, orange, rose, violet, rouge, jaune, noir, gris et blanc. Ce choix est basé sur l'étude réalisée par [Berlin 1969] selon laquelle ces couleurs sont universelles. L'objectif est ici de partitionner l'espace L^*a^*b qui est perceptuellement uniforme. L'avantage de cet espace est que l'utilisation de la distance euclidienne pour mesurer la différence perceptuelle entre deux couleurs est reconnue. Chaque catégorie de couleur est modélisée par une fonction, qui mesure le degré d'appartenance d'une valeur numérique donnée à cette catégorie spécifique. Ce modèle paramétrique est basé sur l'utilisation de fonctions sigmoïdes triples avec un centre elliptique (triple sigmoid with elliptical center, TSE, en anglais). La figure 5.2 permet de visualiser les régions floues.

Nous avons retenu cette approche floue car elle permet plus de flexibilité par rapport à un partitionnement dur qui associe une et une seule couleur à un pixel donné. Nous n'avons pas implémenté ici d'approche s'appuyant directement sur cette notion. Cependant, nous pourrions imaginer plusieurs applications industrielles concrètes. Dans le cadre d'une segmentation en couches couleur par exemple, un pixel pourrait apparaître sur plusieurs couches. Les résultats de la segmentation pourraient aussi être affinés ou corrigés en prenant en compte à la fois les degrés d'appartenance aux couleurs d'un pixel donné, et ceux des pixels localisés dans son voisinage.

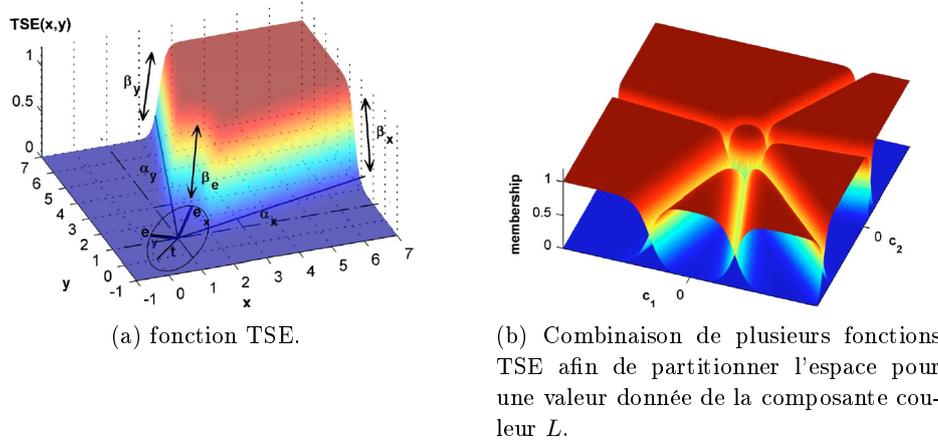


FIGURE 5.2 – Régions d'appartenance floue proposée par [Benavente 2008] pour partitionner l'espace L^*a^*b , et utilisées par [Párraga 2009]. Crédits : [Benavente 2008] (Figures 6 et 7)

Les paramètres du modèle fixent les frontières entre les couleurs. Ils doivent donc être déterminés judicieusement car ils détermineront la forme des régions d'appartenance aux différentes catégories de couleurs. La grande force de ces modèles est qu'ils prennent en compte la capacité du système visuel humain à discriminer les couleurs. Pour cela, l'étape de paramétrage s'appuie sur un ensemble de données, obtenu préalablement au moyen d'expériences psychovisuelles. Les participants ont eu pour tâche d'associer un terme à une couleur présentée sur un patch. Cette couleur était choisie de façon à se trouver sur les zones de frontières entre deux couleurs, qui sont problématiques, et donc loin des couleurs représentatives de chaque catégorie. On leur demandait ensuite de choisir parmi les deux termes lequel correspondait effectivement à la couleur qu'ils étaient en train de visualiser.

Le modèle présenté ici utilise un nombre fixe de couleurs. Or les documents à notre disposition sont très hétérogènes. Si les couleurs choisies et les frontières du modèle sont mal définies, ou inappropriées pour une famille de documents donnée, certaines régions de l'image seront mal classées. De plus, l'ajout ou la suppression d'une classe nécessite la mise en place d'une nouvelle série d'expériences permettant d'acquérir les données nécessaires à une redéfinition du modèle. Par conséquent, nous avons envisagé une méthode plus adaptative basée sur une classification hiérarchique descendante. Celle-ci sera présentée dans la section suivante.

5.2.3 Approche basée sur une classification hiérarchique descendante

Sans connaissance *a priori* concernant les caractéristiques colorimétriques des documents, il nous est apparu qu'imposer un nombre fixe de couleurs était restric-

tif et peu flexible. Par exemple dans le cas des 11 couleurs, plusieurs nuances de bleues seront catégorisées comme une seule et même couleur bleue. Nous avons donc envisagé une approche de regroupement qui permettrait de détecter les nuages de points associés aux couleurs dominantes dans l'espace colorimétrique. Nous présentons les principales caractéristiques de ces techniques dans la section 3.4.2. Pour plus de détails, le lecteur est invité à se référer à [Jain 2010].

Nous nous sommes tournés vers une approche basée sur une *classification hiérarchique* afin de rendre l'étape de détection du nombre de couleurs adaptative. En effet, ce type de méthode ne requiert pas la spécification du nombre initial de classes. De plus, cette notion de hiérarchie est intéressante, car elle conserve les relations classes mères-classes filles, ce que ne permet pas une technique de partitionnement plus classique telle que l'algorithme des K-moyennes. En fonction de l'application souhaitée, plusieurs niveaux de granularité peuvent éventuellement être proposés.

D'après les observations réalisées sur notre base, nous avons constaté un nombre très réduit de couleurs sur l'ensemble de nos documents. Par conséquent, nous avons préféré une approche *descendante* (tous les points de l'image sont placés dans une classe qui est ensuite divisée récursivement), à une approche *ascendante* (chaque point est considéré comme une classe ; la segmentation est obtenue par fusions successives). De plus, l'hypothèse de base est que les couleurs dominantes de l'image peuvent être associées à des nuages de points dans l'espace colorimétrique. Ces ensembles de points ou clusters doivent être les plus homogènes et les plus compacts possibles. Une approche descendante permet de tester assez facilement ces caractéristiques et de ne diviser l'ensemble de points que si c'est nécessaire.

Dans un premier temps, tous les pixels sont placés dans une seule et unique classe. Puis celle-ci est divisée en deux, et de nouveaux centres sont initialisés. Les pixels qui appartenaient à cette classe sont alors redistribués dans les deux nouvelles classes ainsi formées. Cette opération est réalisée grâce à un algorithme de partitionnement ; nous avons choisi d'appliquer ici les *K-Moyennes*. Après la convergence, des caractéristiques sont mesurées et comparées à des valeurs de seuil. Des règles d'arrêt décident alors si la classe peut encore être divisée ou non. Ces caractéristiques donnent des informations sur l'homogénéité et la connexité des ensembles considérés. L'objectif est d'obtenir des classes les plus homogènes possibles d'un point de vue colorimétrique, et formant des régions compactes sur l'image. Le processus est itéré jusqu'à ce qu'il n'y ait plus aucune classe à diviser, ou après N itérations (N étant fixé par l'utilisateur). Les couleurs dominantes correspondent aux feuilles du dendrogramme. La figure 5.3 résume le processus de segmentation.

Nous allons maintenant présenter plus en détails le processus de division des classes, ainsi que les caractéristiques et les règles d'arrêt utilisées.

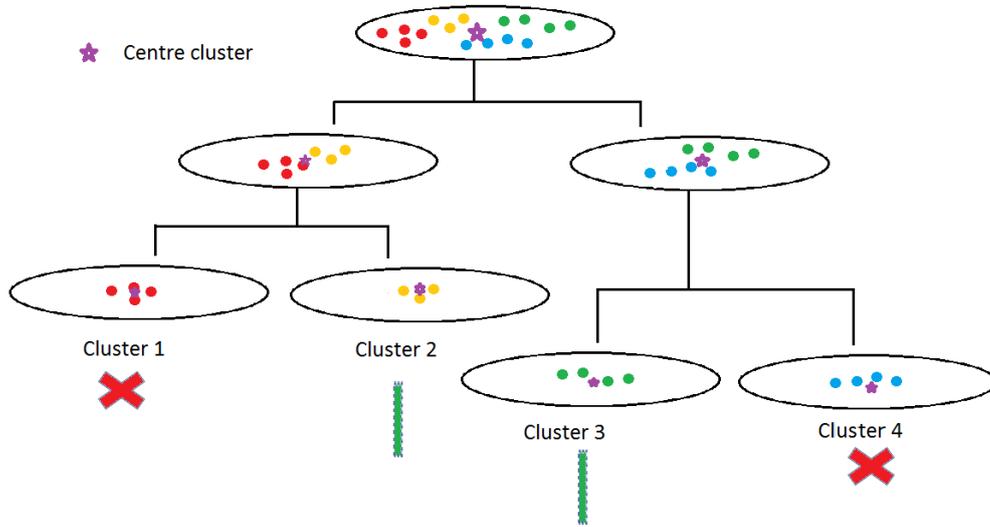


FIGURE 5.3 – Classification hiérarchique descendante.

Division des ensembles de points

Dans cette section, nous allons expliciter le processus de division des ensembles de pixels. L'objectif de cette étape de créer deux nouveaux sous-ensembles de points à partir d'un ensemble qui est considéré trop hétérogène d'un point de vue colorimétrique, et pas assez compact, pour être un nuage de points valable, et être conservé comme une couleur principale de l'image.

Une fois les centres initialisés, les pixels de la classe sont redistribués dans les deux nouveaux sous-ensembles grâce à une méthode de partitionnement. Nous avons choisi d'appliquer ici l'algorithme des *K-Moyennes* dont le but est de minimiser la somme des carrés des distances de chaque point par rapport au centre de la classe qui lui est associée. Cette méthode de regroupement est communément utilisée pour sa rapidité et sa simplicité. Nous utilisons la distance euclidienne pour tous les calculs dans tous les espaces colorimétriques.

La difficulté principale consiste à initialiser les centres des nouveaux clusters. En effet, les résultats dépendent fortement des valeurs choisies. Le risque est d'obtenir deux ensembles qui ne seraient pas suffisamment séparés l'un de l'autre. Au cours de l'étape de regroupement, chaque point est affecté au groupe dont il est le plus proche. Le processus est itéré jusqu'à convergence ou pendant N itérations.

Soit Z contenant tous les pixels de la classe parent. Cet ensemble est ensuite divisé en deux. Les deux nouveaux centres Z_1 et Z_2 sont calculés de la manière suivante

selon l'approche décrite dans [Ray 1999] :

$$Z_1 = (C_1 - a_1, C_2 - a_2, C_3 - a_3) \quad (5.1a)$$

$$Z_2 = (C_1 + a_1, C_2 + a_2, C_3 - a_3) \quad (5.1b)$$

où C_i est la i -ème composante couleur et, a_i , une constante. La valeur de ces constantes est calculée en prenant en compte les valeurs minimales et maximales de chaque composante couleur des points de l'ensemble considéré. La constante a_i est égale à la moitié de la plus petite des deux valeurs ($C_i - \min_i$) et ($\max_i - C_i$), avec \min_i et \max_i , qui sont respectivement la valeur minimale et maximale de la i -ème composante couleur. De cette façon, les deux nouveaux ensembles sont bien séparés à l'intérieur du groupe parent. Afin de rester à l'intérieur du cluster, nous conservons la valeur de a_i , pour i , la composante couleur pour laquelle la variance est la plus élevée. Sinon, a_i est égale à 0.

Règles d'arrêt

Au cours du processus, nous mesurons certaines propriétés de chacun des groupes. Ces caractéristiques sont comparées à des valeurs de seuils. Des règles d'arrêt décident si la classe doit encore être divisée ou non. Tout l'enjeu est d'obtenir des classes suffisamment éloignées les unes des autres, le plus homogènes possible d'un point de vue colorimétrique, et formant des régions compactes dans l'image. Nous utilisons le *degré de compacité spatio-colorimétrique* (en anglais, *spatial-color compactness degree*) proposé par [Macaire 2006]. Une région est définie comme un sous-ensemble S de pixels fortement connexes qui ont une couleur homogène.

Dans un premier temps, le degré de connexité permet de mesurer la compacité d'un sous-ensemble de points dans l'image. Ainsi, une valeur proche de 0, indique que les points de l'ensemble sont dispersés à travers l'image. Au contraire, une valeur proche de 1, indique que les points sont fortement connexes. Soit P , un pixel de l'image I . $N_S(P)$ est le sous-ensemble des 8 pixels voisins de P qui appartiennent à S .

La connexité entre P et S est mesurée ainsi :

$$\gamma_S = \frac{\text{Card}\{N_S(P)\}}{8} \quad (5.2)$$

Le degré de connexité de S est alors la moyenne normalisée des voisins des pixels de l'ensemble S qui appartiennent eux-aussi à S :

$$CD(S) = \frac{\sum_{P \in S} \gamma_S(P)}{\text{Card}\{S\}} \quad (5.3)$$

Dans un deuxième temps, [Macaire 2006] combinent ce degré avec un degré mesurant l'homogénéité colorimétrique. Le degré d'homogénéité, $\sigma(S)$, mesure la dispersion colorimétrique des points d'un sous-ensemble S . L'objectif est d'extraire des groupes homogènes, c'est à dire des ensembles dont les points sont faiblement dispersés dans

l'espace couleur. Cette valeur peut être estimée comme la racine carrée de la trace de la matrice de covariance des points appartenant à ce sous-ensemble S . Cette valeur globale est comparée à une valeur locale.

Pour chaque pixel P appartenant à S , nous déterminons la mesure de dispersion $\sigma(N_S(P))$ du sous-ensemble $N_S(P)$ constitué de tous les voisins de P qui appartiennent aussi à S . La dispersion locale $\sigma_{local}(S)$ correspond alors à la moyenne de toutes les mesures de dispersion $\sigma(N_S(P))$ estimé pour tous les sous-ensembles $N_S(P)$, de tous les pixels P de S :

$$\sigma_{local}(S) = \frac{1}{Card\{S\}} \cdot \sum_{P \in S} \sigma(N_S(P)) \quad (5.4)$$

Lorsque la mesure de dispersion locale $\sigma_{local}(S)$ de S est proche de sa mesure de dispersion $\sigma(S)$, l'ensemble S constitue probablement une classe unique. Au contraire, si $\sigma_{local}(S)$ est plus petite que $\sigma(S)$, l'ensemble peut être divisé en plusieurs sous-ensembles distincts. Le degré d'homogénéité est calculé de la manière suivante :

$$\begin{cases} HD(S) = \frac{\sigma_{local}(S)}{\sigma(S)} \text{ if } \sigma(S) \neq 0 \\ HD(S) = 1 \text{ otherwise} \end{cases} \quad (5.5)$$

Enfin, le degré de compacité spatio-colorimétrique $SCD(S)$ est mesuré ainsi :

$$SCD(S) = CD(S) \cdot HD(S) \quad (5.6)$$

Une valeur élevée (proche de 1) signifie que les pixels appartiennent à un sous-ensemble qui est à la fois fortement connexe dans l'image, et homogène en termes de couleurs. Cependant, les images de documents contiennent de grandes régions telles que les régions de fond pour lesquelles cette valeur reste élevée du fait du nombre important de pixels fortement connexes les uns avec les autres. Ces régions sont donc divisées, ce qui entraîne une sur-segmentation de l'image. Pour réduire cet effet, nous avons ajouté un critère pour comparer la différence entre les valeurs maximales et minimales des composantes couleurs du centre de chacun des groupes.

Soient R_{max}, V_{max} , et B_{max} , les valeurs d'intensité maximales des composantes couleurs R, V et B ; R_{min}, V_{min} , et B_{min} correspondent à leurs valeurs d'intensité minimales. Les distances d_R , d_V et d_B , respectivement associées aux canaux R, V, et B, sont calculées à partir de la distance euclidienne. d_{max} est alors la distance maximale :

$$d_R = R_{max} - R_{min} \quad (5.7a)$$

$$d_V = V_{max} - V_{min} \quad (5.7b)$$

$$d_B = B_{max} - B_{min} \quad (5.7c)$$

$$d_{max} = \max(d_R, d_V, d_B) \quad (5.7d)$$

Si pour un groupe donné, la différence maximale d_{max} , est inférieure à un seuil (fixé expérimentalement), nous supposons qu'une division supplémentaire produirait des sous-groupes trop proches l'un de l'autre. Le processus de division est alors stoppé pour ce groupe.

5.3 Expérimentations

5.3.1 Approche de quantification basée sur des mesures psychovisuelles

Au cours de ces travaux, nous avons simplement testé cette approche telle quelle sans adaptation ni modification. Le nombre de couleurs est fixé à 11 termes qui sont universellement utilisés d'après l'étude réalisée par [Berlin 1969]. Le modèle a été construit spécifiquement pour partitionner l'espace colorimétrique L^*a^*b . Nous avons utilisé les paramètres déterminés au moyen d'expériences psychovisuelles réalisées par [Párraga 2009] lors de la création de leur modèle.

L'objectif était d'analyser les avantages et les limites de cette technique pour la segmentation de documents couleur. Cependant, les images de documents sont extrêmement variables en termes de couleurs. La présence de nuances d'une même teinte, de bruit, ou même de dégradés peut poser des problèmes de modélisation : détermination du nombre de couleurs à utiliser, définition de frontières pertinentes, etc. Du fait de ces limites, nous nous sommes tournés vers un autre type d'approche, plus adaptative, basée sur une classification hiérarchique descendante.

5.3.2 Approche basée sur une classification hiérarchique descendante

Cette classification peut être appliquée à différents espaces couleurs. Il en existe de nombreux mais il est communément admis qu'il n'en existe aucun de générique. Une étude détaillée des espaces couleurs numériques a été réalisée dans la section 3.3. Sur le plan théorique, les nuages de points que nous recherchons peuvent être plus ou moins marqués en fonction de l'espace et des images à traiter. Afin de vérifier cette hypothèse, nous avons appliqué l'approche décrite à trois espaces, chacun appartenant à une des grandes familles d'espaces couleurs :

- RGB (système de primaires), car cet espace est celui qui est historiquement mis en place sur de nombreux matériels d'acquisition, et est par conséquent très communément utilisé ;
- HSV (système luminance-chrominance), car il permet de séparer l'information de luminance de l'information de chrominance ;
- L^*a^*b (système perceptuel), car il permet de reproduire les écarts visuellement observables entre deux couleurs.

Nous avons fixé expérimentalement les valeurs gérant les critères d'arrêt du processus de division. Celui-ci s'arrête dès que l'un des critères ci-dessous est vérifié :

- le nombre maximal d'itérations fixé à 10 est atteint ;
- le degré de compacité spatio-colorimétrique, $SCD(S)$, est supérieure à 0.009 ;
- la différence maximale entre les valeurs d'intensité maximales et minimales des composantes couleurs (L,a et b), d_{max} est inférieure à 110.

Les images de documents réels et synthétiques ont été traités de manière identiques avec les mêmes critères d'arrêt. Nous allons maintenant donner quelques pistes pour déterminer les seuils de ces mesures. Le nombre maximal d'itérations permet simplement de stopper le processus avant qu'il n'atteigne un nombre trop élevé de couches couleur détectées. Plus la valeur du degré de compacité spatio-colorimétrique est proche de 1 et plus l'ensemble est à la fois compact et homogène. Cependant, cette valeur est très complexe à déterminer en particulier pour des documents si hétérogènes en termes de caractéristiques colorimétriques. Du fait des limites que nous présenterons dans la section 5.4, nous avons finalement préféré développer une approche de segmentation différente qui avait un plus grand potentiel en termes de qualité. Pour terminer, la valeur d_{max} permet de réguler le processus de division en l'empêchant de créer des groupes trop petits. Cependant, elle doit être fixée judicieusement : trop élevée, elle prendra le pas sur la règle de seuillage du degré de compacité et la division sera arrêtée uniquement sur ce critère ; trop petite, elle ne sera pas efficace.

5.4 Résultats et discussion

Approche de quantification basée sur des mesures psychovisuelles

Documents réels

Les résultats peuvent être visualisés sur la figure 5.4. La couleur affectée à chaque pixel est celle pour laquelle le degré d'appartenance du pixel à cette couleur est le plus élevé. Nous pouvons observer visuellement que cette approche est très sensible au bruit. Grâce à l'intégration de la logique floue dans ce modèle, il pourrait être possible d'améliorer en partie les résultats. Pour cela, il faudrait tenir compte du voisinage des pixels, et des valeurs de leurs degrés d'appartenance aux couleurs lors de la prise de décision. Nous n'avons pas mis en place ce processus car il nécessiterait une étape de traitement supplémentaire, et la mise en place de règles de décision qui pourraient ne pas être généralisables, et qui ne pourraient pas tout corriger. En effet, cela ne résoudrait pas les problèmes liés au choix du nombre de couleurs. De plus, le bruit introduit de nouvelles couleurs sur l'image. Une approche basée uniquement sur des caractéristiques colorimétriques n'est pas suffisamment précise pour séparer les couleurs qui devraient être conservées, de celles associées à ces dégradations, et qui par conséquent devraient être supprimées.

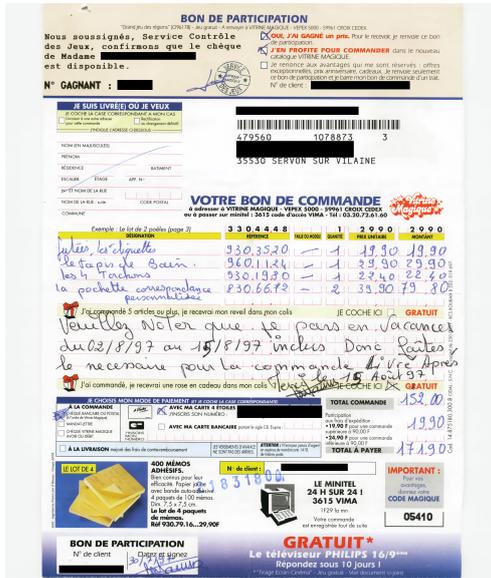
Le grand avantage de cette quantification est d'utiliser des termes aisément compréhensibles par un humain pour décrire la couleur. De plus, cette approche repose sur des expériences psychovisuelles, ce qui permet de modéliser le comportement du système visuel humain, et en particulier sa capacité à discriminer les couleurs. Cependant, ce type de modèle doit être construit avec attention. Les points clés sont la détermination du nombre de couleurs à utiliser, et la définition de frontières pertinentes entre ces couleurs. Or notre base contient des documents très hétérogènes pour lesquels nous n'avons aucune connaissance *a priori*. Ceux-ci contiennent plus ou moins de couleurs, qui sont plus ou moins bien contrastées. L'ajout ou la suppression d'une nouvelle teinte dans le modèle impose la mise en place de nouvelles séries d'expériences pour fixer les paramètres. Au cours du cycle de vie d'un document, un utilisateur peut ajouter différentes surcharges, ce qui ajoute une difficulté supplémentaire, car il n'est pas toujours possible de contrôler ce facteur.

En conclusion, ce modèle fonctionne assez bien lorsque les couleurs sont bien contrastées. En revanche, certains problèmes se posent pour étiqueter des valeurs qui se trouvent à la frontière entre deux classes du modèle. De plus, il faut définir un nombre de couleurs pertinent. Le risque de sur-segmentation ou de sous-segmentation est élevé si le modèle utilisé n'est pas adapté. Ce type d'approche paraît peu robuste au bruit, et peu adaptative dans un contexte de segmentation bas niveau. En revanche, il pourrait être intéressant d'utiliser un tel modèle pour ajouter des données sémantiques de plus haut niveau, après une première étape de segmentation. L'idée serait par exemple d'utiliser cette approche pour fournir des données de description pour annoter des régions déjà segmentées, ou encore les couches couleur construites par notre système. Le nombre de couches extraites étant très réduit, le bruit aurait moins d'influence sur l'utilisation de cette technique. Celle-ci pourrait fournir des informations permettant d'interroger facilement une base de données, avec des termes du langage humain.

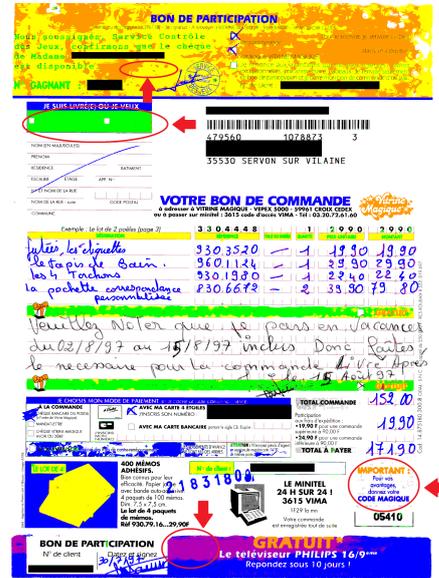
Images synthétiques

Les mesures de performance de la méthode de quantification sur la base d'images synthétiques sont résumées dans le tableau 5.1. La faible valeur de précision est principalement due au fait que la couleur a été dégradée par l'ajout de bruit sur les images synthétiques. Ce bruit a pour effet de créer de nouvelles couleurs. Cette approche ne prend pas en compte le contexte spatial des points avant de leur affecter une étiquette. Il n'est donc pas possible de vérifier si la couleur d'un pixel est cohérente par rapport à son voisinage. Visuellement, nous observons une sur-segmentation qui apparaît notamment le long des contours.

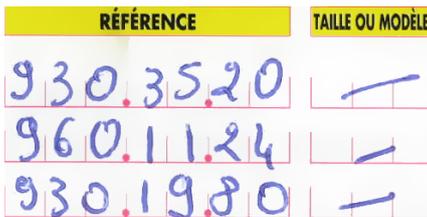
La valeur élevée de rappel montre que le nombre de couches couleur détectées est proche de la réalité, même si une partie des pixels sont mal classés. Cependant, le processus détecte généralement plus de couches que le nombre réel. La méthode de nommage de couleurs impose un nombre fixe de termes de couleurs. Toutes ces



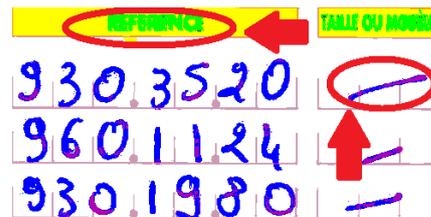
(a) Image originale



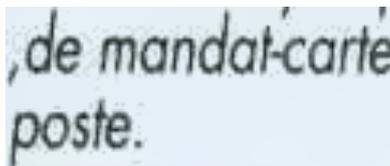
(b) Résultats, en fausses couleurs.



(c) Image originale.



(d) Résultats de la segmentation.



(e) Image originale.



(f) Résultats de la segmentation.



(g) Image originale.



(h) Résultats de la segmentation.

FIGURE 5.4 – Approche de nommage de couleur proposée par [Benavente 2002]. Dans un souci de simplification, nous faisons apparaître tous les résultats sur une image unique, avec de fausses couleurs. Un ensemble de couches binaires peut être obtenu pour chaque couleur à partir de cette image. Le label affecté aux pixels est celui pour lequel le degré d'appartenance des pixels au label est le plus élevé. Les données personnelles ont été masquées. Les cercles en rouge mettent en évidence certaines erreurs de segmentation.

	Quantification
Précision	0.70
Rappel	0.90

TABLEAU 5.1 – Performance de la quantification sur la base d’images synthétiques.

couleurs ne sont pas obligatoirement présentes sur une même image. Par conséquent, nous avons supprimé les catégories qui ne contenaient aucun pixel de l’image. Une analyse des résultats montre que les couches supplémentaires qui ont été extraites correspondent la plupart du temps aux régions de transition qui sont des zones très bruitées.

Le choix du modèle a un autre impact aussi bien sur la précision que sur le rappel. Si certains de nos documents synthétiques ont plusieurs nuances d’une même catégorie de couleurs (ex : plusieurs nuances de bleues), celles-ci seront classées comme faisant partie d’une classe unique. La précision va alors diminuer. Le nombre de couches couleur détecté sera alors inférieur à la réalité.

Approche basée sur une classification hiérarchique descendante

5.4.1 Documents réels

Les figures 5.6, 5.7 et 5.8 présentent les résultats obtenus respectivement avec les espaces RVB, HSV et L*a*b. Nous pouvons observer que cette approche est elle-aussi très sensible au bruit. Les résultats montrent une tendance à la sur-segmentation. Ceci est notamment dû aux seuils utilisés par notre approche qui ont été déterminés expérimentalement, et qui ne s’adaptent pas aux documents.

Nous avons appliqué les mêmes règles d’arrêt à tous les espaces couleurs pour pouvoir les comparer. Nous n’avons pas noté de différences visuelles notables dans les résultats. Certaines régions sont effectivement mieux segmentées selon l’espace couleur. Par exemple, l’espace L*a*b semble être meilleur pour segmenter certaines zones texte (figure 5.8). Mais dans le même temps, l’en-tête beige du document est mieux segmenté en RVB (figure 5.6) ou en HSV (figure 5.7). Parmi les espaces testés, nous n’en avons trouvé aucun qui se démarque des autres.

D’autre part, la classification ne prend en compte que les données colorimétriques à un niveau pixel. De plus, la numérisation et la compression ont introduit du bruit sur les images sous forme d’artéfacts et de fausses couleurs (figure 5.5). Dans le cas d’un caractère sur une région de fond par exemple, notre approche produira donc au moins trois couches : une pour le fond, une pour le premier-plan, et une pour la zone de transition. Si l’effet du bruit est trop important, les pixels de fond et de



FIGURE 5.5 – La numérisation et la compression introduisent du bruit sur les images sous forme d’artéfacts et de fausses couleurs qui apparaissent notamment près des contours.

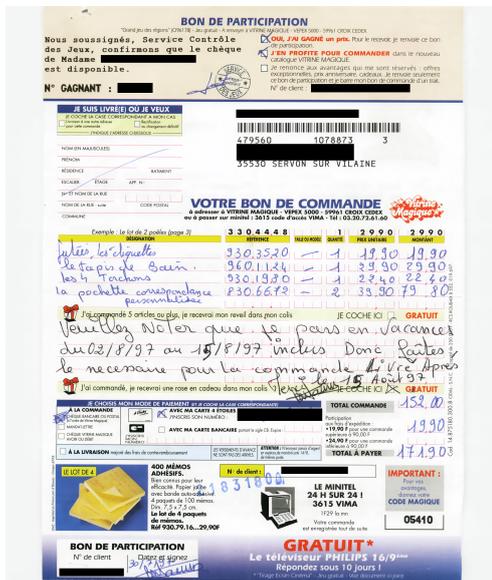
premier-plan seront répartis sur plusieurs couches couleur, ce qui rendra les résultats difficilement exploitables (figure 5.6f).

Notre segmentation est une approche globale qui entraîne parfois des erreurs de détection : certaines couleurs peuvent ne pas être détectées du tout. Ce phénomène se produit notamment lorsqu’une petite région apparaît sur une très grande zone de fond, ou lorsque des couleurs sont faiblement contrastées. Il apparaît dans deux cas. Tout d’abord, lorsque la distribution globale d’une couleur est faible par rapport à une autre, l’information risque d’être noyée au sein de la masse de données disponible. D’autre part, notre approche divise les ensembles considérés par deux. Or, en réalité, ceux-ci peuvent contenir plus de couleurs. Cela provoque alors des erreurs de répartition des points. Une fois que la décision est prise, il n’est plus possible de revenir en arrière.

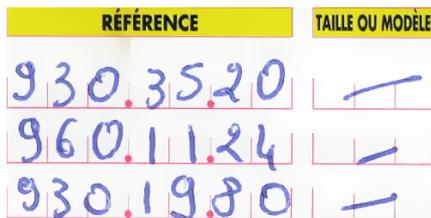
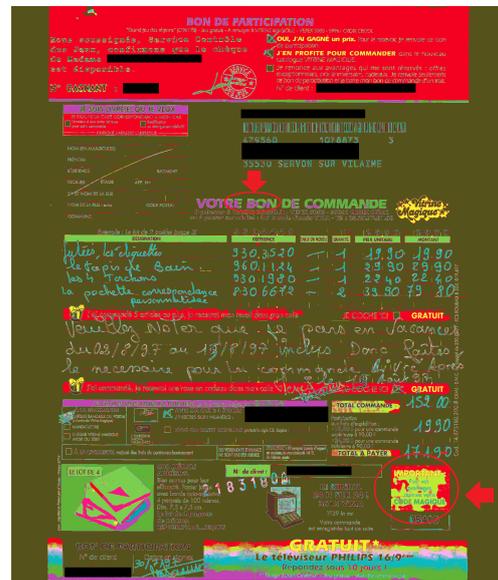
Malgré les inconvénients que nous venons d’évoquer, les résultats sont intéressants. Les couleurs bien contrastées ont été généralement détectées. Les régions de fond, moins bruitées, sont relativement bien segmentées. Les régions principales ainsi extraites donnent des indications sur la disposition des éléments sur le document, et peuvent alors servir de descripteurs. L’avantage principal de cette classification est de pouvoir déterminer automatiquement le nombre de couleurs en fonction du document analysé. Cette approche prend donc en compte l’hétérogénéité de notre base d’images. Les problèmes de sur-segmentation pourraient être atténués par l’intégration de données spatiales ou contextuelles, et par l’utilisation de paramètres et de règles d’arrêt adaptatifs. D’autre part, appliquer des traitements localement plutôt que globalement pourrait aussi améliorer la précision de la segmentation.

5.4.2 Images synthétiques

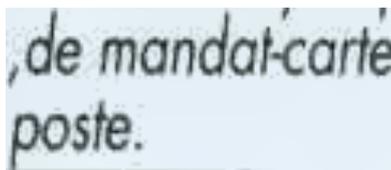
Les résultats de l’évaluation de la performance de la classification hiérarchique descendante sont résumés dans le tableau 5.2. La faible valeur de la précision confirme la grande sensibilité de cette approche au bruit. Les valeurs de rappel, notamment pour l’espace L^*a^*b , indiquent que le processus a trouvé un nombre



(a) Image originale



(c) Image originale.



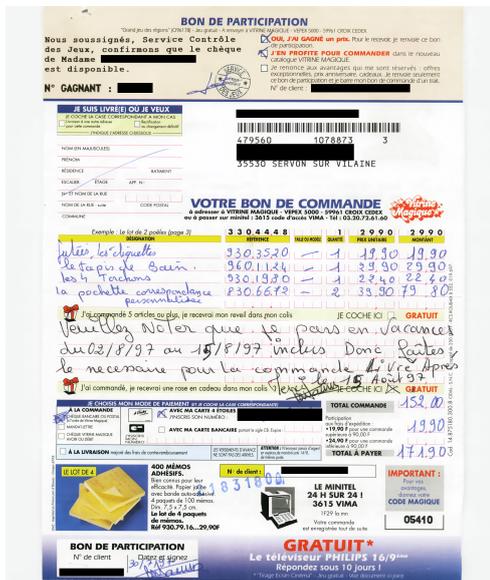
(e) Image originale.



(g) Image originale.



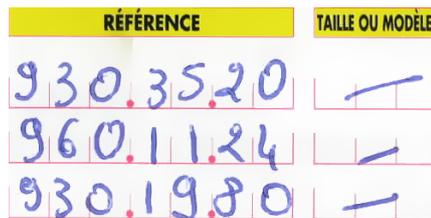
FIGURE 5.6 – Approche de classification hiérarchique descendante appliquée dans l'espace RVB. Les cercles en rouge mettent en évidence certaines erreurs de segmentation.



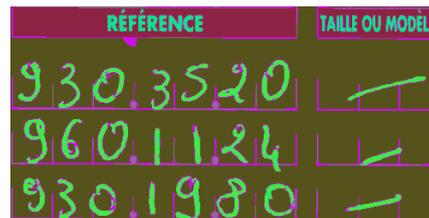
(a) Image originale



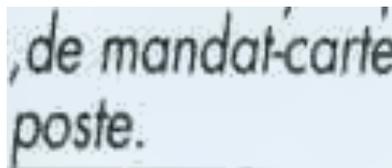
(b) Résultats (en fausses couleurs).



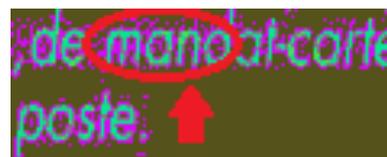
(c) Image originale.



(d) Résultats (en fausses couleurs).



(e) Image originale.



(f) Résultats (en fausses couleurs).



(g) Image originale.



(h) Résultats (en fausses couleurs).

FIGURE 5.8 – Approche de classification hiérarchique descendante appliquée dans l'espace L^*a^*b . Les cercles en rouge mettent en évidence certaines erreurs de segmentation.

Classification hiérarchique	RVB	HSV	L*a*b
Précision	0.76	0.67	0.55
Rappel	0.94	0.93	0.98

TABLEAU 5.2 – Performance de la classification hiérarchique sur la base d’images synthétiques.

cohérent de couches couleur. Cependant, beaucoup de pixels sont mal classés. Une analyse des cas pour lesquels la segmentation échoue montre que les erreurs sont causées par le bruit (dégradations de la couleur, zones de transitions, etc), mais aussi par la non détection de certaines couleurs. Cette dernière erreur peut être provoquée par le processus de division qui fait un découpage strict d’un ensemble en deux nouveaux sous-groupes. Or, un ensemble de points donné peut contenir plus que deux couleurs. Certains points seront alors mal classés. Dans ce cas, non seulement les régions qui leur sont associées ne seront pas détectées, mais en plus, ils vont venir perturber le processus de division des nouveaux ensembles ainsi créés.

La différence importante de précision entre les espaces couleurs étudiés s’explique par le fait que les paramètres et règles de décision ont été fixés expérimentalement avec des images en RVB. Pour pouvoir les comparer, nous avons conservé le même paramétrage. Néanmoins, ils ont des propriétés différentes, et la précision pourrait être améliorée en rendant cette étape adaptative. Cependant, cette technique paraît peu généralisable car les images de documents étant très hétérogènes d’un point de vue colorimétriques, il n’est pas possible de prédire quel sera l’espace couleur dans lequel les régions de l’image seront le mieux segmentées. Du fait de l’ensemble des limites liées au processus de découpage et de partitionnement, nous avons préféré expérimenter une approche de segmentation différente et nous n’avons pas poursuivi cette réflexion.

5.5 Conclusion

L’objectif de notre projet est la segmentation d’images de documents administratifs en couches couleur. Les deux problématiques majeures sont la détection du nombre de couleurs observables sur une image, de manière automatique, et sans connaissance *a priori*, et la définition de frontières pertinentes entre ces couleurs. La difficulté provient de la nécessité d’appliquer une méthode générique pour traiter des documents complexes et hétérogènes.

A travers ce chapitre, nous avons présenté deux approches colorimétriques : une approche de nommage de couleurs, et une approche basée sur une classification hiérarchique descendante. La première utilise un modèle paramétrique construit à partir de données psychovisuelles. Ce modèle permet d’associer un terme de couleur

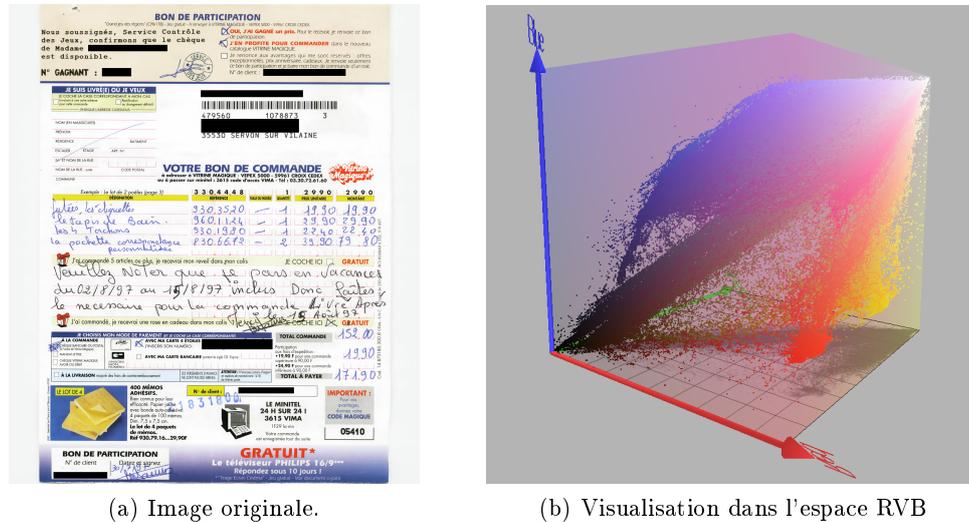


FIGURE 5.9 – Visualisation d'un document réel dans l'espace RVB.

du langage humain à une valeur numérique. Du fait des limites de cette approche notamment au niveau de son manque d'adaptabilité, nous avons ensuite testé une méthode basée sur une classification hiérarchique qui détecte automatiquement le nombre de couleurs d'un document.

Les deux approches colorimétriques que nous venons de présenter donnent des résultats intéressants. Cependant, nous arrivons à la même conclusion que l'état-de-l'art : une segmentation basée uniquement sur la couleur est peu robuste au bruit présent sur les images de documents. Théoriquement, il est possible d'extraire les couleurs dominantes d'une image en recherchant les nuages de points observables dans l'espace couleur. Il faut pour cela que les couleurs soient bien séparées. En réalité, à cause de la densité de l'information et du bruit, ces points forment plutôt des traînées. Les documents complexes contiennent aussi des dégradés de couleurs et des nuances très proches ce qui complique encore le processus de détection. La figure 5.9 présente la visualisation d'un document réel dans l'espace RVB.

L'approche de nommage de couleurs résout une partie des problèmes en définissant un nombre fixe de couleurs. Pour cela, elle se base sur des travaux linguistiques, et des expériences psychovisuelles. L'avantage est que ce modèle simule le comportement du système visuel humain. Les images de documents étant créées spécifiquement pour pouvoir être lues par un humain, l'utilisation de ce type de modèle peut être justifiée. Mais, le choix de ce nombre de couleurs n'est pas si évident. Il est difficile de classer correctement les valeurs localisées à la frontière entre deux couleurs. De plus, des nuances d'une même couleur peuvent être étiquetées avec le même label même si elles appartiennent à des régions différentes de l'image.

La classification hiérarchique est plus adaptative. Le nombre de couleurs est obtenu automatiquement grâce à un processus de divisions successives d'ensembles de points. Une couleur dominante est déterminée par un ensemble de critères d'arrêt mesurant l'homogénéité colorimétrique, et la compacité de ces ensembles. La difficulté provient du paramétrage de la méthode. La qualité des résultats dépend fortement de paramètres et de règles d'arrêt qui ont été déterminés expérimentalement. Du fait de la variabilité des images à notre disposition en termes de caractéristiques couleurs, ces seuils sont difficiles à fixer. Cette classification est une approche globale qui fournit des couches couleur observables sur l'ensemble de l'image. Les données sont considérées à un niveau pixel. Malgré des problèmes de sur-segmentation liés notamment au bruit, les résultats pourraient être utilisés pour fournir un descripteur qui servirait de base à une segmentation plus fine.

Pour améliorer les résultats, il serait intéressant de combiner les informations couleurs avec des données provenant de sources différentes. Un pixel partage des données proches avec son voisinage. Nous avons donc envisagé le développement d'une approche spatio-colorimétrique pour combiner des informations couleurs avec des informations spatiales. Notre méthode est basée sur une segmentation en région locales appelées *superpixels*. Elle sera présentée dans le chapitre 6.

Segmentation spatio-colorimétrique

Sommaire

6.1	Introduction	117
6.2	Contributions	118
6.2.1	Pré-traitements	118
6.2.2	SLIC original	119
6.2.3	SLIC Adaptatif	121
6.2.4	Procédure de regroupement	124
6.3	Expérimentations	126
6.4	Résultats et discussion	127
6.4.1	Documents réels	127
6.4.2	Images synthétiques	128
6.5	Conclusion	131

6.1 Introduction

Les approches colorimétriques sont très sensibles au bruit car elles ne tiennent pas compte des relations spatiales entre les éléments de l'image (chapitre 5). Ainsi, les régions détectées dans un espace couleur donné ne correspondent pas toujours à des régions réelles de l'image. D'après l'état-de-l'art (section 3.4.2), la combinaison d'informations colorimétriques avec des données spatiales rend le processus de segmentation potentiellement plus robuste.

Nous avons choisi d'intégrer ces informations spatiales en appliquant une approche s'inspirant des méthodes de segmentation en régions locales appelées *superpixels*. La combinaison de données spatiales et colorimétriques est compliquée. L'idée derrière cette approche est donc de sur-segmenter volontairement l'image, pour ensuite appliquer des processus complexes directement au niveau régions, en interprétant des informations "pseudo-locales".

Il existe encore très peu de travaux portant sur l'utilisation des superpixels pour la segmentation d'images de documents. Pourtant, ces approches semblent avoir un grand potentiel pour cette tâche. De par leur caractère structuré, ces images se

prêtent assez bien à ce type de segmentation. Les régions générées par ces méthodes sont compactes. De plus, elles adhèrent relativement bien aux frontières de l'image, ce qui en fait toute leur force.

La contribution proposée au cours de ces travaux est basée sur l'algorithme *SLIC* (en anglais, *Simple Linear Iterative Clustering*) qui a été initialement proposé par [Achanta 2012]. Notre choix s'est porté sur lui pour ses bonnes performances en termes de qualité de la segmentation. Cependant, cette algorithme a été développé spécifiquement pour des images naturelles. L'application sur des images de documents pose donc le problème de la présence de grandes régions homogènes que l'on n'observe pas ou très peu sur les images naturelles.

Notre contribution porte donc sur l'adaptation de l'approche originale aux caractéristiques structurelles propres aux images de documents. La version adaptative du SLIC produit ainsi des régions de tailles différentes en fonction du contenu de l'image : grandes sur des zones de fond homogènes, et petites au niveau des régions de premier-plan, plus variables, et qui nécessitent une segmentation plus fine. Pour terminer, nous présenterons la procédure de regroupement de ces régions locales, en couches couleur globales. Ce processus est basé sur la représentativité des couleurs en termes de fréquence d'apparition dans l'image.

6.2 Contributions

6.2.1 Pré-traitements

Le processus de numérisation et la compression introduisent du bruit dans les images sous forme d'artéfacts et de fausses couleurs, qui apparaissent notamment le long des contours (Figure 6.1a). Afin de réduire les perturbations liées à ce bruit, nous avons lissé notre ensemble d'images. Pour cela, nous avons choisi le filtre bilatéral proposé par [Tomasi 1998]. Ce filtre préserve les contours, ce qui nous permet de ne pas perdre trop d'informations au niveau des zones sensibles que sont les régions textuelles. L'idée est de calculer la valeur d'un pixel par rapport à celles de ses voisins, en tenant compte à la fois de leur proximité spatiale et colorimétrique.

L'hétérogénéité des images en termes de structure et de qualité, et le manque de connaissance sur le processus d'acquisition, rendent très difficile le paramétrage automatique de la méthode de lissage. Il pourrait être intéressant de détecter le type de document à traiter, ou la qualité de l'image, de façon à adapter la méthode en fonction de ces données. Dans ce mémoire cependant, nous nous sommes focalisés sur l'étape de segmentation. Les paramètres du filtre bilatéral ont été fixés expérimentalement de façon à corriger au mieux le bruit sur la majorité de nos images sans pour autant perturber l'extraction d'information par un lissage trop important. La figure 6.1b présente les résultats.

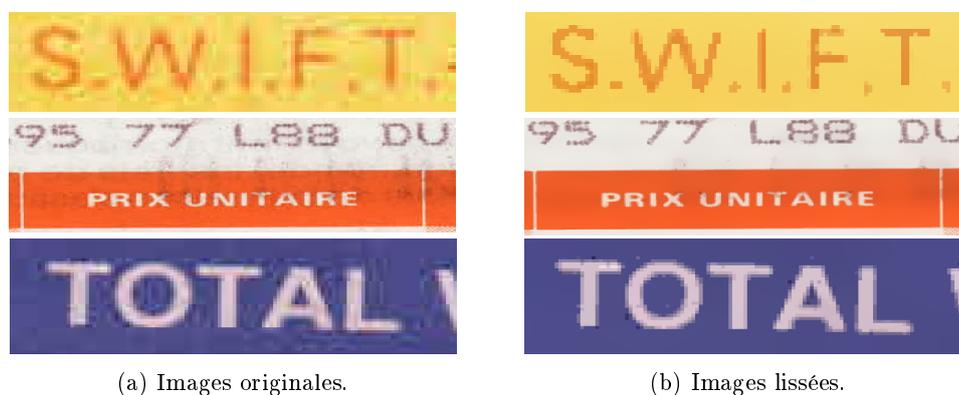


FIGURE 6.1 – Résultats du lissage des images de documents par un filtre bilatéral.

Nous avons appliqué ce filtre plusieurs fois de manière successive. Les images synthétiques bruitées de notre vérité-terrain ont été traitées de la même façon. Le protocole expérimental a montré que les valeurs ci-dessous pour le filtre bilatéral permettaient d’obtenir le meilleur lissage sur notre jeu de données :

- taille de la fenêtre autour du pixel courant = 15 ;
- $\sigma_{couleur} = 15$ (plus la valeur est élevée, et plus les couleurs distantes en termes colorimétriques, localisées dans le voisinage du pixel seront mélangées) ;
- $\sigma_{spatial} = 60$ (plus la valeur est élevée, et plus les pixels éloignés auront d’influence les uns sur les autres à partir du moment où leur couleur est suffisamment proche) ;
- nombre d’itérations = 5.

6.2.2 SLIC original

Les approches basées sur la génération de superpixels sont souvent utilisées dans des applications de vision par ordinateur comme pré-traitements. L’idée principale est de réaliser une sur-segmentation pour regrouper des pixels en régions locales, les superpixels (SPs), qui sont perceptuellement cohérents. L’objectif est de réduire les données redondantes de l’image, tout en évitant la perte d’information. En permettant de manipuler des régions plutôt que des pixels, ces approches réduisent l’effet du bruit sur la segmentation.

Nous avons choisi d’appliquer l’algorithme SLIC, développé par [Achanta 2012], pour sa capacité à produire des régions adhérant bien aux contours de l’image. En effet, la segmentation doit être suffisamment précise, en particulier au niveau des zones textuelles, pour ne pas perturber la phase de reconnaissance. SLIC est une adaptation de l’algorithme des *K-Moyennes* appliqué localement. Les images sont traitées dans l’espace couleur L^*a^*b . Pour rendre la segmentation robuste, cette approche combine à la fois des informations colorimétriques (les composantes couleurs), et des données spatiales (les coordonnées des centres). Fondamentalement,



(a) SLIC - initialisation des centres

(b) SLIC - contours

FIGURE 6.2 – Segmentation en superpixels d'un document réel basée sur l'approche SLIC.

c'est l'utilisation de ces deux sources de données par une technique de partitionnement appliqué dans une petite fenêtre de recherche qui permet à SLIC d'adhérer si bien aux contours. Par défaut, le seul paramètre est k , le nombre désiré de SPs qui auront approximativement la même taille (Figure 6.2b).

Pendant la phase d'initialisation, k SPs sont échantillonnés sur une grille régulière (Figure 6.2a). Soit S , l'intervalle de la grille. Chaque SP, dont le centre de gravité est $[x_i, y_i]$, est associé au vecteur $C_i = [l_i, a_i, b_i, x_i, y_i]$. Les centres sont déplacés vers le point associé à la valeur de gradient la plus faible dans un voisinage 3×3 , ceci afin d'éviter l'ajout de bruit en initialisant un centre sur un contour. Pour accélérer le processus, le clustering limite sa recherche à une région $2S \times 2S$ autour du centre du SP (Figure 6.3).

A chaque itération, une distance est mesurée entre chaque pixel et les SPs dont la fenêtre de recherche le recouvre. Puis les pixels sont associés au cluster pour lequel la valeur de distance est minimale. L'algorithme 1 décrit le processus complet. Tous les calculs sont réalisés à partir de la distance euclidienne. L'information colorimétrique portée par les centres des SPs est le vecteur moyen $[l, a, b, x, y]^T$ calculé à partir de tous les pixels contenus par ce SP. Une distance couleur, d_c , est donc calculée ainsi :

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \quad (6.1)$$

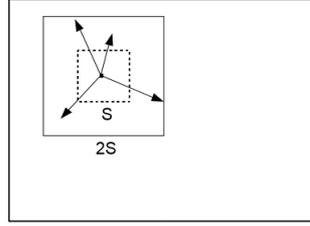


FIGURE 6.3 – Fenêtre de recherche du SLIC. Crédits : [Achanta 2012] (Figure 2b)

Une distance spatiale, d_s , est ensuite mesurée en prenant en compte les centres de gravité des clusters :

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (6.2)$$

La mesure de distance globale D combine une distance couleur, d_c , et une distance spatiale, d_s :

$$D = \sqrt{d_c^2 + \left(\frac{d_s}{S}\right)^2 m^2} \quad (6.3)$$

S correspond à l'intervalle d'échantillonnage, soit la distance spatiale maximale à l'intérieur d'un cluster. Cette valeur permet de normaliser la distance d_s . La variable m pondère l'importance de la proximité spatiale par rapport à la similarité colorimétrique. Lorsque la valeur de m est élevée, la proximité spatiale est plus importante, et les superpixels ainsi générés seront plus compacts. Lorsque cette valeur est plus faible, les superpixels adhéreront beaucoup plus aux contours de l'image, mais leur taille, et leur forme seront plus irrégulières. Avec l'espace L^*a^*b , m est généralement fixé dans l'intervalle $[1, 40]$. Nous avons conservé la valeur par défaut et initialisé m à 10 dans tous les tests qui ont été réalisés ici.

6.2.3 SLIC Adaptatif

L'algorithme SLIC présenté précédemment nécessite la connaissance du nombre de SPs désirés. Il est bien sûr possible de le faire dépendre de caractéristiques de l'image telle que sa taille par exemple. Cependant, ce paramétrage peut se révéler délicat pour une base d'images génériques. Notre but est de traiter des documents aux contenus divers et dont les structures sont variables. De plus, cette approche produit des régions qui ont toutes approximativement la même taille. Si la valeur du pas utilisé pour l'échantillonnage est trop grande, beaucoup de petits éléments peuvent être manqués. D'un autre côté, les images de documents contiennent beaucoup de grandes régions homogènes pour lesquelles l'utilisation de petits SPs n'est pas pertinente.

Au cours de ces travaux, notre contribution a consisté en une étude de l'adaptation de la méthode du SLIC à la structure particulière des images de documents. Pour cela, nous avons proposé d'utiliser des SPs de tailles différentes qui dépendent du contenu des documents. Ainsi, l'objectif est d'obtenir des SPs de grande taille dans

Algorithme 1 : Segmentation en superpixels : approche SLIC ([Achanta 2012])

```

/* Initialisation */
Initialiser les centres des clusters  $C_k = [l_k, a_k, b_k, x_k, y_k]^T$ , grâce à un
échantillonnage des pixels sur une grille régulière d'intervalle  $S$ .

Déplacer les centres des clusters vers les points pour lesquels la valeur du
gradient est la plus faible dans un voisinage  $3 \times 3$ .

Initialiser les labels  $l(i) = -1$  pour chaque pixel de  $i$ .

Initialiser les distances  $d(i) = \infty$  pour chaque pixel de  $i$ .

répéter
  /* Affectation */
  pour tous les centres des clusters  $C_k$  faire
    pour tous les pixels  $i$  dans un voisinage  $2S \times 2S$  autour de  $C_k$  faire
      Mesurer la distance  $D$  entre  $C_k$  et  $i$ .
      si  $D < d(i)$  alors
         $d(i) = D$ 
         $l(i) = k$ 
      fin
    fin
  fin

  /* Mise à jour */
  Calculer les valeurs des centres des nouveaux clusters.

  Mesurer l'erreur résiduelle  $E$ .

jusqu'à  $E \leq \text{seuil}$ ;

```

les zones homogènes de l'image, tandis que les régions pour lesquelles la variation de la couleur est élevée localement, et qui nécessitent plus de précision, contiendront beaucoup plus de SPs de petite taille (Figure 6.5d).

Pour réaliser cet objectif, nous avons proposé un processus d'initialisation qui ne distribue plus les centres des SPs de manière homogène (Figure 6.4). L'idée est de détecter les éléments faisant probablement partie du premier-plan et de les séparer du fond. Les pixels de l'image sont ainsi répartis en deux masques en fonction de leurs caractéristiques. Dans un second temps, l'initialisation se fait comme pour le SLIC mais en utilisant les deux masques pour changer le paramètre S contrôlant la taille de la fenêtre de recherche : une valeur élevée pour les pixels de fond, et une plus faible pour les éléments de premier-plan. Enfin, le clustering est appliqué de la même manière que le SLIC original. Pour réduire le nombre final de couleurs, tous les SPs ayant la même couleur sont fusionnés.



FIGURE 6.4 – SLIC Adaptatif - initialisation des centres.

Cette séparation fond/premier-plan peut être réalisée de différentes manières. Par exemple, il serait tout à fait possible d'utiliser une procédure classique de binarisation. Cependant, notre objectif est de profiter au maximum de l'information couleur disponible sur l'image. Les zones de premier-plan correspondent aux régions de gradient fort. Nous appliquons donc tout d'abord un gradient couleur proposé par [Di Zenzo 1986] (figure 6.5a). L'avantage est que cette approche manipule l'information couleur de manière globale sous la forme d'un vecteur. Un gradient classique au contraire traiterait chaque composante couleur de manière indépendante, puis fusionnerait les résultats (stratégie marginale), perdant ainsi la corrélation entre ces canaux. Nous utilisons un filtre de Canny [Canny 1986] pour détecter les contours principaux de l'image. A la fin de cette étape, nous obtenons une image binaire (figure 6.5b).

Le masque de premier-plan peut ne pas être suffisamment précis. En effet, les contours correspondent à des zones de transition. A ce niveau, ce masque ne détecte pas forcément toutes les couleurs "stables" de premier-plan et de fond. Par conséquent, pour éviter de manquer des éléments, nous avons choisi d'étendre le masque de premier-plan au voisinage des pixels détectés lors de la détection des contours. Pour cela, nous appliquons l'algorithme de lissage *Run Length Smearing* proposé par [Wong 1982]. Cette dernière méthode a été choisie car elle est adaptée à la structure des images de documents. Elle est communément utilisée en analyse de documents pour extraire des blocs de texte. La plupart des éléments d'une image de

document ayant une orientation globalement horizontale, nous l'appliquons uniquement dans cette direction. La figure 6.5c présente le masque de premier-plan obtenu. Le résultat de la segmentation finale peut être observé sur la figure 6.5d.

Pour finir, le choix des valeurs d'échantillonnage pour les deux masques est importante. Trop petite, le processus générera beaucoup trop de régions. Trop grande, certains éléments risquent d'être manqués. Nous avons déterminé ces valeurs expérimentalement comme des ratios de la largeur de l'image.

6.2.4 Procédure de regroupement

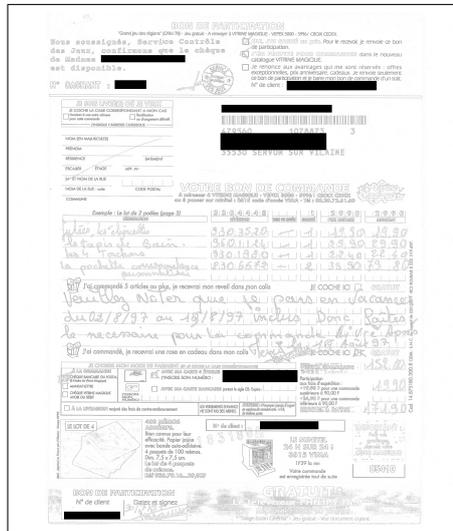
A la fin de l'étape de segmentation, nous obtenons un ensemble de régions locales qui sont perceptuellement cohérentes. L'avantage d'une approche basée sur l'algorithme SLIC est qu'elle nous a permis de regrouper les pixels en utilisant des informations locales qui ont réduit l'effet du bruit. De plus, la version adaptative profite des caractéristiques structurelles des images de documents contrairement au processus original qui a été développé spécifiquement pour les images de scènes naturelles.

Il nous faut maintenant regrouper ces régions en couches couleur globales. L'information colorimétrique portée par un SP est la valeur moyenne des composantes L^*a^*b de tous les pixels qui appartiennent à cette région. La difficulté est alors de combiner à la fois des données locales et globales. Il nous faut pouvoir regrouper des éléments similaires d'un point de vue colorimétrique même s'ils se trouvent spatialement très éloignés sur le document.

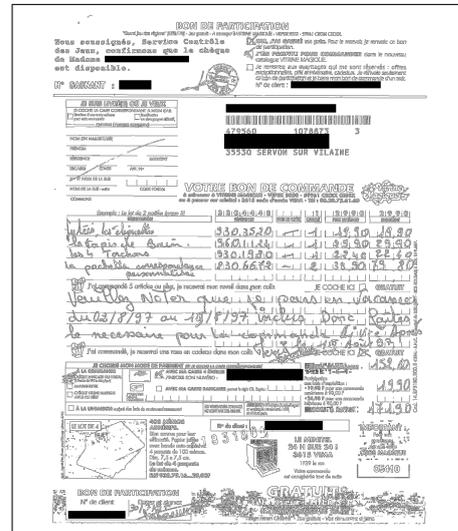
Nous avons donc choisi d'analyser la représentation globale des couleurs portées par chaque SP afin d'extraire les couleurs dominantes. Nous partons de l'hypothèse que le système visuel humain va s'adapter à l'environnement et captera les couleurs dans leur globalité : les plus représentées, et les plus contrastées. Tout d'abord, la fréquence d'apparition (en nombre de SPs) de chaque couleur est calculée. Puis, l'idée est de créer une nouvelle couche pour la couleur dont la fréquence est la plus élevée. Tous les SPs ayant cette couleur, ou une couleur très proche (en termes de distance colorimétrique), seront regroupés. C'est à dire que l'étiquette associée à la couche couleur globale sera affectée à tous ces SPs. Une fois étiquetés, ces SPs ne sont plus pris en compte. Nous recherchons ensuite la couleur la plus représentative parmi celles portées par les SPs non étiquetés. La procédure de regroupement s'arrête lorsqu'il n'y a plus aucun SP à étiqueter.

Le processus de regroupement est résumé de la façon suivante :

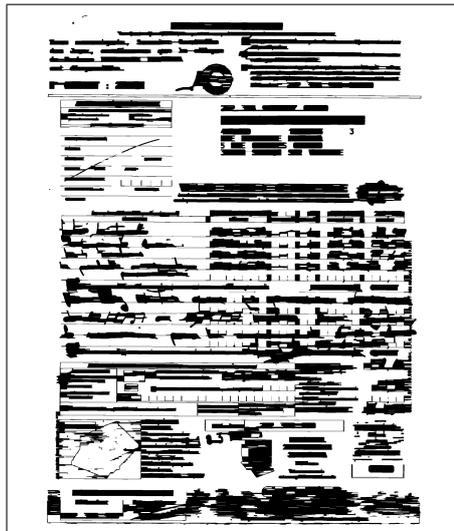
- trouver la couleur la plus représentative en termes de fréquence d'apparition ;
- regrouper tous les SPs dont la distance couleur est inférieure à un seuil fixé par rapport à cette couleur représentative ;
- itérer jusqu'à ce qu'il n'y ait plus de SPs à étiqueter.



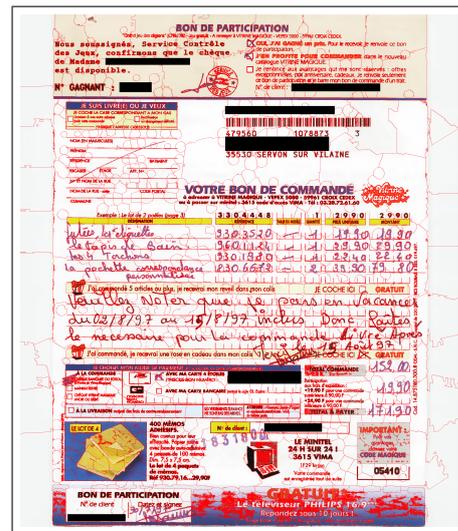
(a) Gradient de Di Zenzo (pour une meilleure visibilité, les valeurs ont été inversées).



(b) Après une détection de contours par un filtre de Canny appliqué sur le gradient de Di Zenzo présenté en (a).



(c) Génération du masque de premier-plan par l'application du Run-Length Smearing sur (b).



(d) Contours des régions générées par l'application du SLIC adaptatif.

FIGURE 6.5 – Segmentation en superpixels d'un document réel basée sur l'approche du SLIC adaptatif.

L'approche de fusion a été appliquée aux résultats obtenus par la segmentation basée sur le SLIC original, et à ceux générés par notre approche afin de comparer ces deux méthodes. La distance utilisée pour le regroupement des couleurs a été déterminée expérimentalement de façon à limiter une sur-segmentation des régions textuelles.

6.3 Expérimentations

Nous avons présenté précédemment notre approche de segmentation en régions locales appelées superpixels. Celle-ci s'adapte au contenu de l'image pour fournir des régions de tailles différentes : grandes au niveau des zones homogènes, et plus petites sur les zones nécessitant une segmentation plus précise. Nous nous sommes en effet aperçu que le SLIC original était plutôt adapté aux problèmes de vision par ordinateur. Sa stricte application aux images de documents soulève le problème des grandes régions qui sont moins présentes sur les images de scènes naturelles.

Un certain nombre de paramètres ont été utilisés. Nous avons conservé les valeurs par défaut des paramètres déterminant la compacité ($m = 10$), et le nombre d'itérations ($iter_{max} = 10$) définis par le SLIC original. Le clustering est réalisé dans l'espace L^*a^*b tout comme dans le processus original. Lors de la phase de test, nous avons eu quelques difficultés à fixer le nombre initial de superpixels. Les résultats présentés pour la segmentation basée sur le SLIC original ont été générés avec un pas S égal à 100. Cette valeur a donc été fixée expérimentalement de façon à obtenir une segmentation satisfaisante pour la majorité des images à notre disposition.

En ce qui concerne le SLIC adaptatif, le nombre de superpixels n'est plus déterminé par un paramètre qui doit être fixé par un utilisateur, mais dépend directement du contenu de l'image. L'idée principale est de répartir les centres des SPs de manière non homogène sur l'ensemble de l'image. Pour cela, les éléments du document qui se trouvent probablement au premier-plan sont extraits. Lors de l'initialisation, des SPs avec des tailles différentes sont initialisés en fonction du masque auquel ils appartiennent. Les valeurs de ces pas ont été déterminés expérimentalement pour chaque masque. C'est à dire que tous les SPs dont les centres appartiendront au premier-plan auront tous la même valeur de pas, qui sera plus petite que celle associée aux zones de fond. Par conséquent, plus l'image contient d'éléments de premier-plan, et plus le processus générera de petits SPs. Au contraire, sur les grandes zones homogènes, les SPs seront peu nombreux et de grandes tailles. A titre d'exemple, dans le cas du document présenté sur la figure 6.7a, le pas S est initialisé à 51 pour les éléments de premier-plan et à 255 pour les éléments de fond.

Les paramètres contrôlant le lissage (rl_{th}) et les pas utilisés pour les régions de fond (S_{fd}) et de premier-plan (S_p), sont des ratios de la largeur, l , de l'image afin de les rendre le plus adaptatif possible :

$$rl_{th} = \frac{l}{70} \quad (6.4)$$

$$S_{fd} = \frac{l}{10} \quad (6.5)$$

$$S_p = \frac{l}{50} \quad (6.6)$$

Pour le processus de génération du masque de premier-plan, les paramètres du filtre de Canny ont été fixés expérimentalement pour convenir au plus grand nombre de documents de notre base. Ils ont été choisis de façon à préférer la génération d'un masque de premier-plan contenant éventuellement plus d'éléments que ce qui est nécessaire, plutôt que de manquer des petites régions, et ainsi prendre le risque de sous-segmenter l'image. Les valeurs suivantes ont donc été appliquées : une fenêtre de recherche de taille 3x3 a été utilisée lors de cette étape, le processus d'hystérésis applique un seuil bas $S_b = 80$, et un seuil haut $S_h = 200$.

Pour le processus de regroupement des SPs en couches couleurs globales, après avoir déterminé la couleur la plus représentative en termes de fréquence d'apparition, nous regroupons tous les SPs dont la distance couleur est inférieure à un seuil. Nous avons déterminé expérimentalement qu'une valeur de 40 convenait le mieux pour notre jeu de données, et permettait de limiter les phénomènes de sur-segmentation apparaissant notamment au niveau des régions textuelles.

6.4 Résultats et discussion

6.4.1 Documents réels

Les figures 6.6 et 6.7 présentent les résultats de la segmentation basée sur les deux versions du SLIC (originale et adaptative) qui sont suivies par l'approche de fusion en couches couleur. Une approche basée sur une segmentation en SPs telle que SLIC donne en effet des résultats intéressants dans le cas des images de documents. Nous pouvons noter une bonne adhérence des régions locales aux contours des images (figures 6.6a, 6.6c, 6.6e, et 6.6g).

Comme nous pouvons l'observer sur les figures 6.7a, 6.7c, 6.7e, et 6.7g, la version adaptative produit des SPs de différentes tailles : grande au niveau des zones homogènes, et plus petite dans les régions avec un gradient fort. Cette approche bénéficie ainsi de la structure particulière des images de documents qui contiennent beaucoup de grandes régions homogènes. L'avantage majeur est que nous pouvons ainsi initialiser au mieux le nombre de SPs souhaité pour une image donnée, en faisant dépendre ce paramètre directement du contenu de l'image. Les valeurs des pas sont des ratios de la largeur du document.

Pour obtenir une bonne segmentation, il faut que les résultats de l'extraction des masques de fond et de premier-plan aient une qualité suffisante. C'est à dire qu'une erreur se produisant lors de cette étape (régions de premier-plan non détectées) peut avoir des répercussions sur les résultats de segmentation obtenus à partir du SLIC. Nous avons choisi une approche basée sur l'exploitation de l'information couleur à travers un gradient vectoriel, et l'application de techniques communément utilisées en analyse d'images de documents telles qu'une approche de lissage. Le fait d'avoir beaucoup de SPs de petite taille dans des régions très bruitées peut amplifier l'im-

paquet du bruit par rapport à la version originale. Nous pouvons notamment observer ce phénomène sur la figure 6.7e. Cependant, le processus de regroupement réduit ce bruit, ce qui permet d'obtenir une segmentation finale similaire à celle du SLIC original (figure 6.6e).

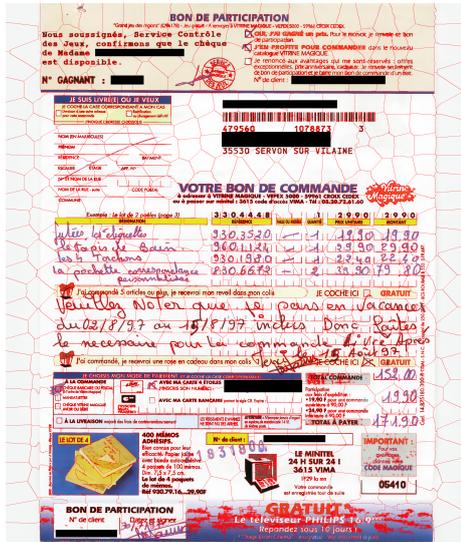
Concernant le processus de regroupement, nous avons trouvé globalement un nombre plus réduit de couches couleur. La segmentation des régions est globalement meilleure, en particulier au niveau des zones de transition qui posaient problème jusque-là. Cependant, nous observons encore quelques erreurs sur les figures 6.6b et 6.7b. Des couleurs peu contrastées comme par exemple la région beige sur fond blanc de l'en-tête ne sont pas séparées. Ceci est principalement dû à la valeur des seuils de regroupement qui ont été fixés expérimentalement, et qui sont génériques. Concernant ce problème, nous n'avons pas trouvé de seuil générique permettant d'obtenir à la fois une bonne segmentation sur les régions faiblement contrastées, tout en limitant les effets de sur- ou de sous-segmentation apparaissant sur d'autres régions de l'image (régions de transition, zones de texte, graphiques, etc).

D'autre part à ce stade, nous n'utilisons plus l'information spatiale. En effet, alors que le SLIC appliquait un processus de clustering local qui permettait de combiner à la fois des informations colorimétriques et spatiales, à l'échelle du document complet, l'information spatiale pourrait tendre à séparer des régions de même couleur localisées à deux endroits très éloignés l'un de l'autre. Or ici, notre objectif est d'obtenir des couches couleur globales regroupant des régions similaires en termes de couleurs peu importe leur localisation sur l'image.

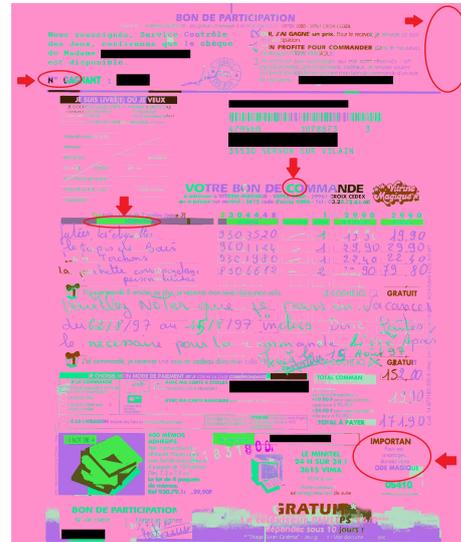
6.4.2 Images synthétiques

Le tableau 6.1 compare les performances du SLIC original et de la version adaptative. Les résultats obtenus avec la version adaptative du SLIC sont comparables à ceux obtenus avec la version originale. La valeur de la précision est légèrement supérieure pour le SLIC adaptatif mais cette différence n'est pas notable. Il est important de noter qu'il est possible de jouer sur la taille des SPs tout en conservant la précision de la version originale du SLIC. Il est donc pertinent de faire dépendre la taille des SPs du contenu de l'image.

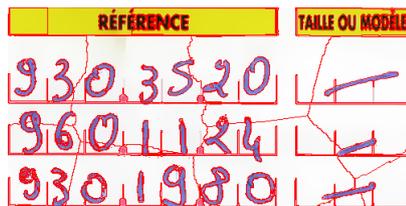
Nous rappelons dans le tableau 6.2 les résultats obtenus à partir de la méthode de classification hiérarchique qui a été décrite dans le chapitre 5. Nous pouvons observer une légère baisse de la valeur du rappel pour les deux approches basées sur SLIC. Cependant, nous notons une très nette augmentation de leurs valeurs de précision comparées à celles obtenues par l'approche de classification hiérarchique appliquée dans les espaces RVB, HSV et L^*a^*b . Notre approche adaptative du SLIC apporte donc un gain important pour la qualité de la segmentation qui s'explique par l'ajout d'informations spatiales. En effet, l'étape de regroupement est faite directement à partir des régions locales générées par l'algorithme SLIC. Ainsi, les problèmes liés



(a) Image segmentée par l'approche du SLIC.



(b) Résultats (en fausses couleurs).



(c) Image segmentée par l'approche du SLIC.



(d) Résultats (en fausses couleurs).



(e) Image segmentée par l'approche du SLIC.



(f) Résultats (en fausses couleurs).

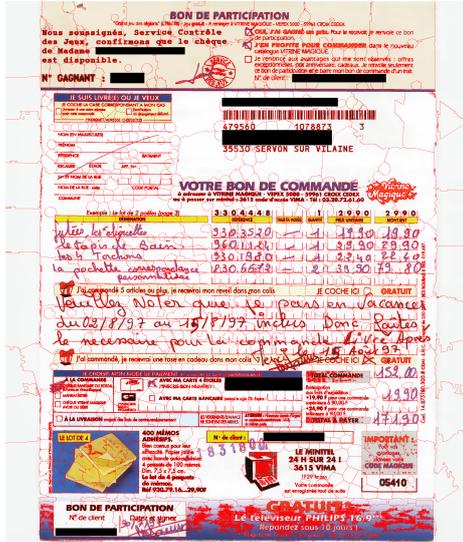


(g) Image segmentée par l'approche du SLIC.

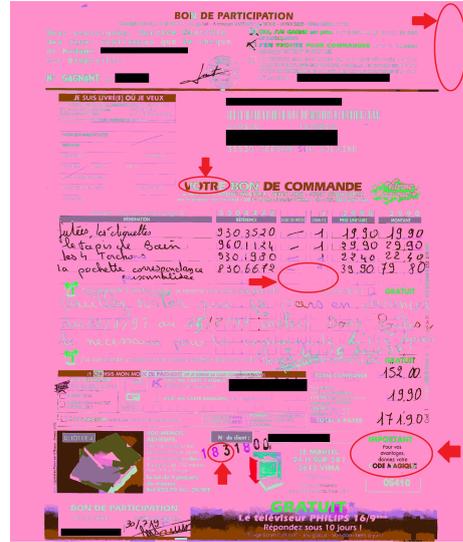


(h) Résultats (en fausses couleurs).

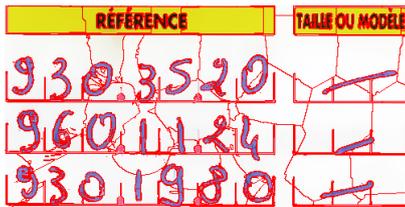
FIGURE 6.6 – Approche de segmentation basée sur le SLIC proposé par [Achanta 2012], suivi par un regroupement des SPs en couches couleur. Le pas S est initialisé à 100. Les cercles en rouge mettent en évidence certaines erreurs de segmentation.



(a) Image segmentée par l'approche du SLIC Adaptatif.



(b) Résultats (en fausses couleurs).



(c) Image segmentée par l'approche du SLIC Adaptatif.



(d) Résultats (en fausses couleurs).



(e) Image segmentée par l'approche du SLIC Adaptatif.



(f) Résultats (en fausses couleurs).



(g) Image segmentée par l'approche du SLIC Adaptatif.



(h) Résultats (en fausses couleurs).

FIGURE 6.7 – Approche de segmentation basée sur le SLIC adaptatif, suivi par un regroupement des SPs en couches couleur. Le pas S est initialisé à 51 pour les éléments de premier-plan et à 255 pour les éléments de fond. Les cercles en rouge mettent en évidence certaines erreurs de segmentation.

	SLIC	SLIC Adaptatif
Précision	0.84	0.85
Rappel	0.93	0.93

TABLEAU 6.1 – Mesure de performance de l’approche basée sur le SLIC original et sur le SLIC adaptatif appliqués sur la base d’images synthétiques.

Classification hiérarchique	RVB	HSV	L*a*b
Précision	0.76	0.67	0.55
Rappel	0.94	0.93	0.98

TABLEAU 6.2 – Performance de la classification hiérarchique sur la base d’images synthétiques.

à la segmentation des zones de transition (caractères-fond) s’en trouvent en partie résolus.

6.5 Conclusion

A travers ce chapitre, nous avons présenté une approche spatio-colorimétrique basée sur une segmentation en superpixels. Cette approche a pour objectif de regrouper localement les pixels en régions perceptuellement homogènes. L’application locale d’une méthode de partitionnement permet la manipulation de données à la fois colorimétriques et spatiales autrement très difficiles à combiner à un niveau global. Notre contribution porte sur l’extension de l’algorithme SLIC proposé par [Achanta 2012] aux propriétés particulières des images de documents.

La version adaptative du SLIC produit des SPs de différentes tailles en fonction du contenu du document. L’idée est de répartir les centres des SPs de manière non homogène en se basant directement sur les caractéristiques structurales de l’image. Les zones homogènes contiennent ainsi de grands SPs alors que les régions pour lesquelles le gradient est élevé, et qui varient beaucoup, contiennent de petits SPs. L’intérêt principal de cette approche est de ne plus fixer le nombre de SPs souhaité mais de l’adapter directement au contenu.

La segmentation en SPs s’est révélée très intéressante pour les images de documents car elle a permis de produire des régions qui adhéraient bien aux contours réels de l’image. La précision en a été augmentée. La segmentation des zones textuelles est meilleure, ce qui est un net avantage pour l’étape de reconnaissance ultérieure. Malgré tout la question du regroupement de ces régions locales en couches couleur

globales pose encore quelques problèmes. Lors de l'étape de regroupement, seule la représentativité de chaque couleur est étudiée. A ce stade, nous n'utilisons plus de données spatiales. Or, la richesse du contenu des images de documents rend le paramétrage des méthodes difficilement adaptable. Les couleurs observables sur les documents administratifs sont très variables, et plus ou moins contrastées d'une image à une autre. Nous sommes donc confrontés à des problèmes de sensibilité au bruit, similaires à ceux que nous avons avec une approche colorimétrique.

Pour les résoudre, nous avons envisagé une analyse multi-résolution qui permettrait d'observer l'image à différents niveaux de détails. L'objectif serait d'extraire les couleurs dominantes à une résolution basse sur laquelle seule la structure globale de l'image de document peut être observée. Une fois les couleurs principales extraites, la segmentation pourrait être affinée en observant des niveaux de détails de plus en plus fins ; c'est à dire en étudiant l'image à des résolutions plus élevées. Le chapitre 7 sera dédié à la description de cette analyse.

Analyse multi-résolution

Sommaire

7.1	Introduction	133
7.2	Contributions	134
7.2.1	Pré-traitements	135
7.2.2	Procédure globale	136
7.2.3	Procédure de fusion	138
7.2.4	Procédure d'étiquetage	139
7.3	Expérimentations	142
7.4	Résultats et discussion	144
7.4.1	Documents réels	144
7.4.2	Images synthétiques	145
7.5	Conclusion	152

7.1 Introduction

Nous avons présenté dans le chapitre 5, deux approches de segmentation colorimétriques dont l'objectif était la segmentation de documents administratifs en couches couleur. Ces approches peuvent être assimilées à des méthodes de réduction couleurs. Malgré des résultats intéressants, elles se sont révélées sensibles au bruit. De plus, l'utilisation de la couleur comme seule source d'information ne permet pas une segmentation précise des régions réelles présentes sur les images de documents. C'est pourquoi nous nous sommes orientés vers une technique intégrant des données spatiales pour rendre le processus plus robuste.

Une approche basée sur l'extraction de superpixels, appelée SLIC, a été présentée dans le chapitre 6. Notre contribution portait sur l'adaptation de cette méthode afin de prendre en compte la structure propre aux images de documents. Nous avons ainsi montré qu'une approche spatio-colorimétrique pouvait améliorer grandement la précision de la segmentation, en générant des régions qui adhéraient bien aux contours réels de l'image.

Une image en haute résolution contient souvent beaucoup trop de détails ce qui rend les traitements complexes à une résolution unique. Ceci est particulièrement

vrai dans nos travaux, notre but étant le traitement de documents de structuration variable et dont les contenus sont très variés. Nous avons notamment rencontré des difficultés à définir une approche de regroupement des régions locales en couches couleur globales qui tienne compte de cette hétérogénéité. Les points clés concernent les règles de décision pertinentes et le choix des seuils. Comment rendre ces seuils adaptatifs, et sur quels critères peut-on fusionner deux régions ? La question se pose lorsque les images à notre disposition sont si différentes en termes de couleurs et de contrastes.

Pour faciliter l'étape de segmentation et améliorer sa performance, nous avons envisagé une analyse multi-résolution afin de simuler le comportement du système de vision humain. Ce dernier perçoit généralement une scène en plusieurs étapes qui vont extraire des informations de différents types. Tout d'abord, un humain aura tendance à regarder une image dans son ensemble pour ne capter que sa structure globale. Ceci est dû à un phénomène qualifié de vision périphérique. Dans un deuxième temps, il se focalisera sur une région d'intérêt pour en extraire les détails. On parle alors de vision fovéale. L'idée principale qui nous intéresse ici est qu'une analyse multi-résolution permet de percevoir les éléments d'un document à différents niveaux de détails, et simule ainsi ce comportement. Un des avantages est de pouvoir adapter les traitements à une résolution donnée.

L'approche que nous proposons, et qui sera présentée à travers ce chapitre, a pour objectif d'extraire les couleurs globales grâce à une segmentation appliquée en basse résolution. A ce niveau, seules les régions principales de l'image sont visibles. Il est donc possible de les localiser grossièrement sans être perturbé par les détails et le bruit présents sur l'image originale. Les résultats doivent être ensuite affinés. Cette étape en revanche doit être appliquée à des résolutions plus élevées afin de capter des éléments fins tels que les régions textuelles. L'objectif du processus que nous allons maintenant présenter est de fusionner peu à peu les régions segmentées en couches couleur, tout en corrigeant le cas échéant les régions pour lesquelles la segmentation ne serait pas suffisamment précise.

7.2 Contributions

Les images de documents contiennent énormément d'éléments différents en termes de caractéristiques, en plus du bruit introduit au cours du processus de numérisation. Une analyse multi-résolution permet d'observer une image à différents niveaux de détails. En basse résolution, seules les structures globales de l'image sont visibles. Plus la résolution sera élevée, et plus les détails observables seront fins.

Pour réduire la quantité d'informations disponibles et limiter l'effet du bruit, nous commençons par localiser grossièrement les couleurs dominantes sur une image en basse résolution. Nous utilisons pour cela le SLIC adaptatif décrit dans la section

6.2.3. L'analyse multi-résolution consiste ensuite à affiner la segmentation obtenue. En effet, il nous faut reconnaître et reconstituer certains éléments fins tels que les graphiques, les zones textuelles, etc. Cette analyse va donc considérer des niveaux de résolution de plus en plus élevés. D'une part, l'idée est de fusionner peu à peu les régions en couches couleur. D'autre part, l'objectif est de vérifier la pertinence des étiquettes qui ont été affectées aux pixels de l'image, et le cas échéant de les corriger.

Dans cette section, nous commencerons par présenter les pré-traitements qui ont été appliqués sur les images de notre base dans le but de réduire le bruit impactant les résultats de la segmentation. Puis, nous présenterons le processus global pour ensuite mettre l'accent sur les différentes étapes de notre analyse multi-résolution.

7.2.1 Pré-traitements

Le processus de numérisation et la compression introduisent du bruit dans les images sous forme d'artéfacts et de fausses couleurs, qui apparaissent notamment le long des contours (Figure 7.1a). Afin de réduire les perturbations liées à ce bruit, nous avons lissé notre ensemble d'images. Pour cela, nous avons choisi le filtre bilatéral proposé par [Tomasi 1998]. Ce filtre préserve les contours, ce qui nous permet de ne pas perdre trop d'informations au niveau des zones sensibles que sont les régions textuelles. L'idée est de calculer la valeur d'un pixel par rapport à celles de ses voisins, en tenant compte à la fois de leur proximité spatiale et colorimétrique.

L'hétérogénéité des images en termes de structure et de qualité, et le manque de connaissance sur le processus d'acquisition, rendent très difficile le paramétrage automatique de la méthode de lissage. Il pourrait être intéressant de détecter le type de document à traiter, ou la qualité de l'image, de façon à adapter la méthode en fonction de ces données. Dans ce mémoire cependant, nous nous sommes focalisés sur l'étape de segmentation. Les paramètres du filtre bilatéral ont été fixés expérimentalement de façon à corriger au mieux le bruit sur la majorité de nos images sans pour autant perturber l'extraction d'information par un lissage trop important. La figure 7.1b présente les résultats.

Nous avons appliqué ce filtre plusieurs fois de manière successive. Les images synthétiques bruitées de notre vérité-terrain ont été traitées de la même façon. Le protocole expérimental a montré que les valeurs ci-dessous pour le filtre bilatéral permettaient d'obtenir le meilleur lissage sur notre jeu de données :

- taille de la fenêtre autour du pixel courant = 15 ;
- $\sigma_{couleur} = 15$ (plus la valeur est élevée, et plus les couleurs distantes en termes colorimétriques, localisées dans le voisinage du pixel seront mélangées) ;
- $\sigma_{spatial} = 60$ (plus la valeur est élevée, et plus les pixels éloignés auront d'influence les uns sur les autres à partir du moment où leur couleur est suffisamment proche) ;
- nombre d'itérations = 5.

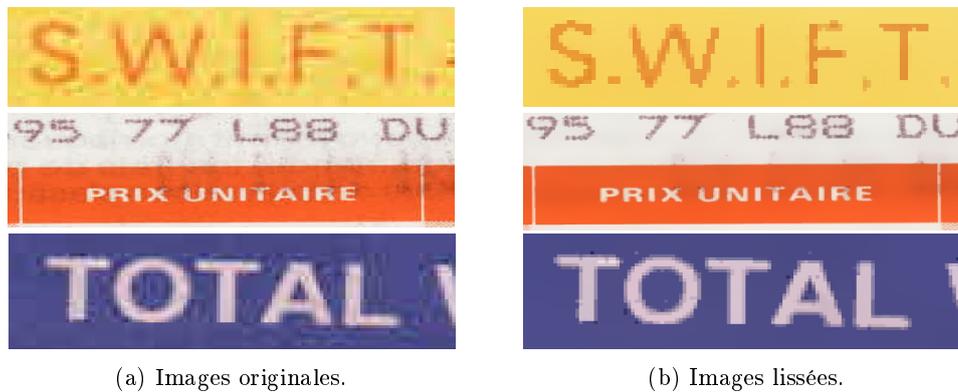


FIGURE 7.1 – Résultats du lissage des images de documents par un filtre bilatéral.

7.2.2 Procédure globale

Un état-de-l'art des approches multi-résolution en analyse de documents a été présenté dans la section 2.3.2. L'idée est de simplifier progressivement l'image pour obtenir une pyramide d'images. A chaque changement de niveau, l'image est lissée et sa taille est divisée par deux. Certains éléments sont plus ou moins visibles en fonction du niveau de détails sur lequel on se focalise. L'intérêt principal est que les traitements peuvent être adaptés à une résolution donnée. Par exemple, il peut être difficile de distinguer les caractères du bruit dans un bloc de texte en haute résolution. En revanche, en basse résolution, les lignes de texte apparaissent. Celles-ci sont plus faciles à détecter, et serviront de base pour localiser les régions d'intérêt à une résolution plus élevée.

Le processus complet d'analyse se résume en une phase de segmentation, appliquée uniquement en basse résolution, suivi par une procédure de nettoyage et de vérification (étape d'étiquetage), appliquée de la même manière à différentes résolutions. La figure 7.2 présente ce processus. Les termes de couleurs, de clusters, ou de régions seront employés pour un ensemble de pixels partageant des caractéristiques colorimétriques proches, et constituant une couche couleur. Nous travaillons dans l'espace couleur L^*a^*b qui est un espace perceptuellement uniforme pour lequel la pertinence de la distance euclidienne est reconnue pour la mesure de distance entre deux couleurs.

Avant de poursuivre sur la procédure globale, nous allons expliciter un pré-traitement qui est appliqué à chaque niveau de résolution, pour aider à la prise de décision de fusionner ou non deux couleurs proches (voir section 7.2.3). Au sujet de cette notion de proximité, il est difficile de trouver une bonne valeur de seuil. Nous avons donc décidé de diviser l'espace couleur en blocs et de fusionner les couleurs si elles se trouvent à une distance inférieure à N blocs l'une de l'autre. Pour cela, nous appliquons une quantification uniforme qui génère des blocs uniformes de taille Z . Les

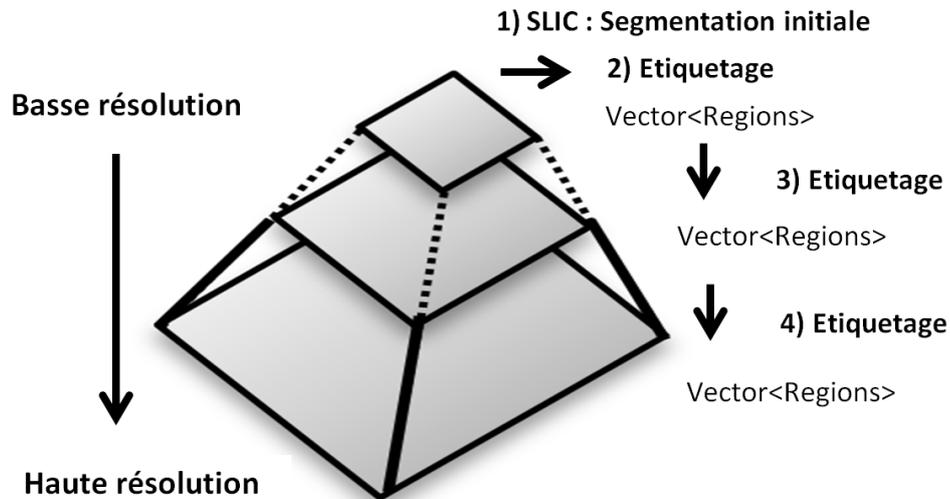


FIGURE 7.2 – Analyse multi-résolution.

pixels de l'image sont ensuite distribués dans ces blocs. Chacun des canaux est codé sur l'intervalle $[0, 255]$. Nous avons déterminé expérimentalement qu'un pas $Z = 4$ permettait de réduire grandement le nombre de couleurs sans perdre trop d'informations. Nous passons donc de 256^3 , soit plus de 16 millions de couleurs potentielles, à $(256/4)^3$, soit 64 fois moins de couleurs.

Pour visualiser l'image à différents niveaux de détails, nous avons utilisé une pyramide gaussienne. A chaque changement de résolution, l'image est sous-échantillonnée par une opération de convolution avec une gaussienne, puis elle est redimensionnée. Nous cherchons dans un premier temps à détecter uniquement les couleurs principales. C'est pourquoi, nous étudions tout d'abord l'image dans sa globalité, en basse résolution. A ce niveau, les résultats de la segmentation sont grossiers, mais l'effet du bruit a moins d'impact sur le processus. De plus, la taille de l'image étant réduite, les traitements sont plus rapides.

La méthode de segmentation appliquée ici est le SLIC adaptatif qui a été présenté dans la section 6.2.3. L'idée est de réaliser une sur-segmentation pour réduire la complexité de l'image. Elle combine à la fois des informations colorimétriques et des informations spatiales. Cette approche a été choisie pour ses capacités à s'adapter au contenu des images de documents, et à produire des régions qui respectent les structures géométriques locales de l'image. Sans être précises, ses structures donnent une bonne indication des caractéristiques colorimétriques de l'image.

La segmentation produit un ensemble de régions locales qui doivent ensuite être fusionnées en couches couleur globales. Il faut aussi améliorer la précision des résultats. En effet, à ce niveau, les éléments les plus fins tels que les zones de texte ne sont pas extraits. La segmentation est alors affinée, et éventuellement corrigée,

grâce à une analyse multi-résolution qui a pour objectif d'étudier des niveaux de détails de plus en plus précis. Cette étape qui est appelée procédure d'étiquetage va être progressivement appliquée à des résolutions plus élevées. Elle est décomposée en deux étapes :

- une étape de fusion, dont le but est de nettoyer les résultats en fusionnant les couleurs trop proches, ou qui ne constituent pas des régions compactes ;
- l'étape d'étiquetage proprement dite, qui vérifie la pertinence des labels attribués aux pixels, et si nécessaire les corrige.

7.2.3 Procédure de fusion

L'étape de segmentation permet d'obtenir un ensemble de régions homogènes et qui adhèrent aux contours des grandes structures de l'image. Cependant, il reste quelques problèmes de bruit qui entraînent une sur-segmentation de l'image. Le bruit est notamment visible au niveau des zones de transition où l'on verra apparaître beaucoup de clusters dont les pixels sont très dispersés à travers l'image (figure 7.3b). L'étape de fusion a pour objectif de nettoyer les résultats et de réduire peu à peu le nombre de couches couleur à une quantité raisonnable. Seules les plus pertinentes en fonction de leurs propriétés colorimétriques, ou de compacité vont être conservées.

Dans un premier temps, les petites régions sont supprimées. Le but est de détecter les zones de transition et de les corriger en leur affectant une nouvelle étiquette déterminée en fonction des couleurs de fond et de premier-plan qui se trouvent dans leur voisinage. Pour cela, nous étudions la distribution de la couleur de chaque pixel dans une fenêtre 3×3 centrée sur ce pixel. Lorsque cette valeur est faible à l'intérieur de la zone de recherche, nous recherchons la nouvelle étiquette en nous basant sur l'hypothèse que seulement deux couleurs doivent apparaître dans une fenêtre de cette taille. Nous supposons que la couleur de fond est celle qui a la plus grande distribution, et que la couleur de premier plan est celle qui a la plus grande valeur de distance par rapport au fond détecté.

Nous cherchons ensuite à fusionner les couleurs qui sont très proches. En pré-traitement, les images ont été quantifiées de manière uniforme. L'espace couleur a donc été divisé en blocs de taille Z . Ce paramètre a été fixé expérimentalement à 4, qui est une valeur suffisamment petite pour ne pas perdre trop d'informations, tout en réduisant grandement le nombre de couleurs. Plutôt que de proposer une valeur de seuil pour fusionner deux couleurs en fonction de leur distance, nous avons préféré considérer une distance en nombre de blocs qui nous a paru plus souple. Nous partons de l'hypothèse que deux couleurs localisées à une distance de N blocs l'une de l'autre dans l'espace colorimétrique ne peuvent pas être distinguées par le système visuel humain. En basse résolution où les couleurs sont dégradées, nous prenons un pas de un bloc. Nous augmentons cette valeur pour l'image en pleine résolution.



(a) Image originale.



(b) Résultats de la segmentation appliquée en basse résolution.



(c) Après l'étape de fusion appliquée sur 7.3b.

FIGURE 7.3 – Application de l'étape de fusion sur les régions obtenues à partir d'une segmentation appliquée en basse résolution. L'objectif est le nettoyage de l'image en réduisant notamment les clusters de bruit associés aux zones de transition.

Pour terminer, nous cherchons à supprimer les clusters de bruit qui contiennent des pixels qui sont très dispersés dans l'image. Nous mesurons pour chacun des clusters le degré de connexité proposé par [Macaire 2006] et nous le seuillons pour ne conserver que les régions fortement connexes. L'idée est de compter pour chaque pixel appartenant à un cluster donné, le nombre de pixels ayant la même étiquette dans son voisinage. Le degré de connexité correspond à la valeur moyenne de cette mesure. Plus cette valeur est proche de 1, et plus le cluster est compact.

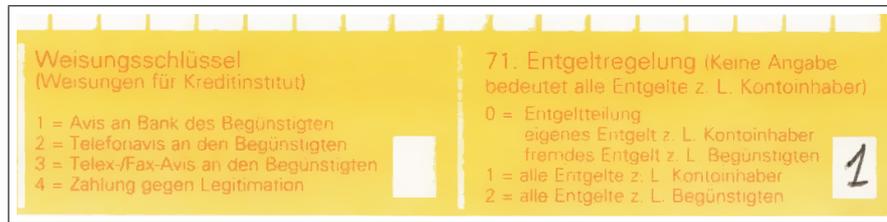
Un exemple du résultat de ce processus peut être observé sur la figure 7.3c.

7.2.4 Procédure d'étiquetage

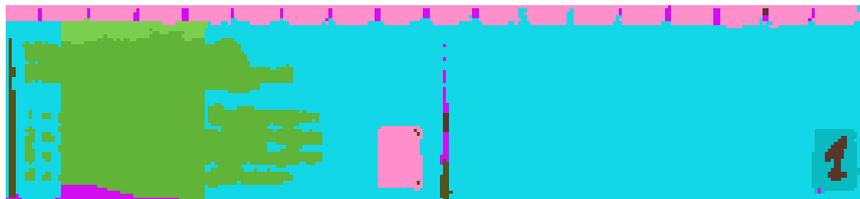
Après les étapes de segmentation et de nettoyage, les régions locales doivent être peu à peu regroupées en couches couleur globales. Plus le processus monte en résolution, et plus les détails de l'image apparaissent. La segmentation peut alors être affinée afin de retrouver les régions les plus fines. Par exemple, en basse résolution, il n'est pas possible de reconnaître des régions textuelles. Cette procédure est appliquée plusieurs fois de manière itérative.

L'étape d'étiquetage a plusieurs objectifs :

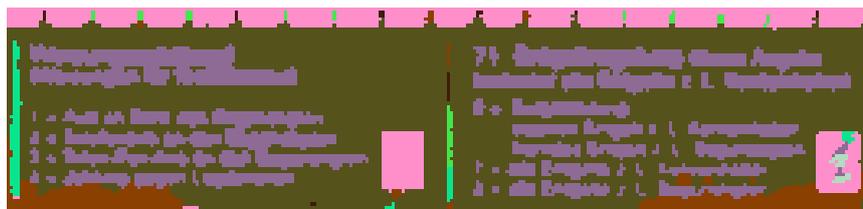
- retrouver des régions qui n'ont pas été correctement segmentées, notamment en basse résolution (figure 7.4c) ;
- fusionner les éléments de même couleur qui sont localisés à l'intérieur d'une fenêtre de recherche, ici même ligne ou colonne (figure 7.5c) ;
- reconstituer les caractères, notamment en haute résolution (figure 7.6c).



(a) Image originale.



(b) Après l'étape de fusion appliquée en basse résolution.



(c) Après la phase d'étiquetage appliquée sur 7.4b.

FIGURE 7.4 – La phase d'étiquetage permet de retrouver des régions qui n'avaient pas été correctement segmentées.

Exemple : Le lot de 2 poêles (page 3)					
DÉSIGNATION	RÉFÉRENCE	TAILLE OU MODÈLE	QUANTITÉ	PRIX UNITAIRE	MONTANT
	3 3 0 4 4 4 8		1	2 9 9 0	2 9 9 0

(a) Image originale.

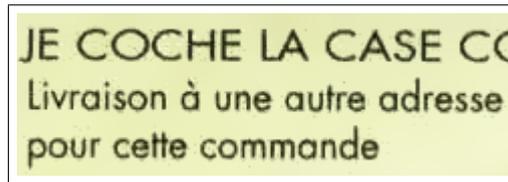


(b) Après l'étape de fusion appliquée en basse résolution.



(c) Après la phase d'étiquetage appliquée sur 7.5b.

FIGURE 7.5 – La phase d'étiquetage permet de fusionner tous les éléments d'une même couleur qui sont localisés à l'intérieur d'une fenêtre de recherche (ici même ligne ou colonne).



(a) Image originale.



(b) Après l'étape de fusion appliquée en haute résolution.



(c) Après la phase d'étiquetage appliquée sur 7.6b.

FIGURE 7.6 – La phase d'étiquetage permet de reconstituer les caractères.

Dans cet objectif, nous utilisons les labels de couleurs qui ont été obtenus à une résolution basse, et les informations colorimétriques fournies par l'image L^*a^*b de la résolution supérieure. Nous parcourons les pixels de l'image pour vérifier que les étiquettes qui leur ont été affectées sont correctes. Les résultats peuvent en effet être améliorés si nous avons une idée des couleurs qui sont observables dans le voisinage du pixel étudié. L'idée est donc de parcourir les couleurs détectées dans une fenêtre de recherche autour du pixel, et de mesurer leur distance colorimétrique. Nous utilisons ici une distance euclidienne. Cette valeur est comparée à la distance mesurée entre la valeur $[L, a, b]$ du pixel et la couleur qui lui a été affectée. S'il existe une couleur plus proche, l'étiquette du pixel est modifiée. Le processus est appliqué de gauche à droite, et de haut en bas, puis une deuxième fois de droite à gauche, et de bas en haut. Ceci est fait pour éviter les apparitions d'effets de lignes ou de colonnes causées par l'ordre de balayage. A la fin du processus, une étape de fusion telle que celle décrite dans la section 7.2.3 est appliquée, ce qui permet d'éliminer une partie des couches couleur non pertinentes (zones de transition).

Au sujet de la fenêtre de recherche, nous supposons que la plupart des éléments d'une image de documents ont une orientation horizontale ou verticale : en-têtes de colonnes, alignement de labels, graphiques, etc. Notre recherche est donc effectuée dans deux directions en définissant des zones rectangulaires, centrées sur le pixel courant, telles que présentées sur la figure 7.7. Pour accélérer le processus, nous considérons seulement un pixel sur N . En basse résolution, où nous nous concentrons sur la structure globale, nous parcourons une grande zone de recherche. Cela nous permet ainsi de retrouver des régions qui n'ont pas été correctement segmentées (figure 7.5c), et de fusionner des éléments alignés sur la page tels que des en-têtes (figure 7.5c). En revanche, en haute résolution, nous n'étudions que le voisinage proche du pixel pour éviter d'introduire trop de bruit. L'objectif est plutôt de reconstituer certains éléments fins comme les caractères (figure 7.6c).

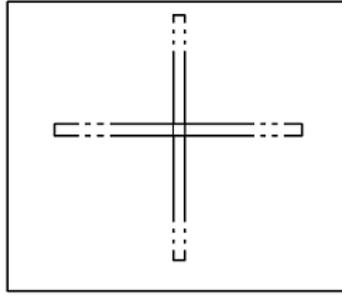


FIGURE 7.7 – Fenêtre de recherche utilisée pendant la phase d’étiquetage.

7.3 Expérimentations

Nous avons réalisé deux séries d’expériences. Premièrement, nous avons appliqué le SLIC adaptatif suivi de la procédure d’étiquetage décrite dans ce chapitre : successions de fusion et d’étiquetages. Nous avons ensuite comparé les résultats à ceux obtenus à partir d’une étape de fusion basée sur l’étude de la représentativité des couleurs, qui a été décrite dans le chapitre 6. Notre objectif était de vérifier l’hypothèse selon laquelle la prise en compte du contexte à travers l’étude du voisinage autour des pixels corrigerait la segmentation, ce que ne permet pas un processus global.

Dans un deuxième temps, nous avons analysé les performances de l’analyse multi-résolution. Nous avons donc expérimenté l’analyse présentée dans ce chapitre avec une pyramide à trois niveaux de résolutions différents. Les résultats ont ensuite été comparés avec la segmentation générée à partir du SLIC adaptatif appliquée uniquement sur une image en haute résolution. Le regroupement des régions locales en couches couleur est effectué avec la même procédure.

Dans le cas de l’analyse multi-résolution, nous avons modifié l’initialisation des paramètres utilisés par le SLIC adaptatif pour les rendre dépendants de la résolution. Soient l , la largeur de l’image, et $level$, le niveau de résolution. Nous travaillons avec une pyramide à trois niveaux de résolutions, telles que l’image en haute résolution est associée au niveau 0, et l’image en basse résolution correspond au niveau 2. Ainsi, les paramètres contrôlant le lissage (rl_{th}) et les pas utilisés pour les régions de fond (S_{fd}) et de premier-plan (S_p), sont calculés comme indiqué ci-dessous :

$$rl_{th} = \frac{level^2 * l}{70} \quad (7.1)$$

$$S_{fd} = \frac{level^2 * l}{10} \quad (7.2)$$

$$S_p = \frac{level^2 * l}{50} \quad (7.3)$$

Les paramètres utilisés lors de l'analyse multi-résolution ont été fixés expérimentalement. Ils concernent les seuils utilisés pendant l'étape de fusion. Nous avons tenté de rendre le seuillage de la distance couleur plus souple en considérant les blocs de l'espace couleur qui ont été générés à partir d'une quantification uniforme. Les différents seuils ont été fixés de façon à être suffisamment faibles pour ne pas perdre d'informations. L'objectif était d'éliminer uniquement les régions associées au bruit de façon évidente telles que les zones de transition entre les éléments de premier-plan et le fond. A titre d'information, nous indiquons ci-dessous les valeurs des paramètres et leur impact :

- la distribution *dist* : plus la valeur est élevée, et plus les éléments fins risquent d'être perdus (trop peu de pixels voisins ayant la même étiquette dans la fenêtre de recherche). Ici, tous les pixels ayant une distribution inférieure à 3 sont reclassés. Ce seuil a été appliqué de la même façon pour toutes les résolutions ;
- la compacité *C* : plus cette valeur est proche de 1 et plus le cluster est compact. Ici, nous avons fixé le seuil à 0.55 pour toutes les résolutions ;
- la distance couleur *dc* : plus cette valeur est élevée et plus des couleurs éloignées d'un point de vue colorimétrique seront fusionnées. Nous avons fixé ce seuil à 1 pour les images en basse et en moyenne résolutions, et à 3 pour la résolution la plus élevée.

Pour l'étape d'étiquetage, les seuils concernent le nombre d'itérations et les fenêtres de recherche. Ce sont les valeurs les plus importantes du processus. En effet, plus le nombre d'itérations sera élevé et plus le processus sera lent, et en particulier pour les images en haute résolution. Cependant, cette valeur permet de converger vers un nombre raisonnable de couches couleur en supprimant les régions de bruit. D'autre part, la fenêtre de recherche et le pas d'échantillonnage doivent être suffisamment précis pour ne pas louper de couleurs, ni introduire trop de bruit. En basse résolution, nous avons remarqué que nous pouvions nous permettre une fenêtre de recherche de grande taille. En effet, l'objectif est d'extraire les structures globales de l'image. En revanche en haute résolution, cette fenêtre doit être très petite car sinon le processus introduit du bruit car nous travaillons à un niveau pixel. Nous perdons ainsi le bénéfice de l'analyse multi-résolution.

L'avantage principal de cette approche est qu'elle nous permet de profiter des éléments qui se répètent pour corriger une éventuelle erreur de segmentation. Cette vérification est faite à un niveau pixel, ce qui augmente le temps de traitement. Cependant, cette étape s'est révélée nécessaire, et en particulier en haute résolution, pour pouvoir corriger des éléments essentiels (ex : reconstruction de caractères). Pour information, nous donnons les valeurs de ces paramètres pour chaque résolution dans les tableaux 7.1 (analyse multi-résolution) et 7.2 (méthode de comparaison).

Soient *nbIter*, le nombre d'itérations, *lFenH*, la taille de la zone de recherche verticale au-dessus et en dessous d'un pixel donné, *lFenL*, la taille de la zone de recherche

	nbIter	lFenH	lFenL	N	dist	C	dc
Résolution 2 (basse)	5	$\frac{l_2}{2}$	$\frac{h_2}{2}$	20	3	0.55	1
Résolution 1 (moyenne)	5	$\frac{l_1}{8}$	$\frac{h_1}{8}$	40	3	0.55	1
Résolution 0 (haute)	3	5	5	1	3	0.55	3

TABLEAU 7.1 – Valeurs des paramètres utilisés pendant l'étape d'étiquetage appliquée lors de l'analyse multi-résolution pour chaque niveau de résolution.

	nbIter	lFenH	lFenL	N	dist	C	dc
Etape 1	1	$\frac{l_0}{2}$	$\frac{h_0}{2}$	40	3	0.55	1
Etape 2	1	$\frac{l_0}{8}$	$\frac{h_0}{8}$	80	3	0.55	1
Etape 3	1	5	5	1	3	0.55	3

TABLEAU 7.2 – Valeurs des paramètres utilisés lors de l'application du SLIC adaptatif sur l'image en haute résolution dans un but de comparaison. L'étiquetage s'est déroulé en trois étapes dans le but de faire varier la taille de la fenêtre de recherche.

horizontale à gauche et à droite d'un pixel donné, et N , le pas de balayage utilisé pour accélérer la recherche. Nous notons l_i et h_i , les valeurs respectives de la largeur, et de la hauteur de l'image avec i , le niveau de résolution correspondant.

7.4 Résultats et discussion

7.4.1 Documents réels

Les figures 7.8 et 7.9 comparent des couches couleur obtenues à partir de deux approches de fusion des régions locales générées par le SLIC adaptatif : une approche basée sur la représentativité globale des couleurs (décrite dans la section 6), et une autre qui prend en compte le contexte à travers l'étude du voisinage autour des pixels.

La première méthode recherche les couleurs les plus représentatives en termes de fréquence d'apparition (en nombre de SPs), et fusionne ensuite toutes les couleurs qui se trouvent à une distance inférieure à un seuil (fixé à 40). Celui-ci a été déterminé expérimentalement. Cette méthode donne des résultats intéressants. Cependant, nous pouvons voir sur les résultats que cette technique est très sensible au bruit. Certaines régions ne sont que partiellement détectées (figure 7.8b). La figure 7.9b montre que des régions de transition (bruit) sont segmentées en même temps que les peignes. Les figures 7.8c et 7.9c montre que l'étude des couleurs présentes

	V1 : représentativité	V2 : étude du voisinage
Précision	0.85	0.92
Rappel	0.93	0.94

TABLEAU 7.3 – Performance de l'extraction des couches couleur extraites à partir d'une segmentation basée sur le SLIC adaptatif suivie par l'application de deux approches de fusion : une approche basée sur la représentativité globales des couleurs (V1), et une autre basée sur l'étude des couleurs présentes dans le voisinage de chaque pixel (V2).

dans une fenêtre de recherche autour du pixel étudié corrige ces défauts. Globalement, les résultats obtenus de cette manière sont nettement moins bruités.

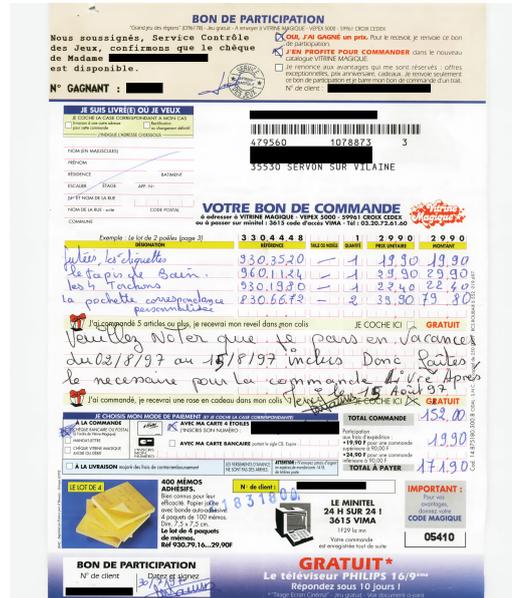
Dans un deuxième temps, nous avons appliqué l'analyse multi-résolution présentée dans ce chapitre. Les résultats sont observables sur les figures 7.10, 7.11 et 7.12. Nous pouvons voir que les régions générées sont beaucoup plus consistantes que dans le cas des approches basées uniquement sur la couleur. Les résultats sont comparables voire même meilleurs que ceux obtenus à partir du SLIC adaptatif appliqué uniquement sur une résolution. Nous pouvons encore noter quelques artéfacts de bruit près des caractères (figure 7.12d), mais leur impact a été fortement diminué. Malgré quelques erreurs, les zones fines telles que les régions textuelles ont été relativement bien segmentées. Comme nous l'avons vu sur la figure 7.4c, l'analyse multi-résolution permet de corriger certaines zones n'ayant pas été détectée pendant l'étape de segmentation en basse résolution.

La section suivante analysera les résultats de l'évaluation de ces méthodes sur la base de vérité-terrain.

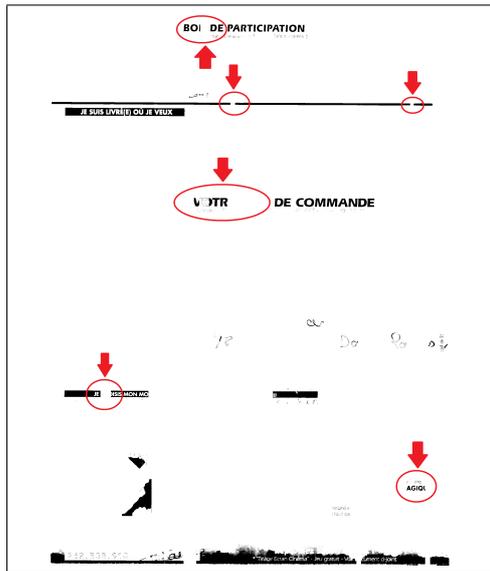
7.4.2 Images synthétiques

Le table 7.3 compare les deux approches qui ont été proposées pour fusionner des régions locales en couches couleur globales. Les valeurs de précision et de rappel confirment que la prise en compte du voisinage des pixels améliore les résultats. Nous passons ainsi d'une précision de 0,85 pour l'approche présentée dans le chapitre 6 qui était basée sur la représentativité en termes de fréquence d'apparition, à une précision de 0,92 pour la version V2 basée sur l'étude du voisinage.

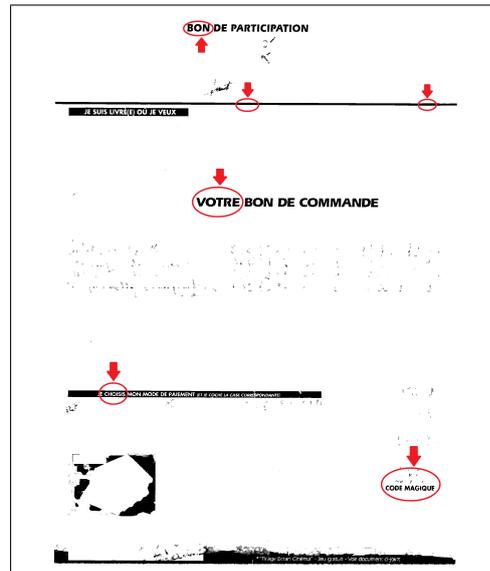
Le tableau 7.4 compare les résultats obtenus à partir de l'approche basée sur une analyse multi-résolution, avec un processus appliqué à une résolution unique. Nous pouvons noter une baisse de la valeur de précision dans le cas de l'analyse multi-résolution, mais qui est égal à la précision obtenue à partir de la combinaison du SLIC adaptatif et de la méthode de fusion basée sur la représentativité. Nous ob-



(a) Image originale.



(b) V1 : Couche bleu foncée.



(c) V2 : Couche bleu foncée.

FIGURE 7.8 – Comparaison de couches couleur obtenues à partir de deux approches de fusion des régions locales générées par le SLIC adaptatif : une approche basée sur la représentativité globale des couleurs (V1), et une autre basée sur l'étude des couleurs présentes dans le voisinage de chaque pixel (V2), partie 1.

BON DE PARTICIPATION
 Nous soussignés, Service Contrôle des Jeux, certifions que le chèque de Madame [redacted] est disponible.
 N° GAGNANT : [redacted]

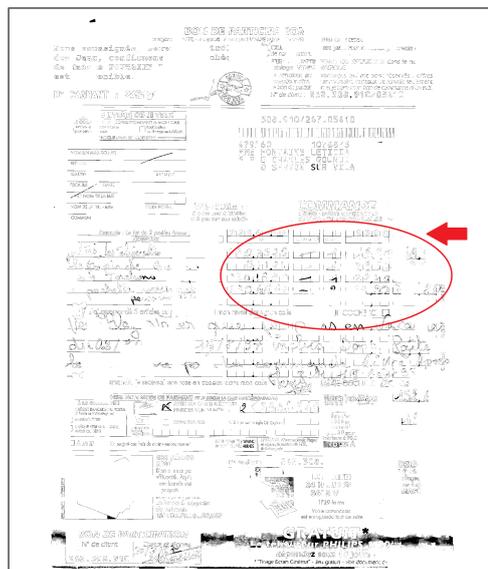
JE SUIS LIVRÉ(E) OU JE VEUX
 479560 1078873 3
 35530 SERVON SUR VILAINE

VOTRE BON DE COMMANDE
 à adresser à VITRINE MAGIQUE - VITRINE 5000 - 09951 CREUX CREUX
 ou à passer sur MINTEL - 3615 code d'accès TOTAL. Tél. 02.30.72.12.01

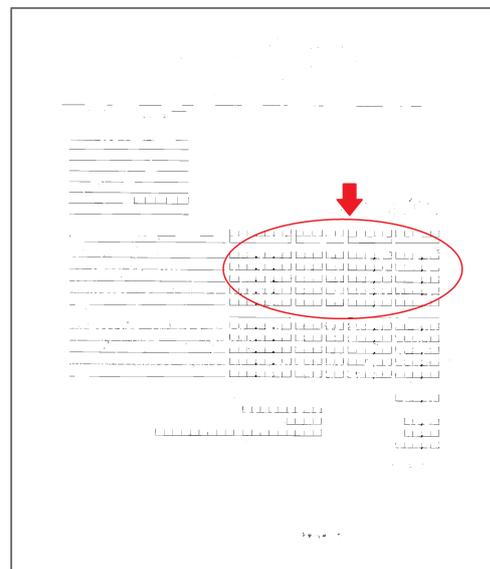
Exemple (à la fin de 2 pages (page 3))	3	3	0	4	4	4	8	1	2	9	9	0	2	9	9	0
Jeux laquette	330	3520														19 90
Le tapis de bain	960	1124														29 90
Les 4 Tachans	530	1820														22 40
La pochette correspondance	830	6673														79 80

TOTAL COMMANDE 152,00
TOTAL À PAYER 174,90

(a) Image originale.

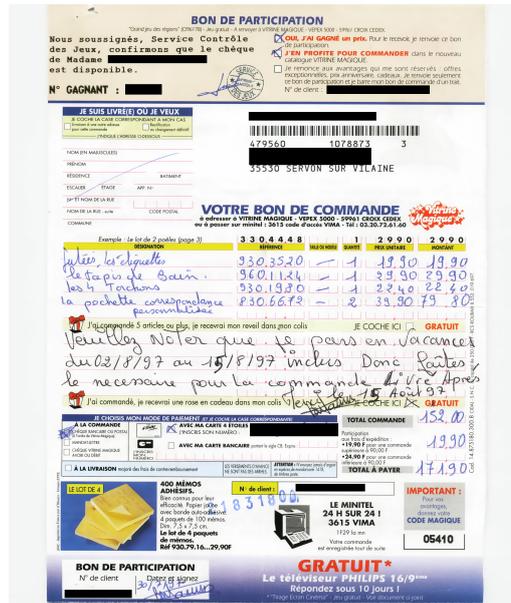


(b) V1 : Couche rose.



(c) V2 : Couche rose.

FIGURE 7.9 – Comparaison de couches couleur obtenues à partir de deux approches de fusion des régions locales générées par le SLIC adaptatif : une approche basée sur la représentativité globale des couleurs (V1), et une autre basée sur l'étude des couleurs présentes dans le voisinage de chaque pixel (V2), partie 2.

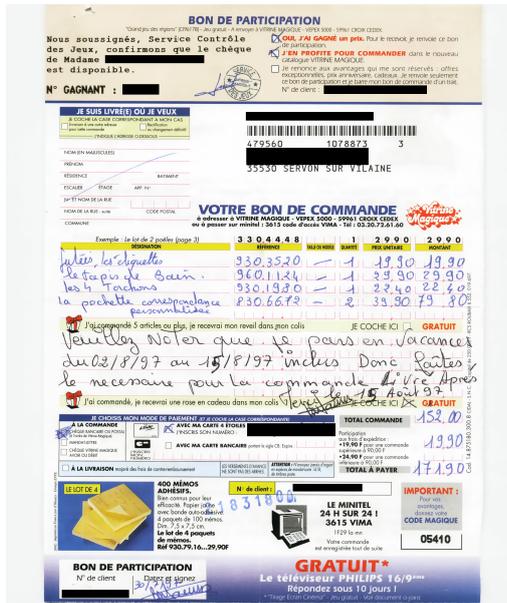


(a) Image originale.

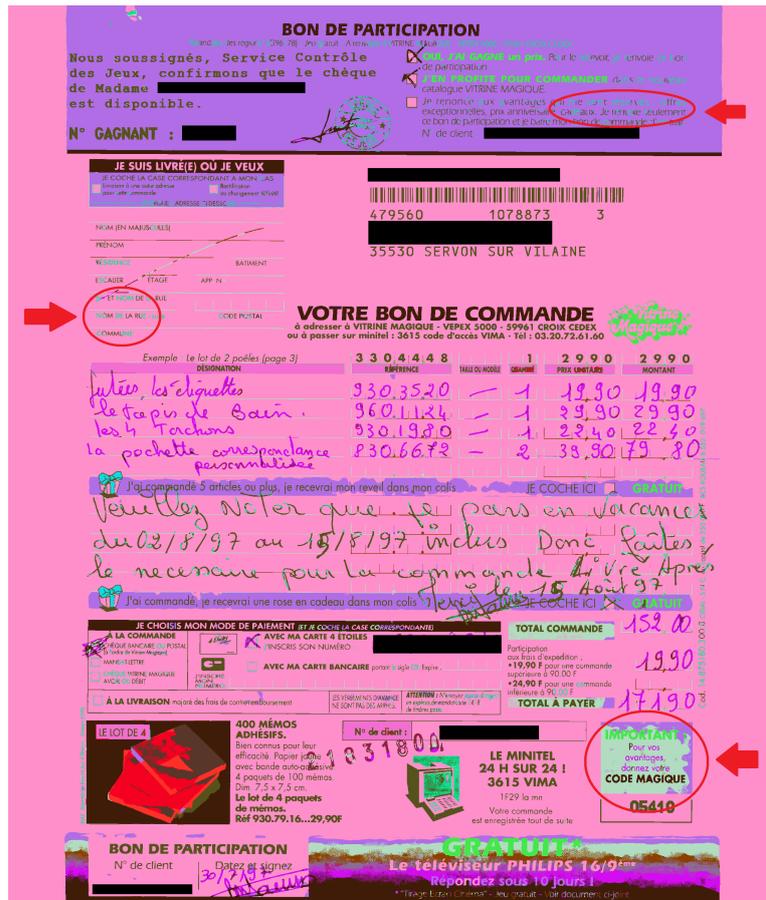


(b) Résultats (en fausses couleurs).

FIGURE 7.10 – Résultats de la segmentation basée appliquée en basse résolution sur un document réel. Les cercles en rouge mettent en évidence certaines erreurs de segmentation.



(a) Image originale.



(b) Résultats (en fausses couleurs).

FIGURE 7.11 – Résultats de la segmentation finale obtenue à partir de l'analyse multi-résolution appliquée à un document réel. Les cercles en rouge mettent en évidence certaines erreurs de segmentation.



FIGURE 7.12 – Résultats de la segmentation basée sur l’analyse multi-résolution appliquée à un document réel. Les cercles en rouge mettent en évidence certaines erreurs de segmentation.

servons une augmentation de la valeur de rappel pour l’analyse multi-résolution. Cela signifie que cette analyse permet de produire moins de couches couleur. Mais qu’il reste encore des zones associées à du bruit sur ces couches. Notre hypothèse était que la précision de la segmentation serait meilleure que celle obtenue à une résolution unique. Cette différence peut s’expliquer par l’utilisation de seuils mal adaptés notamment au niveau des fenêtres de recherche.

La valeur de la précision pour l’analyse multi-résolution peut sembler contradictoire par rapport aux conclusions précédentes : 0.85 pour l’analyse multi-résolution contre 0.92 qui l’application en haute résolution uniquement. L’analyse des cas pour lesquels la valeur de la précision était basse a mis en évidence que l’approche appliquée en haute-résolution uniquement conduisait à une sur-segmentation : parfois jusqu’à deux fois plus de couches couleur détectées que la vérité-terrain. Au contraire, la segmentation basée sur l’analyse multi-résolution permettait de détecter un nombre de couches couleur très proches de la réalité. Nous avons un rap-

	Multi-résolution	Haute résolution
Précision	0.85	0.92
Rappel	0.97	0.94

TABLEAU 7.4 – Comparaison de l’approche basée sur une analyse multi-résolution, avec un processus appliqué à une résolution unique.

	Multi-résolution	Haute résolution
SLIC adaptatif : segmentation	141	175
Basse résolution : Etiquetage	23	-
Résolution moyenne : Etiquetage	20	-
Haute résolution : Etiquetage	16	25

TABLEAU 7.5 – Nombre de labels extraits à chaque étape du processus pour le document réel, figure 7.11a.

pel de 0,97 qui montre une amélioration importante. Cependant, l’analyse multi-résolution peut entraîner une sous-segmentation ce qui explique la différence de précision entre les deux méthodes. Cependant, dans notre contexte, nous préférons une sous-segmentation, en particulier lorsqu’elle implique des couleurs très proches, quand les régions obtenues sont cohérentes, plutôt qu’une sur-segmentation, qui produira de nombreuses couches couleur bruitées difficiles à exploiter. En effet, dans le cas d’une sous-segmentation, les éléments apparaîtront sur une seule image qui pourra être exploitée relativement facilement. Dans le cas d’une sur-segmentation au contraire, les régions apparaîtront sur plusieurs couches, et des traitements supplémentaires seront nécessaires pour les reconstituer.

Le tableau 7.5 indique le nombre de labels obtenus à chaque étape du processus. Leur nombre final dépend des seuils employés. Il est intéressant de noter que nous obtenons un nombre de couleurs raisonnable dès la première étape appliquée en basse-résolution. L’intérêt principal est que l’image étant beaucoup plus petite en basse résolution qu’en haute résolution, les traitements seront plus rapides. L’analyse multi-résolution permet ensuite d’affiner la segmentation d’un point de vue globale, tout d’abord, pour terminer par une vision locale en haute résolution. A une résolution donnée, l’étape d’étiquetage est même capable de corriger certaines régions non détectées quand la couleur qui devrait leur être associée apparaît dans la fenêtre de recherche. Cette étape d’étiquetage s’est révélée nécessaire pour reconstituer proprement de petites régions telles que les zones textuelles.

7.5 Conclusion

A travers ce chapitre, nous avons présenté une approche multi-résolution qui permet d'analyser une image de documents à plusieurs niveaux de détails. Nous nous sommes tournés vers ce type d'approche devant la difficulté de traiter des documents à une résolution unique. Cette analyse consiste à segmenter l'image en basse résolution de façon à en extraire la structure globale. Puis la segmentation est affinée par l'application successive à des résolutions plus élevées d'étapes de nettoyage et de vérification. Ce processus simule ainsi le fonctionnement du système visuel humain qui regarde tout d'abord une scène dans son ensemble avant de se focaliser sur des détails précis.

Le processus de segmentation en basse résolution est basé sur le SLIC adaptatif que nous avons proposé au cours de ces travaux de recherche. En effet, cette approche a été adaptée spécifiquement à la structure des images de documents. De plus, sa capacité à adhérer aux contours de l'image le rend relativement robuste à la présence de bruit. L'étape de nettoyage a pour objectif d'éliminer les couleurs non cohérentes. Celles-ci correspondent la plupart du temps aux zones de transition. Elles peuvent être réduites voire supprimées de façon heuristique sur la base de critères colorimétriques, de compacité, ou en fonction des distributions locales des couleurs. L'étape d'étiquetage quant à elle a pour objectif de vérifier que la couleur affectée à un pixel est correcte. Pour cela, l'idée est de regarder dans une zone de recherche voisine s'il n'en existe pas de plus appropriées. Le cas échéant, l'étiquette est modifiée. L'intérêt de cette approche est de pouvoir profiter des éléments qui se répètent sur un document : colonnes de labels, lignes de texte, etc.

Une fois la segmentation obtenue, l'analyse multi-résolution est faite au niveau pixel. C'est donc un processus relativement coûteux en temps de calcul. Cependant, cette étape permet de localiser des régions n'ayant pas été détectées au moment de la segmentation. De plus, elle est essentielle pour reconstituer des régions fines qui nécessitent une segmentation précise (texte, graphiques, etc).

Un nombre raisonnable de couleurs est détecté dès la basse résolution, ce qui permet de diminuer les temps de calculs, la taille de l'image étant beaucoup plus petite qu'en haute résolution. Les régions produites sont globalement plus consistantes. Nous avons comparé cette approche avec une segmentation appliquée uniquement en haute résolution mais pour laquelle le processus de nettoyage et de vérification sont similaires. L'analyse quantitative montre que la précision est meilleure dans le cas de la segmentation en haute résolution. Cependant, les résultats indiquent une tendance à la sur-segmentation. Au contraire, les résultats produits par l'analyse multi-résolution sont plus proches de la réalité, même si cette méthode peut produire des couches légèrement sous-segmentée.

Conclusion générale

Sommaire

8.1 Conclusion	153
8.2 Perspectives scientifiques	156
8.2.1 Affectation des pixels à plusieurs couches	156
8.2.2 Interaction avec l'utilisateur	157
8.2.3 Apprentissage et paramétrage automatique	157
8.2.4 Dégradés de couleurs et textures	158
8.3 Applications industrielles futures possibles	158

8.1 Conclusion

Aujourd'hui, les industriels sont de plus en plus intéressés par l'exploitation de documents administratifs couleur. Les travaux de recherche portant sur ce type d'images sont encore très récents, et peu nombreux. Mais l'on voit peu à peu apparaître des solutions permettant de les traiter. Nombre de ces approches cherchent à conserver une séparation fond/premier-plan tout en rendant les processus de binarisation plus intelligents, et capables de tenir compte, dans une certaine mesure, des propriétés colorimétriques des images. D'autres sont dédiées à des tâches particulières, ou à des ensembles de documents partageant des caractéristiques similaires.

A travers ce mémoire, nous avons identifié les problèmes liés aux processus classiques de dématérialisation. Ceux-ci simplifient les images, en prenant ainsi le risque de perdre une partie des données. Cette perte peut être préjudiciable et en particulier lorsque des éléments de différentes couleurs se chevauchent. Par conséquent, nous avons proposé d'utiliser toute l'information colorimétrique disponible pour améliorer la performance de la segmentation. L'enjeu était de pouvoir traiter de manière automatique, et non supervisée, des documents très hétérogènes, et extrêmement riches en termes de contenus.

Nos contributions ont porté sur la segmentation des documents administratifs en couches couleur principales : une couche couleur est définie comme un ensemble de pixels ayant à la fois une apparence colorimétrique très proche, et présentant une unité logique ou sémantique. Chaque couche couleur est fournie sous la forme d'une image binaire qui peut être réinjectée dans le processus classique. Le but était de

proposer une alternative à l'étape de binarisation, et non de remplacer tout la procédure de numérisation existante. Chacune de ses couches porte une partie des données colorimétriques qui sont présentes sur l'image, et dont il est possible de profiter. En termes de segmentation, la plus-value est certaine : séparation d'éléments de couleurs différentes se chevauchant, localisation, génération de régions plus précises, etc. Les deux problématiques auxquelles nous nous sommes confrontés sont l'identification de manière automatique du nombre de couleurs observables sur l'image, et l'étape de segmentation elle-même.

Au sujet du nombre de couleurs, une image de document est supposée n'en contenir que très peu. En réalité, la densité des informations disponibles (textures, dégradés de couleurs, bruit, etc) fait qu'il est très difficile de déterminer une valeur pertinente. Cela est d'autant plus vrai qu'il est impossible de prédire les surcharges pouvant apparaître sur l'image (surlignements, annotations, etc). Le bruit impacte fortement la qualité de la segmentation. L'implémentation d'une méthode robuste, et suffisamment générique pour pouvoir traiter n'importe quel document, a été un véritable challenge. Du fait des limites que nous nous sommes imposées, la méthode devait fonctionner sans aucune connaissance *a priori* pour aider à la reconnaissance et donc à la segmentation. Lors de ces travaux, nous avons proposé et évalué trois catégories d'approches : colorimétriques, spatio-colorimétriques, et basée sur une analyse multi-résolution.

Les approches de segmentation qui se basent uniquement sur des caractéristiques colorimétriques fournissent une vision grossière des couleurs présentes sur les documents. Elles sont relativement efficaces lorsque les couleurs sont bien contrastées ; c'est à dire lorsqu'elles sont bien séparées dans l'espace couleur choisi. Cependant, ce cas de figure n'apparaît que rarement sur les images de documents couleurs. La manipulation, la numérisation, et la compression introduisent en effet des dégradations au niveau colorimétrique. De plus, la présence de texture, de dégradés de couleurs, et de plusieurs nuances d'une même teinte perturbent l'identification des nuages de points qui sont en théorie associés aux couleurs dominantes. Dans la réalité, ces points apparaîtront plutôt sous forme de traînées dans l'espace couleur choisi. La plupart du temps, la segmentation n'est donc pas suffisamment précise, et elle est très sensible au bruit.

Afin d'améliorer la qualité de la segmentation, nous avons ensuite envisagé la combinaison d'informations provenant de plusieurs sources à travers une approche spatio-colorimétriques. En effet, un pixel partage des caractéristiques similaires avec ses voisins. L'introduction de données spatiales dans le processus permet la génération de régions qui sont plus proches de celles réellement observées sur l'image. Nous nous sommes tournés vers l'algorithme de segmentation SLIC ([Achanta 2012]) qui génère un ensemble de régions locales perceptuellement cohérentes appelées super-pixels. Cette approche est en effet réputée pour sa capacité à produire des régions compactes et qui adhèrent bien aux contours de l'image. Ce résultat est directement

lié à l'utilisation d'une approche de partitionnement de type K-moyenne dans une fenêtre de recherche de petite taille, ce qui permet d'utiliser à la fois des données couleurs et spatiales.

Notre contribution a porté sur l'adaptation du SLIC à la structure particulière des images de documents. En effet, le SLIC original requiert le paramétrage du nombre initial de superpixels. Nous avons donc proposé une approche qui initialise de manière automatique cette valeur en fonction du contenu des documents. Ces images contiennent de grandes régions homogènes qui peuvent contenir de grands superpixels sans perdre d'informations. Au contraire, les zones pour lesquelles le gradient est fort (zone de premier-plan) nécessitent une segmentation beaucoup plus précise, et par conséquent plus de superpixels, et plus petits. Après avoir séparé les éléments se trouvant probablement au premier-plan, de ceux du fond, des superpixels de différentes tailles sont initialisés et placés sur l'image.

Le challenge a alors été le regroupement des régions locales en couches couleur globales. Cette étape implique la définition de règles de fusion, et de critères d'arrêt qui peuvent être délicats à mettre en place. Une image en haute résolution contient énormément de détails, ce qui complexifie les traitements. Nous avons alors proposé une analyse multi-résolution pour percevoir les éléments du document à différents niveaux de détails. Les couleurs principales sont ainsi extraites en basse résolution où l'on observe uniquement la structure globale du document. Les résolutions plus élevées sont ensuite utilisées pour affiner les résultats de la segmentation. Nous avons noté qu'une analyse multi-résolution était pertinente dans ce contexte car elle permet d'imiter le comportement visuel humain.

Le tableau 8.1 résume l'ensemble des mesures de performances des approches présentées dans ce mémoire. Au vu des résultats présentés, nous recommandons la combinaison de différentes sources d'informations pour obtenir une segmentation précise des images de documents administratifs. Une bonne segmentation nécessite au minimum des données spatiales et colorimétriques. De plus, un processus manipulant à la fois des données locales et globales est à préférer à une approche uniquement globale. En effet, cela le rend moins sensible aux variations locales. L'approche multi-résolution proposée dans ce mémoire répond aux limites que nous nous sommes imposées qui étaient le traitement de documents génériques, de manière automatique, et non supervisée. Une approche possible pour l'améliorer serait de piloter la segmentation à partir des structures logiques attendues. Cela impliquerait cependant une étape de reconnaissance qui serait appliquée en parallèle de la segmentation. Cette étape devrait être guidée par des connaissances *a priori* décrivant les éléments recherchés.

Les perspectives scientifiques de ces travaux vont maintenant être discutées dans la section 8.2. Nous listerons dans la section 8.3 quelques applications qui pourraient bénéficier des informations supplémentaires fournies par cette segmentation.

	Précision	Rappel
Nommage de couleurs	0.70	0.90
Classification hiérarchique		
— RVB	0.76	0.94
— HSV	0.67	0.93
— L*a*b	0.55	0.98
SLIC	0.84	0.85
SLIC Adaptatif (méthode de fusion basée sur la représentativité des couleurs)	0.85	0.93
SLIC Adaptatif (méthode de fusion basée sur l'étude du voisinage)	0.92	0.94
Analyse multi-résolution	0.85	0.97

TABLEAU 8.1 – Tableau récapitulatif des performances des approches évaluées lors de ces travaux de recherche.

8.2 Perspectives scientifiques

Ces travaux ont ouvert la voie à de nouvelles pistes d'exploration :

- affectation des pixels à plusieurs couches ;
- interaction avec l'utilisateur ;
- apprentissage et paramétrage automatique ;
- dégradés de couleurs et textures.

8.2.1 Affectation des pixels à plusieurs couches

Actuellement, chaque pixel de l'image ne peut appartenir qu'à une seule couche couleur. Or en cas de superpositions, les points localisés dans la zone de chevauchement devraient appartenir à plusieurs éléments logiques. Ces régions présentent la plupart du temps des variations colorimétriques importantes dues aux mélanges des encres utilisées. La teinte qui en résulte est parfois très éloignée des couleurs originales. Les pixels qui sont associés à cette zone sont donc susceptibles d'être identifiés comme une couche couleur à part entière. La problématique réside dans la détection de ces régions qu'il est très difficile de séparer du bruit sans connaissances *a priori*.

Il serait donc intéressant de pouvoir affecter un pixel à plusieurs couleurs. Cela impliquerait d'être capable de reconnaître des régions logiques de haut-niveaux telles que des mots, des paragraphes, des graphiques, etc. Tout l'enjeu n'est plus seulement d'identifier les couleurs dominantes, mais aussi de détecter des éléments homogènes de par leurs caractéristiques.

En cas de doute, un pixel pourrait apparaître sur plusieurs couches. L'objectif étant de favoriser la redondance, plutôt que la perte d'information. L'idéal serait cependant d'intégrer une approche basée sur la logique floue qui affecterait ce pixel à une couleur donnée avec un degré d'appartenance. Par la suite, la prise de décision serait faite en combinant des informations provenant de différentes sources qui ne seraient plus uniquement colorimétriques ou spatiales : contexte, texture, homogénéité des caractéristiques, relations topologiques, etc.

8.2.2 Interaction avec l'utilisateur

La définition d'une couche couleur pertinente peut poser problèmes car elle est dépendante de la qualité de la segmentation désirée, et de l'application. Segmenter des documents hétérogènes de manière précise, dans le but d'en extraire un contenu très riche, est très coûteux en temps de calculs. Dans un contexte industriel, cela peut limiter l'utilisation de ce type d'approche, et ce d'autant plus qu'elle a pour objectif d'être appliquée comme pré-traitement. Dans la mesure où cela serait possible, il pourrait être intéressant d'autoriser une interaction minimale avec un utilisateur. Dans le cas de traitements par lots, cette interaction serait peu importante, car faite une seule fois, et augmenterait grandement les performances des processus.

Plusieurs paramètres peuvent être fixés. Au niveau des couleurs plus spécifiquement, un utilisateur pourrait par exemple définir celles qui sont le plus contrastées sur le document. Les zones de transition associées au bruit seraient ainsi plus facilement détectées et supprimées. Dans la mesure du possible, et si l'application le permet, l'idéal serait de profiter au maximum des connaissances *a priori*. Si l'objectif est l'extraction de régions spécifiques pour lesquelles la couleur, ou la localisation, est connue, il pourrait être intéressant de limiter le processus de segmentation à ces zones d'intérêt.

8.2.3 Apprentissage et paramétrage automatique

Les documents d'une même famille partagent des caractéristiques communes. Le système pourrait profiter de cette redondance. Lorsqu'un document est rencontré pour la première fois, l'idée serait de le segmenter une première fois grâce à notre approche. Puis les paramètres qui ont conduits aux résultats seraient sauvegardés. Si le système est capable d'identifier un document similaire, l'intérêt serait d'utiliser ces paramètres pour accélérer les traitements.

Cela pourrait passer notamment par la construction d'un modèle qui injecterait des connaissances *a priori* permettant de raffiner la segmentation. Celles-ci pourraient être utilisées pour spécialiser le processus pour un type de document, et une tâche donnée, tout en conservant le caractère générique de la segmentation. Une autre application serait d'adapter de manière automatique les paramètres en fonction de la qualité des résultats. Cela passerait par une étape de validation avec un utilisateur.

8.2.4 Dégradés de couleurs et textures

Les dégradés de couleurs sont relativement fréquents sur les images de documents administratifs. On en rencontre par exemple souvent dans les colonnes de tableaux. Actuellement, nous ne les prenons pas en compte. Cependant, il serait plus pertinent de les considérer comme une classe de couleur à part. Une des perspectives futures est donc leur détection, et leur intégration dans la segmentation. Cette approche pourrait être ensuite étendue à la gestion des textures. Une idée serait d'intégrer des caractéristiques géométriques supplémentaires. En effet, la plupart du temps, les dégradés forment des régions globalement de forme rectangulaire, variant peu localement d'un point de vue colorimétrique, et dont la distance couleur entre les deux extrema est élevée

8.3 Applications industrielles futures possibles

Notre approche est une segmentation bas-niveau qui est appliquée juste après la numérisation, et avant la phase de reconnaissance et de lecture du texte. Nous fournissons en sortie un ensemble d'images binaires : une pour chaque couche couleur. Ces images peuvent être ensuite réinjectées dans la chaîne de traitement classique. Mais elles possèdent des informations colorimétriques supplémentaires qui pourraient être utilisées pour améliorer les processus traditionnels, ou encore pour proposer de nouveaux services.

Nous avons ainsi identifiés plusieurs applications futures possibles qui sont résumées ci-dessous :

- description et indexation ;
- traitement d'une couche particulière ;
- amélioration de la segmentation ;
- localisation d'information.

Description et indexation : La segmentation en couches couleur permet d'identifier un nombre réduit de couleurs principales sur les images de documents. Ces couleurs peuvent être utilisées comme des descripteurs pour des tâches de reconnaissance et d'indexation. L'intérêt est que les documents sont généralement très codifiés. Malgré des variations, les documents d'une même famille contiennent de grandes régions de couleurs similaires localisées globalement aux mêmes endroits.

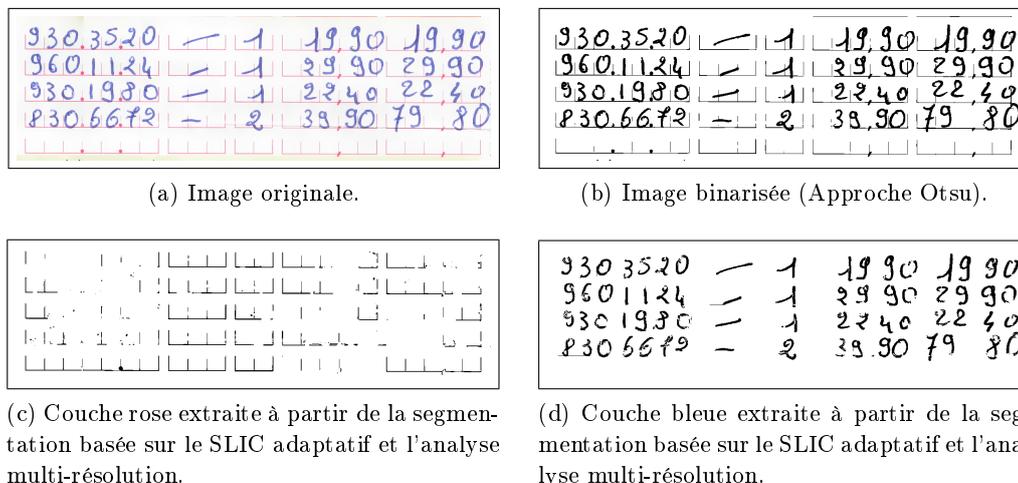


FIGURE 8.1 – Les éléments de couleurs différentes apparaissent sur différentes couches couleur contrairement à une binarisation traditionnelle.

Traitement d'une couche particulière : Une fois les couches couleur obtenues, il est possible de ne traiter que celles qui sont pertinentes pour un objectif donné. Plusieurs applications peuvent en découler : binarisation intelligente, compression, segmentation, etc.

Amélioration de la segmentation : Une binarisation traditionnelle ne permet pas de séparer des éléments de couleurs différentes qui se chevauchent (Figure 8.1b). Cela peut être source d'erreurs de segmentation et de reconnaissance. Au contraire, dans une représentation sous forme de couches couleur, ces éléments apparaissent sur des couches différentes (Figure 8.1), qui peuvent être traitées de manière indépendante. La figure 8.1d montre le résultat de la segmentation de la couche associée à l'écriture manuscrite, qui peut être transmise à un outil de reconnaissance de texte sans les peignes qui se trouvent sur une autre couche (Figure 8.1c).

Localisation d'information : Certaines applications d'analyse de documents doivent localiser des données spécifiques sur un document. Sur ce type d'images, la couleur est généralement porteuse de sens. Sur la figure 8.2a, les régions contenant des montants sont identifiables par les labels qui leur sont associées, et qui ont une couleur bien spécifique. Il arrive souvent que des éléments soient mis en évidence en les soulignant (Figure 8.2). Suivant l'application désirée, une segmentation en couches couleur pourraient aider à extraire ces données, permettant une localisation des régions d'intérêt (Figures 8.2b et 8.2d).



FIGURE 8.2 – La couleur peut fournir des informations permettant de localiser des éléments spécifiques du document.

Bibliographie

- [Achanta 2012] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua et Sabine Süsstrunk. *SLIC superpixels compared to state-of-the-art superpixel methods*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, pages 2274–2281, 2012. (Cité en pages ix, x, 51, 52, 118, 119, 121, 122, 129, 131 et 154.)
- [Agrawal 2009] Mudit Agrawal et David Doermann. *Voronoi++ : A dynamic page segmentation approach based on voronoi and Docstrum features*. In Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, pages 1011–1015, 2009. (Cité en page 25.)
- [Antonacopoulos 2007] Apostolos Antonacopoulos, Basilios Gatos et David Bridson. *Page Segmentation Competition*. In 9th International Conference on Document Analysis and Recognition {ICDAR} 2007, 23-26 September, Curitiba, Paran{á}, Brazil, pages 1279–1283, 2007. (Cité en page 82.)
- [Antonacopoulos 2009a] A Antonacopoulos, D Bridson, C Papadopoulos et S Pletschacher. *A Realistic Dataset for Performance Evaluation of Document Layout Analysis*. In Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on, pages 296–300, 2009. (Cité en pages x, 75 et 76.)
- [Antonacopoulos 2009b] A Antonacopoulos, S Pletschacher, D Bridson et C Papadopoulos. *ICDAR 2009 Page Segmentation Competition*. In Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on, pages 1370–1374, 2009. (Cité en pages 76 et 82.)
- [Badekas 2006] Efthimios Badekas, Nikos Nikolaou et Nikos Papamarkos. *Text binarization in color documents*. International Journal of Imaging Systems and Technology, vol. 16, no. 6, pages 262–274, 2006. (Cité en page 63.)
- [Baird 2014] Henry S. Baird et Karl Tombre. *The Evolution of Document Image Analysis*. In Karl Doermann, David and Tombre, editeur, Handbook of Document Image Processing and Recognition, pages 63–71. Springer London, 2014. (Cité en page 16.)
- [Barlas 2014] P Barlas, C Chatelain, S Adam et T Paquet. *Détection et segmentation des blocs de texte manuscrits et imprimés dans des documents complexes*. In CIFED, Nancy, France, 2014. (Cité en pages ix, 30 et 31.)
- [Benavente 2002] Robert Benavente, Francesc Tous, Ramon Baldrich et Maria Vanrell. *Statistical Modelling of a Colour Naming Space*. Conference on Colour in Graphics, Imaging, and Vision, no. 1, pages 406–411, 2002. (Cité en pages 54 et 108.)
- [Benavente 2006] Robert Benavente, Maria Vanrell et Ramon Baldrich. *A data set for fuzzy colour naming*. Color Research and Application, vol. 31, pages 48–56, 2006. (Cité en pages ix, 54 et 55.)

- [Benavente 2008] Robert Benavente, Maria Vanrell et Ramon Baldrich. *Parametric fuzzy sets for automatic color naming*. J. Opt. Soc. Am. A, vol. 25, no. 10, pages 2582–2593, Octobre 2008. (Cité en pages 99 et 100.)
- [Berlin 1969] Brent Berlin et Paul Kay. Basic Color Terms : Their Universality and Evolution, volume 19. 1969. (Cité en pages 54, 99 et 105.)
- [Bottou 1998] Léon Bottou, Patrick Haffner, Paul G Howard, Patrice Simard, Yoshua Bengio et Yann Le Cun. *High Quality Document Image Compression with {DjVu}*. Journal of Electronic Imaging, vol. 7, no. 3, pages 410–425, 1998. (Cité en pages 60 et 65.)
- [Brun 2002] Luc Brun et Alain Trémeau. Digital Color Imaging Handbook, chapitre 9 : Color, pages 589–637. Electrical and Applied Signal Processing. CRC Press, 2002. (Cité en page 52.)
- [Busin 2008] Laurent Busin, Nicolas Vandenbroucke et Ludovic Macaire. *Color Spaces and Image Segmentation*, 2008. (Cité en pages 38 et 47.)
- [Canny 1986] J Canny. *A computational approach to edge detection*. IEEE transactions on pattern analysis and machine intelligence, vol. 8, pages 679–698, 1986. (Cité en page 123.)
- [Chen 2013] Kai Chen, Fei Yin et Cheng-Lin Liu. *Hybrid Page Segmentation with Efficient Whitespace Rectangles Extraction and Grouping*. In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, pages 958–962, 2013. (Cité en pages 26 et 27.)
- [Cohen 2013] Rafi Cohen, Abedelkadir Asi, Klara Kedem, Jihad El-Sana et Itshak Dinstein. *Robust Text and Drawing Segmentation Algorithm for Historical Documents*. In Proceedings of the 2Nd International Workshop on Historical Document Imaging and Processing, HIP '13, pages 110–117, New York, NY, USA, 2013. ACM. (Cité en page 65.)
- [Comaniciu 2002] D. Comaniciu et P. Meer. *Mean shift : a robust approach toward feature space analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, 2002. (Cité en page 49.)
- [da Silva Barboza 2011] Ricardo da Silva Barboza, Rafael Dueire Lins et Victor Matheus de S. Pereira. *Using Readers' Highlighting on Monochromatic Documents for Automatic Text Transcription and Summarization*. 2013 12th International Conference on Document Analysis and Recognition, vol. 0, pages 212–216, 2011. (Cité en pages 66 et 67.)
- [Dengel 2014] Andreas Dengel et Faisal Shafait. *Analysis of the Logical Layout of Documents*. In Karl Doermann, David and Tombre, éditeur, Handbook of Document Image Processing and Recognition, pages 177–222. Springer London, 2014. (Cité en pages ix, 16 et 31.)
- [Di Zenzo 1986] Silvano Di Zenzo. *A note on the gradient of a multi-image*, 1986. (Cité en pages 88 et 123.)

- [Dombre 2003] Julien Dombre. *Multi-scale representation systems for indexing and restoring color medieval archives*. Theses, Université de Poitiers, 2003. (Cité en pages ix, 20, 28 et 29.)
- [EY 2014] EY et DIMO Gestion. *Livre blanc Dématérialisation des factures fournisseurs*. <http://www.ey.com/FR/fr/Services/Advisory/livre-blanc-EY-dematerialisation-des-factures-fournisseurs>, 2014. (Cité en page 2.)
- [Fairchild 2005] M D Fairchild. *Color Appearance Models*. John Wiley & Sons Ltd, pages 1–409, 2005. (Cité en page 38.)
- [Fernandez-Maloigne 2013] Christine Fernandez-Maloigne. *Advanced color image processing and analysis*. Springer, New York, 2013. (Cité en pages 38 et 48.)
- [Fernández-Mota 2014] David Fernández-Mota, Josep Lladós et Alicia Fornés. *A graph-based approach for segmenting touching lines in historical handwritten documents*. *International Journal on Document Analysis and Recognition (IJ DAR)*, vol. 17, no. 3, pages 293–312, 2014. (Cité en page 23.)
- [Fu 1981] K.S. Fu et J.K. Mui. *A survey on image segmentation*, 1981. (Cité en page 46.)
- [Fukunaga 1975] K. Fukunaga et L. Hostetler. *The estimation of the gradient of a density function, with applications in pattern recognition*. *IEEE Transactions on Information Theory*, vol. 21, no. 1, 1975. (Cité en page 49.)
- [Garain 2005] Utpal Garain, Thierry Paquet et Laurent Heutte. *On foreground-background separation in low quality color document images*. In *Proceedings of the International Conference on Document Analysis and Recognition, IC-DAR*, volume 2005, pages 585–589, 2005. (Cité en page 60.)
- [Gervautz 1990] Michael Gervautz et Werner Purgathofer. *Graphics Gems*. chapitre A Simple M, pages 287–293. Academic Press Professional, Inc., San Diego, CA, USA, 1990. (Cité en page 53.)
- [Hammami 2014] Maroua Hammami, Pierre Héroux et Sébastien Adam. *Extraction de zones informatives dans des images de formulaire en couleur*. In {CO-RIA} 2014 - Conf{é}rence en Recherche d’Infomations et Applications- 11th French Information Retrieval Conference. {CIFED} 2014 Colloque International Francophone sur l’Ecrit et le Document, Nancy, France, March 19-23, 2014., pages 171–184, 2014. (Cité en pages x, 67 et 68.)
- [Heckbert 1982] Paul Heckbert. *Color Image Quantization for Frame Buffer Display*. *SIGGRAPH Comput. Graph.*, vol. 16, no. 3, pages 297–307, 1982. (Cité en page 53.)
- [Heer 2012] Jeffrey Heer et Maureen Stone. *Color naming models for color selection, image editing and palette design*. *CHI’12 (Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems)*, page 1007, 2012. (Cité en page 54.)
- [Helmholtz 1866] Hermann Von Helmholtz. *Handbuch der physiologischen Optik*, volume 9. 1866. (Cité en page 40.)

- [Hering 1875] Ewald Hering. *Zur Lehre vom Lichtsinn*. Wien. Math. Nat. Kl., vol. 70, page 169, 1875. (Cit  en page 41.)
- [Jain 2010] Anil K. Jain. *Data clustering : 50 years beyond K-means*. Pattern Recognition Letters, vol. 31, pages 651–666, 2010. (Cit  en pages 48 et 101.)
- [Kanungo 1993] T. Kanungo, R.M. Haralick et I. Phillips. *Global and local document degradation models*. Proceedings of 2nd International Conference on Document Analysis and Recognition (ICDAR '93), 1993. (Cit  en pages 76, 80, 84, 85 et 87.)
- [Kessi 2015] Louisa Kessi, Lebourgeois Franck, Christophe Garcia et Jean Duong. *AColDPS :Robust and Unsupervised Automatic Color Document Processing System*. In International Conference on Computer Vision Theory and Applications, VISAPP 2015, International Conference on Computer Vision Theory and Applications (VISAPP 2015), pages pp.174–185, Berlin, Germany, 2015. (Cit  en pages 62, 63 et 64.)
- [Kieu 2012] Van Cuong Kieu, Muriel Visani, Nicholas Journet, Jean-Philippe Domenger et R my Mullot. *A character degradation model for grayscale ancient document images*. In Proceedings of the 21st International Conference on Pattern Recognition, {ICPR} 2012, Tsukuba, Japan, November 11-15, 2012, pages 685–688, 2012. (Cit  en pages x, 76, 80, 84, 85, 87 et 90.)
- [Kieu 2013] V C Kieu, Muriel Visani, Nicholas Journet, R my Mullot et Jean Philippe Domenger. *An Efficient Parametrization of Character Degradation Model for Semi-synthetic Image Generation*. In Proceedings of the 2Nd International Workshop on Historical Document Imaging and Processing, HIP '13, pages 29–35, New York, NY, USA, 2013. ACM. (Cit  en pages 84, 85, 87, 88 et 89.)
- [Kise 1998] Koichi Kise, Akinori Sato et Motoi Iwata. *Segmentation of Page Images Using the Area Voronoi Diagram*. Computer Vision and Image Understanding, vol. 70, no. 3, pages 370–382, 1998. (Cit  en pages 25 et 26.)
- [Kise 2014] Koichi Kise. *Page Segmentation Techniques in Document Analysis*. In Karl Doermann, David and Tombre, editeur, Handbook of Document Image Processing and Recognition, pages 135–175. Springer London, 2014. (Cit  en pages ix, 16, 20, 21 et 22.)
- [Lebourgeois 2013] Frank Lebourgeois, Fadoua Drira, Djamel Gaceb et Jean Duong. *Fast Integral MeanShift : Application to Color Segmentation of Document Images*. In IEEE, editeur, Twelfth International Conference on Document Analysis and Recognition (ICDAR 2013), pages 52–56, 2013. (Cit  en pages 53, 62 et 63.)
- [Lemaitre 2008a] A Lemaitre, Jean Camillerapp et B Couasnon. *Approche perceptifve pour la reconnaissance de filets bruit s, Application   la structuration de pages de journaux*. In Antoine Tabbone et Thierry Paquet, editeur, Colloque International Francophone sur l'Ecrite et le Document, pages 61–66, France, 2008. Groupe de Recherche en Communication Ecrite. (Cit  en page 29.)

- [Lemaitre 2008b] Aurélie Lemaitre, Jean Camillerapp et Bertrand Coüasnon. *Multiresolution cooperation makes easier document structure recognition*. International Journal of Document Analysis and Recognition (IJ DAR), vol. 11, no. 2, pages 97–109, 2008. (Cité en page 29.)
- [Lindner 2012] A Lindner, N Bonnier et S Süssstrunk. *What is the color of chocolate ? Extracting color values of semantic expressions*. In Proc. IS&T 6th European Conference on Colour in Graphics, Imaging, and Vision (CGIV 2012), pages 355 – 361, 2012. (Cité en page 56.)
- [Loo 2004] Poh Kok Loo et Chew Lim Tan. *Adaptive Region Growing Color Segmentation for Text Using Irregular Pyramid*. In Simone Marinai et Andreas Dengel, éditeurs, Document Analysis Systems, volume 3163 of *Lecture Notes in Computer Science*, pages 264–275. Springer, 2004. (Cité en page 63.)
- [Luyen 2015] Do Thi Luyen, Elodie Carel, Jean-Marc Ogier et Jean-Christophe Burie. *A Character Degradation Model for Color Document*. In Proceedings of 13th International Conference on Document Analysis and Recognition (ICDAR '2015), page Accepted on June 2015, 2015. (Cité en page 84.)
- [MacAdam 1942] David L. MacAdam. *Visual Sensitivities to Color Differences in Daylight*, 1942. (Cité en page 40.)
- [Macaire 2006] Ludovic Macaire, Nicolas Vandenbroucke et J. G. Postaire. *Color image segmentation by analysis of subset connectedness and color homogeneity properties*. Computer Vision and Image Understanding, vol. 102, pages 105–116, 2006. (Cité en pages 103 et 139.)
- [Maxwell 1860] J C Maxwell. *On the theory of compound colours and the relations of the colours of the spectrum*. Philosophical Transactions of the Royal Society of London, no. 150, pages 57–84, 1860. (Cité en page 40.)
- [Montagne 2005] C Montagne. *Apport de la couleur en vision - Application à la localisation locale et globale d'une base mobile d'aide aux handicapés*. PhD thesis, Univ. Evry, 2005. (Cité en pages ix et 53.)
- [Moroney 2011] N M Moroney et G B Beretta. *Validating large-scale lexical color resources*. In Midterm Meeting of the International Colour Association (AIC)(June 2011), 2011. (Cité en page 55.)
- [Munroe 2010] R. Munroe. *Color Survey Results*. <http://blog.xkcd.com/2010/05/03/color-survey-results/>, 2010. (Cité en pages ix, 54, 55 et 56.)
- [Nagy 1984] George Nagy et Sharad Seth. *Hierarchical representation of optically scanned documents*. Proceedings of International Conference on Pattern Recognition, vol. 1, pages 347–349, 1984. (Cité en page 21.)
- [Nartker 2005] Thomas A Nartker, Stephen V Rice et Steven E Lumos. *Software tools and test data for research and testing of page-reading OCR systems*. In In International Symposium on Electronic Imaging Science and Technology, 2005. (Cité en page 75.)

- [Nikolaou 2009] Nikos Nikolaou et Nikos Papamarkos. *Color reduction for complex document images*. International Journal of Imaging Systems and Technology, vol. 19, no. 1, pages 14–26, 2009. (Cit  en pages 53, 62 et 65.)
- [O’Gorman 1993] Lawrence O’Gorman. *Document spectrum for page layout analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 11, pages 1162–1173, 1993. (Cit  en page 25.)
- [Ouji 2012] Asma Ouji. *Segmentation et classification dans les images de documents num ris s*. Th se de doctorat en informatique, INSA de Lyon, 2012. (Cit  en pages x, 58, 59, 62, 64, 65 et 82.)
- [P rraga 2009] C A P rraga, R Benavente, M Vanrell et R Baldrich. *Psychophysical measurements to model intercolor regions of color-naming space*. Journal of Imaging Science and Technology, vol. 53, page 31106, 2009. (Cit  en pages 99, 100 et 105.)
- [Philips 1993] I T Philips, S Chen, J Ha et R M Haralick. *English Document Database Design and Implementation Methodology*. In In Proceeding of the 2nd Annual Symposium on Document Analysis and Retrieval (UNLV, pages 65–104, 1993. (Cit  en page 75.)
- [Ray 1999] Siddheswar Ray et R.H. Turi. *Determination of number of clusters in k-means clustering and application in colour image segmentation*. In Proceedings of the 4th international conference on advances in pattern recognition and digital techniques, pages 137–143, 1999. (Cit  en page 103.)
- [Rehrmann 1998] Volker Rehrmann et Lutz Prieese. *Fast and Robust Segmentation of Natural Color Scenes*. In Roland T Chin et Ting-Chuen Pong, editeurs, ACCV (1), volume 1351 of *Lecture Notes in Computer Science*, pages 598–606. Springer, 1998. (Cit  en pages ix, 50 et 51.)
- [Sauvola 1998] H. Sauvola J. Kauniskangas. *Mediateam document database*. <http://www.mediateam oulu.fi/downloads/MTDB/>, 1998. (Cit  en page 75.)
- [Sayre 1973] Kenneth M. Sayre. *Machine recognition of handwritten words : A project report*, 1973. (Cit  en page 7.)
- [Shafait 2006] Faisal Shafait, Daniel Keysers et Thomas M. Breuel. *Performance comparison of six algorithms for page segmentation*. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 3872 LNCS, pages 368–379, 2006. (Cit  en page 30.)
- [Smigiel 2004] E Smigiel, A Belaid et H Hamza. *Self-organizing maps and ancient documents*. In Document Analysis Systems VI, num ro 3163 de Lecture Notes in Computer Science ({LNCS}), pages 125–134. Springer, 2004. (Cit  en page 66.)
- [The National Library of Medicine 2010] The National Library of Medicine. *Medical article records groundtruth, Bethesda, USA*. <http://marg.nlm.nih.gov/>, 2010. (Cit  en page 75.)

- [Todoran 2005] Leon Todoran, Marcel Worring et M Smeulders. *The UvA Color Document Dataset*. Int. J. Doc. Anal. Recognit., vol. 7, no. 4, pages 228–240, 2005. (Cit  en page 75.)
- [Tomasi 1998] C. Tomasi et R. Manduchi. *Bilateral filtering for gray and color images*. Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), 1998. (Cit  en pages 98, 118 et 135.)
- [van de Weijer 2009] Joost van de Weijer, Cordelia Schmid, Jakob Verbeek et Diane Larlus. *Learning color names for real-world applications*. IEEE Transactions on Image Processing, vol. 18, pages 1512–1523, 2009. (Cit  en page 55.)
- [Vandenbroucke 2000] N Vandenbroucke. *Segmentation d’images couleur par classification de pixels dans des espaces d’attributs colorim triques adapt s. Application   l’analyse d’images de football*. PhD thesis, Universit  de Lille 1, 2000. (Cit  en pages ix, 38, 43 et 45.)
- [Vantaram 2012] Sreenath Rao Vantaram et Eli Saber. *Survey of contemporary trends in color image segmentation*. Journal of Electronic Imaging, vol. 21, pages 040901–1, 2012. (Cit  en pages x, 46, 169 et 170.)
- [Wong 1982] K Y Wong, R G Casey et F M Wahl. *Document Analysis System*. IBM Journal of Research and Development, vol. 26, pages 647–656, 1982. (Cit  en pages 23 et 123.)
- [Wright 1941] W D Wright. *The sensitivity of the eye to small colour differences*. Proceedings of the Physical Society, vol. 53, no. 2, page 93, 1941. (Cit  en page 40.)
- [Ye 2013] Peng Ye et David Doermann. *Document Image Quality Assessment : A Brief Survey*. In Intl. Conf. on Document Analysis and Recognition ({ICDAR}), pages 723–727, 2013. (Cit  en page 58.)
- [Yin 2001] P. Y. Yin. *Skew detection and block classification of printed documents*. Image and Vision Computing, vol. 19, no. 8, pages 567–579, 2001. (Cit  en pages 23 et 24.)
- [Yin 2009] Fei Yin et Cheng Lin Liu. *Handwritten Chinese text line segmentation by clustering with distance metric learning*. Pattern Recognition, vol. 42, no. 12, pages 3146–3157, 2009. (Cit  en page 23.)
- [Young 1807] T. Young. A course of lectures on natural philosophy, volume 2. Johnson, London, 1807. (Cit  en page 40.)
- [Zagoris 2007] Konstantinos Zagoris, Nikos Papamarkos et Ioannis Koustoudis. *Color Reduction Using the Combination of the Kohonen Self-Organized Feature Map and the Gustafson-Kessel Fuzzy Algorithm*. In Petra Perner, editeur, Machine Learning and Data Mining in Pattern Recognition, volume 4571 of *Lecture Notes in Computer Science*, pages 703–715. Springer Berlin Heidelberg, 2007. (Cit  en page 53.)

Techniques de segmentation bas-niveaux

La segmentation consiste à partitionner une image en régions exploitables pour un objectif donné.

La Figure A.1 présente la taxonomie des algorithmes de segmentation bas-niveaux proposée par [Vantaram 2012]. Ceux-ci ont identifié deux catégories d’approches manipulant des données provenant de sources différentes :

- Les approches colorimétriques ;
- Les approches spatio-colorimétriques.

Les approches colorimétriques sont basées sur l’hypothèse qu’une région colorimétriquement homogène dans l’image peut être associée à un sous-ensemble particulier de points ayant des caractéristiques proches dans l’espace couleur. Elles ne tiennent pas compte de l’organisation spatiale des points à travers l’image étudiée. Ces approches sont généralement très sensibles au bruit.

Malgré les dégradations introduites par la présence de bruit, un pixel partage un certain nombre de caractéristiques proches avec son voisinage. Les approches spatio-colorimétriques combinent à la fois des données couleurs et spatiales afin de rendre la segmentation plus robuste.

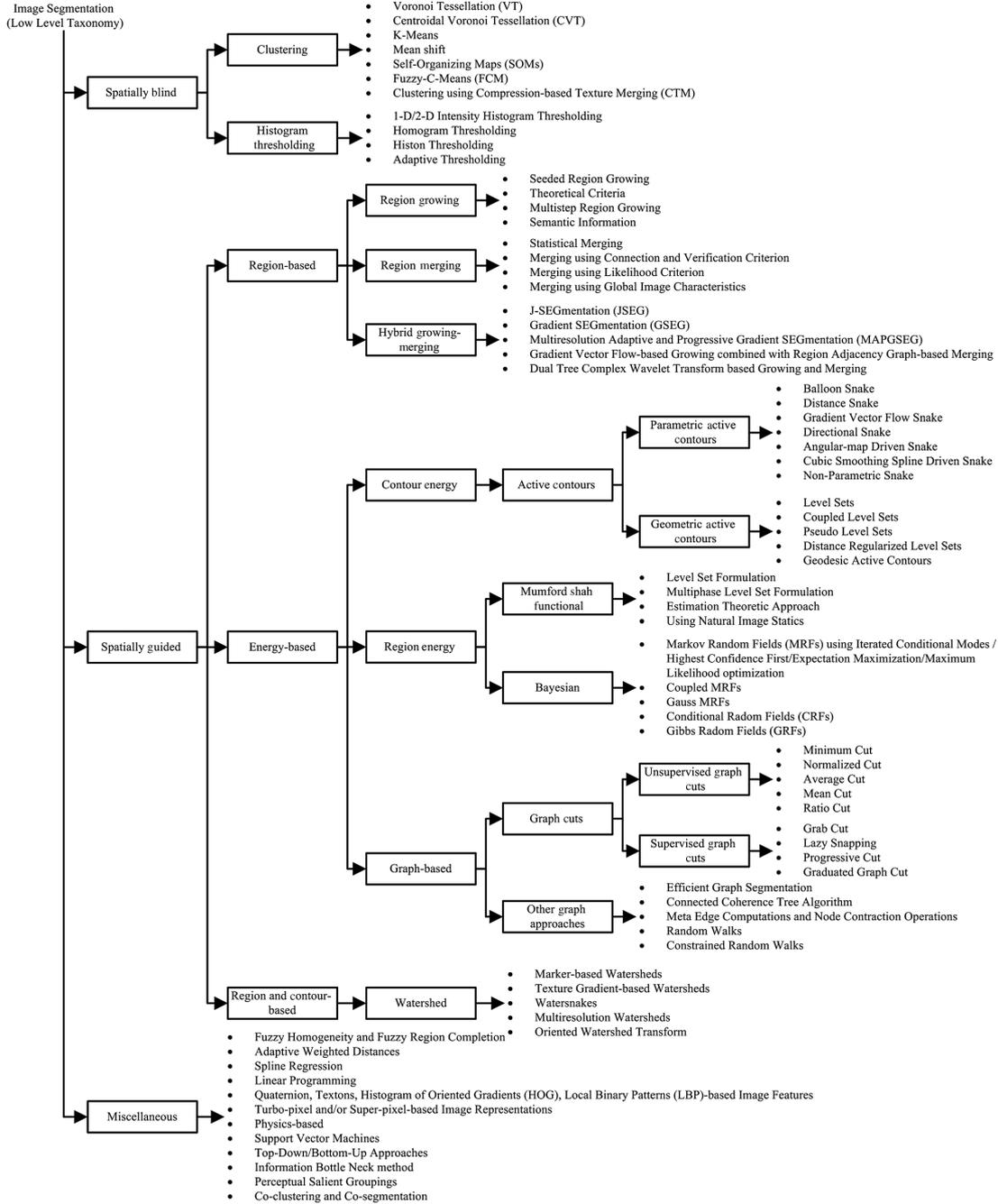


FIGURE A.1 – Taxonomie des algorithmes de segmentation bas-niveaux. Crédits : [Vantaram 2012]

Exemples de documents administratifs réels



(a)



(b)



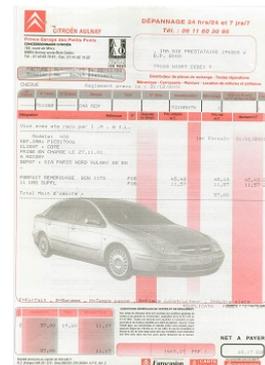
(c)



(d)

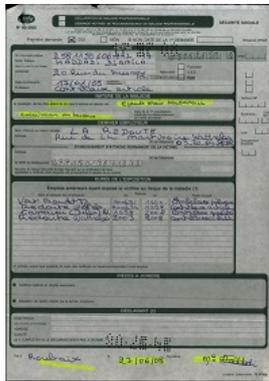


(e)



(f)

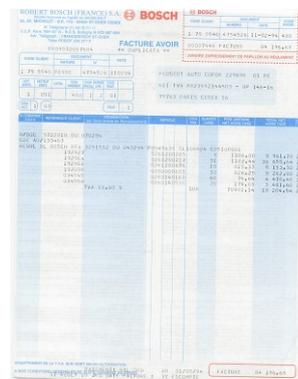
FIGURE B.1 – Ensemble de documents administratifs réels fournis par notre partenaire industriel, partie 1.



(a)



(b)



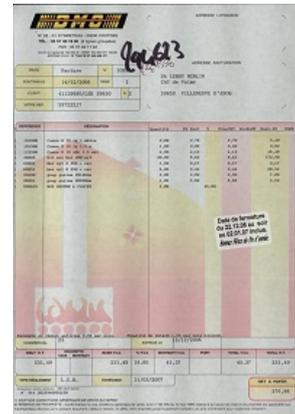
(c)



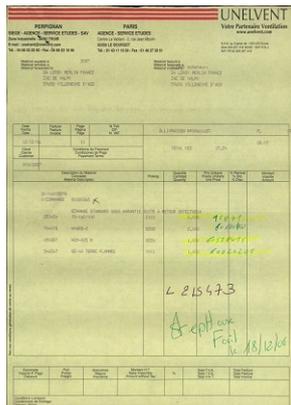
(d)



(e)



(f)



(g)



(h)



(i)

FIGURE B.2 – Ensemble de documents administratifs réels fournis par notre partenaire industriel, partie 2.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

FIGURE B.3 – Ensemble de documents administratifs réels fournis par notre partenaire industriel, partie 3.

Publications

Conférences

- **Multiresolution Approach Based on Adaptive Superpixels for Administrative Documents Segmentation into Color Layers**
Élodie Carel, Jean-Christophe Burie, Vincent Courboulay, Jean-Marc Ogier et Vincent Poulain d'Andecy
Submitted to the 13th International Conference on Document Analysis and Recognition (ICDAR), Accepted on June 2015

- **A Character Degradation Model for Color Document Images**
Do Thi Luyen, Élodie Carel, Jean-Marc Ogier et Jean-Christophe Burie
Submitted to the 13th International Conference on Document Analysis and Recognition (ICDAR), Accepted on June 2015

Workshops

- **Dominant Color Segmentation of Administrative Document Images by Hierarchical Clustering**
Élodie Carel, Vincent Courboulay, Jean-Christophe Burie, Jean-Marc Ogier
Proceedings of the 13th ACM Symposium on Document Engineering (DocEng), Sep 2013, Florence, Italy

Segmentation de documents administratifs en couches couleur.

Résumé : Les entreprises doivent traiter quotidiennement de gros volumes de documents papiers de toutes sortes. Automatisation, traçabilité, alimentation de systèmes d'informations, réduction des coûts et des délais de traitement, la dématérialisation a un impact économique évident. Pour respecter les contraintes industrielles, les processus historiques d'analyse simplifient les images grâce à une séparation fond/premier-plan. Cependant, cette binarisation peut être source d'erreurs lors des étapes de segmentation et de reconnaissance. Avec l'amélioration des techniques, la communauté d'analyse de documents a montré un intérêt croissant pour l'intégration d'informations colorimétriques dans les traitements, ceci afin d'améliorer leurs performances. Pour respecter le cadre imposé par notre partenaire privé, l'objectif était de mettre en place des processus non supervisés. Notre but est d'être capable d'analyser des documents même rencontrés pour la première fois quels que soient leurs contenus, leurs structures, et leurs caractéristiques en termes de couleurs. Les problématiques de ces travaux ont été d'une part l'identification d'un nombre raisonnable de couleurs principales sur une image ; et d'autre part, le regroupement en couches couleur cohérentes des pixels ayant à la fois une apparence colorimétrique très proche, et présentant une unité logique ou sémantique. Fournies sous forme d'un ensemble d'images binaires, ces couches peuvent être réinjectées dans la chaîne de dématérialisation en fournissant une alternative à l'étape de binarisation classique. Elles apportent en plus des informations complémentaires qui peuvent être exploitées dans un but de segmentation, de localisation, ou de description. Pour cela, nous avons proposé une segmentation spatio-colorimétrique qui permet d'obtenir un ensemble de régions locales perceptuellement cohérentes appelées superpixels, et dont la taille s'adapte au contenu spécifique des images de documents. Ces régions sont ensuite regroupées en couches couleur globales grâce à une analyse multi-résolution.

Mots clés : document administratif, couleur, quantification couleur, segmentation non-supervisée, couche couleur, superpixels, analyse multi-résolutions

Segmentation of administrative document images into color layers.

Abstract: Industrial companies receive huge volumes of documents everyday. Automation, traceability, feeding information systems, reducing costs and processing times, dematerialization has a clear economic impact. In order to respect the industrial constraints, the traditional digitization process simplifies the images by performing a background/foreground separation. However, this binarization can lead to some segmentation and recognition errors. With the improvements of technology, the community of document analysis has shown a growing interest in the integration of color information in the process to enhance its performance. In order to work within the scope provided by our industrial partner in the digitization flow, an unsupervised segmentation approach was chosen. Our goal is to be able to cope with document images, even when they are encountered for the first time, regardless their content, their structure, and their color properties. To this end, the first issue of this project was to identify a reasonable number of main colors which are observable on an image. Then, we aim to group pixels having both close color properties and a logical or semantic unit into consistent color layers. Thus, provided as a set of binary images, these layers may be reinjected into the digitization chain as an alternative to the conventional binarization. Moreover, they also provide extra-information about colors which could be exploited for segmentation purpose, elements spotting, or as a descriptor. Therefore, we have proposed a spatio-colorimetric approach which gives a set of local regions, known as superpixels, which are perceptually meaningful. Their size is adapted to the content of the document images. These regions are then merged into global color layers by means of a multiresolution analysis.

Keywords: administrative document, color, color quantization, unsupervised segmentation, color layer, superpixels, multiresolution analysis

Laboratoire Informatique, Image, Interaction
Pôle Sciences & Technologie
Avenue Michel Crépeau

17042 LA ROCHELLE CEDEX 1