Introduction de la vision perceptive pour la reconnaissance de la structure de documents

Aurélie Lemaitre Legargeant

N° d'ordre: D08-23

THÈSE

Présentée devant

l'Institut National des Sciences Appliquées de Rennes

pour obtenir le grade de :

Docteur de l'Institut National des Sciences Appliquées de Rennes
Mention Informatique

par

Aurélie Lemaitre-Legargeant

Équipe d'accueil : Imadoc - IRISA
École Doctorale : Matisse
Titre de la thèse :

Introduction de la vision perceptive pour la reconnaissance de la structure de documents

soutenue le 5 décembre 2008 devant la commission d'examen

MM. :	Rolf	Ingold	Rapporteurs
	Thierry	Paquet	
MM. :	Laurence	Likforman-Sulem	Examinateurs
	Josep	Lládos	
	Jean	Camillerapp	
	Bertrand	Coüasnon	

Remerciements

Tout d'abord, je tiens à remercier tous les membres de mon jury de thèse pour l'intérêt qu'ils ont porté à mon travail. Merci à Rolf Ingold et Thierry Paquet d'avoir accepté la charge de rapporteur et d'avoir réalisé une lecture attentive de ce maunscrit. Je remercie Josep Llados et Laurence Likforman d'avoir bien voulu participer à mon jury de thèse.

Je tiens à remercier tout particulièrement Jean Camillerapp et Bertrand Coüasnon de m'avoir proposé ce sujet de thèse et encadrée pendant ces trois années. Merci pour votre enthousiasme, vos remarques pertinentes, vos conseils avisés et surtout votre grande disponibilité.

Je remercie les membres de l'équipe Imadoc pour leur accueil, leur soutien et l'intérêt qu'ils ont porté à mes travaux. Un grand merci à tous pour les moments sympathiques passés autour des pauses cafés, des séminaires au vert.

Je remercie les amis qui m'entourent, en particulier Xavier pour les midis musicaux, Peggy pour les pauses bavardage, et tous ceux avec qui les repas du midi ne sont jamais monotones.

Enfin, un grand merci à Gaël pour m'avoir soutenue et écoutée, jour après jour, au cours de ces trois années. Merci d'être là et de m'avoir offert ce nom à rallonge sur la couverture...

Table des matières

Introduction

Ces dernières années, une partie de la communauté de l'analyse et de la reconnaissance automatique de documents, s'est orienté vers le traitement des documents historiques. En effet, de vastes campagnes de numérisation sont organisées par les centre d'archives, les bibliothèques, les musées, dans le but de préserver le patrimoine et d'en faciliter l'accès au grand public.

Quelques outils ont été mis en place pour gérer ces bases d'images de documents anciens, mais il reste encore de très grandes masses d'informations qui ne sont pas encore accessibles. Les travaux de recherche actuels visent donc exploiter ces documents dans le but de les indexer, voire même de les retranscrire partiellement ou dans leur globalité.

A l'heure actuelle, il existe dans le commerce des outils de reconnaissance de caractères, appelés OCR (Optical Character Recognition), qui permettent de reconnaître des documents imprimés de bonne qualité. Cependant, les performances de ces outils diminuent rapidement avec des documents anciens ou manuscrits. En effet, ces documents présentent des difficultés particulières, telles que la présence de taches, pliures, déchirures...; l'écriture manuscrite est souvent irrégulière. L'analyse automatique des documents d'archives est donc encore un sujet de recherche.

Afin de réaliser l'analyse automatique de documents, une première étape consiste à en reconnaître la structure. En utilisant cette structure, il sera alors possible de chercher à reconnaître l'écriture uniquement dans des zones appropriées. Nos travaux de recherche s'inscrivent donc dans la reconnaissance automatique de la structure de documents manuscrits ou dégradés.

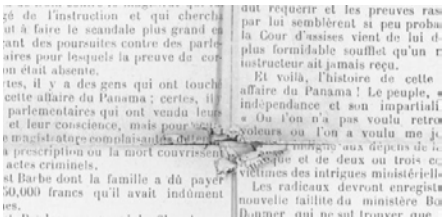
Difficultés rencontrées en reconnaissance de la structure de documents

Antonacopoulos et Downton [AD07] rappellent les difficultés rencontrées dans les documents anciens. Un des principaux obstacles à la reconnaissance de la structure est la présence d'artefacts de différents types. Ces auteurs listent les principales dégradations rencontrées, dont plusieurs sont présentées sur la figure 1. Ils les regroupent en trois

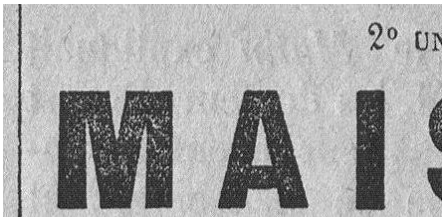
grandes catégories :

- dégradations liées au support papier :
  - pliures (figure 1(a)),
  - déchirures (figure 1(a)),
  - texture forte (figure 1(b)) ;
- dégradations liées au contenu et à l'encre :
  - mouchetage (figure 1(b)),
  - tâches (figure 1(c)),
  - encre pâle (figure 1(a)),
  - encre du verso visible (figure 1(c)) ;
- dégradations liées à l'étape de numérisation :
  - introduction de biais ou de courbure (figure 1(d)).

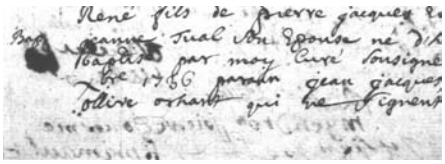
Dans la suite de nos travaux, nous regroupons sous le terme de *bruit* l'ensemble de ces artefacts. Ils sont présents dans les documents anciens ou dégradés, imprimés ou manuscrits.



(a) Pliure verticale, déchirure horizontale et encre pâle



(b) Texture du papier importante et mouchetage dans les caractères



(c) Tâches et encre du verso visible

(d) Courbure liée à la numérisation

FIG. 1 – Difficultés pour la reconnaissance de documents

De nombreux travaux ont été proposés pour la reconnaissance de la structure de documents. Mao *et al.* [MRK03] présentent un état de l'art des méthodes utilisées dans ce domaine. Ils distinguent la reconnaissance de la structure physique, de la reconnaissance de la structure logique. Ils présentent les limites principales des approches recensées, que nous rappelons ci-dessous.

- les méthodes présentées sont dédiées à une séparation des titres, sous-titres, texte,

images ; elles ne sont donc pas suffisamment génériques pour être appliquées à des documents historiques dont la structure peut varier d'une collection à une autre ;
- de nombreux travaux visant à étiqueter la structure logique du document supposent que le découpage physique a été fait. On rejoint ici le paradoxe de Sayre [1] : la segmentation physique requiert parfois la reconnaissance du contenu logique ;
- la plupart des approches ne sont pas capables de gérer correctement le bruit présent dans les documents.

Les travaux résumés dans l'introduction d'Antonacopoulos et Downton [AD07] montrent que le problème de la reconnaissance de documents anciens, dégradés ou manuscrits est encore ouvert.

Contexte de la thèse

Afin de résoudre certains problèmes et de simplifier la mise au point de systèmes de reconnaissance, nous proposons d'introduire la notion de vision perceptive.

La vision perceptive est cette faculté qu'a le cerveau humain de combiner différents niveaux de vision pour interpréter une scène. Ainsi, le cerveau est capable d'utiliser la vision globale d'une scène pour en comprendre son ensemble, puis de focaliser son attention sur un point précis afin d'observer un objet particulier en détail, en tenant compte de son contexte.

Nous proposons d'utiliser ce mécanisme de vision perceptive pour la reconnaissance de la structure de documents. En effet, il semble qu'utiliser une perception globale (à faible résolution) du document permet de limiter l'influence du bruit et de faire ressortir certains éléments structurels. La vision de près (à haute résolution), en revanche, permet de détailler les objets analysés. Notre idée est donc d'utiliser un mécanisme de prédiction/vérification : la vision globale permet de prédire des hypothèses sur la nature et la position d'éléments structurels ; la vision de près, guidée par une connaissance globale du contexte, permet de valider cette hypothèse et d'ajuster localement la position et l'interprétation des éléments structurels.

Nous montrons que cette approche permet de faciliter et d'améliorer la reconnaissance de la structure de documents dans deux principaux cas :
- lorsque le document contient des informations denses vis à vis de la structure à reconnaître (documents bruités ou informations superflues), la vision perceptive permet de sélectionner plus facilement les informations nécessaires à la reconnaissance de la structure ;
- à l'opposé, lorsque le document contient une information structurelle diffuse (documents faiblement structurés), le mécanisme de prédiction/vérification de la vision perceptive permet de mieux reconstituer la structure du document.

[1] Paradoxe de Sayre (1973) : *A letter cannot be segmented before having been recognized and cannot be recognized before having been segmented.*

Contenu du manuscrit

Ce manuscrit est décomposé en trois parties : introduction à la vision perceptive, implémentation de notre approche perceptive, et évaluation des apports pour la reconnaissance de structure de documents.

La première partie introduit le concept de la vision perceptive.

Le chapitre 1 décrit la vision perceptive telle qu'elle est présentée par les neuropsychologues. Ils mettent en évidence deux aspects : le cycle perceptif qui décrit les mécanismes physiologiques mis en œuvre lors de la vision, et l'attention visuelle qui est l'aspect psychologique guidant le cycle perceptif.

Le chapitre 2 recense les approches de la littérature utilisant la vision perceptive, dans le domaine du traitement d'images et particulièrement de la reconnaissance de documents.

Le chapitre 3 met en avant différents problèmes pour lesquels l'utilisation de la vision perceptive améliore intuitivement le traitement des documents.

Cette première partie permet donc de situer nos travaux par rapport aux approches perceptives existant dans la littérature, et de mettre en évidence les apports de la vision perceptive pour le traitement de difficultés habituellement rencontrées en analyse de documents.

La deuxième partie présente la nouvelle méthode générique que nous avons proposée. Son fonctionnement se base sur les principes de la vision perceptive : le cycle perceptif et l'attention visuelle.

Le chapitre 4 met d'abord en évidence les différents éléments requis pour créer un système de vision perceptive. Il conclut sur une solution d'implémentation basée sur trois modules : la réutilisation d'une méthode existante (la méthode DMOS), la création d'outils de multirésolution, la gestion d'objets structurels dits *prégnants*.

Le chapitre 5 expose la méthode existante DMOS (Description et MOdification de la Segmentation), et explique en quoi elle constitue une base solide pour notre implémentation de la vision perceptive.

Le chapitre 6 présente les outils de multirésolution et les formalismes que nous avons intégrés dans la méthode DMOS, afin de permettre l'imitation du cycle perceptif.

Le chapitre 7 décrit la manière dont sont gérés les objets prégnants dans la méthode DMOS enrichie des outils de multirésolution. Ceci représente notre deuxième contribution à la méthode DMOS et mène à l'obtention de notre système global, DMOS-P (DMOS Perceptif).

Dans cette deuxième partie, nous présentons donc le fonctionnement interne de notre méthode DMOS-P. Nous mettons en évidence sa généricité, sa simplicité d'utilisation et sa souplesse d'adaptation, pour la description de mécanismes complexes de coopération perceptive. Ceci est permis principalement par la séparation de la connaissance qui modélise et contrôle les interactions entre les différents niveaux de perception.

La troisième partie présente différents cas d'utilisation de la méthode DMOS-P, pour lesquels nous évaluons les apports de la vision perceptive pour la reconnaissance de la structure de documents.

Le chapitre 8 présente une première application pour la reconnaissance de la structure de courriers manuscrits. Cette application se place dans le contexte d'un projet national, le projet RIMES, ce qui nous permet de situer notre approche par rapport aux méthodes présentées par d'autres équipes de recherche.

Le chapitre 9 montre les apports de notre méthode dans le cas de l'analyse d'un grand volume de documents d'archives, de type décrets de naturalisation. La finalité de cette application est de faciliter l'accès aux documents pour les usagers des archives nationales.

Le chapitre 10 permet de mettre en avant la généricité de notre approche, en l'utilisant dans un contexte différent. En effet, il s'agit ici de fournir un prétraitement pour la reconnaissance de l'écriture manuscrite d'un dialecte indien, le Bangla, et plus généralement de localiser des lignes de base dans de l'écriture manuscrite.

Enfin, le chapitre 11 présente une application de reconnaissance de filets dans des pages de journaux anciens, bruités et abîmés. Ce travail est une étape préliminaire pour la reconnaissance de mots clés et l'indexation de ces pages de journaux. Il a donné lieu à un transfert industriel.

Les applications présentées dans cette troisième partie nous permettent donc de valider les différents aspects de la méthode DMOS-P, à savoir le fonctionnement de chacun des éléments de base puis leur assemblage pour former des mécanismes perceptifs complexes, dédiés à chaque document étudié. De plus, cette troisième partie montre comment notre méthode répond aux difficultés de reconnaissance des documents, comme il est pressenti dans le chapitre 3. Enfin, cette partie expose la validation de notre approche sur des problèmes variés, à grande échelle (plus de 80 000 documents), et par le biais d'un transfert industriel.

Première partie

Introduction à la vision perceptive

Introduction

Cette première partie est consacrée à l'étude de la bibliographie. Notre domaine d'étude regroupe deux vastes domaines de recherche : la vision perceptive et l'analyse de structure de documents. Nous avons donc choisi de focaliser cette étude bibliographique à l'intersection de ces deux domaines.

Dans le premier chapitre, nous abordons le principe de la vision perceptive, telle qu'elle est modélisée par les neuropsychologues. Nous présentons une synthèse des différents travaux proposés, orientée par une possible analogie avec l'étude d'images de documents.

Dans le deuxième chapitre, nous présentons de manière générale les approches utilisées pour la reconnaissance de structure de documents, en se focalisant au sous-ensemble des méthodes basées sur la vision perceptive. Ces méthodes sont représentatives des différents axes étudiés pour la reconnaissance de la structure de documents. En effet, les méthodes perceptives concernent aussi bien des approches ascendantes que descendantes, statistiques que grammaticales, bas-niveau que basées sur une connaissance sémantique.

Dans le troisième chapitre, nous abordons des problèmes plus précis, tels que la reconnaissance des lignes de texte, des traits, pour lesquels nous étudions les approches de la littérature afin de montrer l'intérêt de la vision perceptive.

Grâce à cette étude bibliographique orientée vers la vision perceptive, nous dégageons le besoin et l'utilité de combiner les idées des différents travaux existants pour former un système perceptif complet de reconnaissance de structure de documents qui sera exposé dans la suite du manuscrit.

Chapitre 1

Vision perceptive : approche neuropsychologique

Comme évoqué dans l'introduction, nous entendons par *vision perceptive* la capacité de combiner des visions à différents niveaux de résolution, dans le but de simplifier l'extraction des informations importantes dans une scène. En effet, lorsque l'œil humain regarde une image ou une scène de la vie courante, il ne peut pas analyser et interpréter tous les détails, tant l'information visuelle disponible est vaste. Il réduit donc son intérêt au niveau de *points d'attention*. L'interprétation est alors réalisée par la combinaison des différents niveaux de vision : globale ou locale, selon les besoins. C'est ce qui est appelé *vision perceptive*.

Pour mieux comprendre ce qu'est la vision perceptive, nous détaillons dans ce chapitre les mécanismes neuropsychologiques mis en œuvre par l'être humain. En effet, de nombreux travaux tentent de modéliser ces mécanismes utilisés pour la vision. A l'heure actuelle, les connaissances dans ce domaine permettent de décrire assez précisément le chemin suivi par un message visuel, et les différentes parties du cerveau impliquées [IK01] [Bar05]. Nous nous contenterons d'en présenter une approche globale.

Pour cela, nous nous basons sur le livre d'Itti [IRT05], *Neurobiology of attention*, qui recense, entre autres, les travaux traitant des principes de la vision perceptive. Il en découle que la vision perceptive peut être décrite autour de deux axes :

- un mécanisme physiologique, le *cycle perceptif*, qui permet la capture et le traitement successif des images [Tre92] [Nei76] ;
- un mécanisme psychologique, l'*attention visuelle*, qui spécifie les modalités du parcours du cycle perceptif pour la vision d'une information [IK01].

Nous détaillons ces deux composantes de la vision perceptive.

1.1 Mécanisme physiologique : le cycle perceptif

Treisman [Tre92] et Neisser [Nei76] proposent un modèle basé sur un cycle perceptif. Ce cycle perceptif est un mécanisme qui décrit l'acquisition successive d'images, leur interprétation puis leur confrontation avec des modèles existants en mémoire. Nous

présentons une synthèse des différentes étapes sur la figure 1.1.

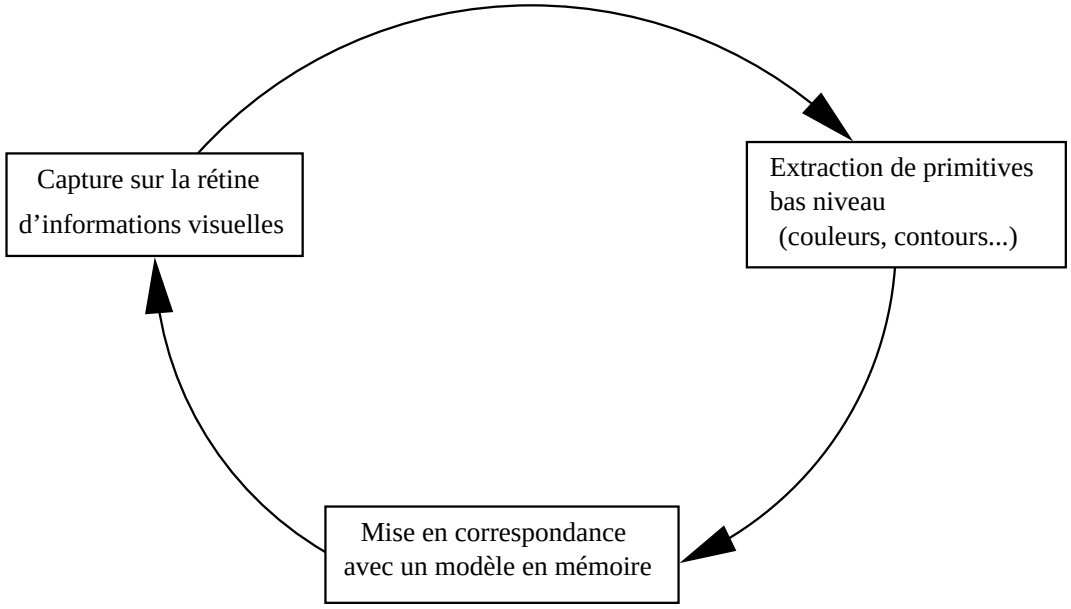


FIG. 1.1 – Les trois étapes du cycle perceptif

Dans les parties suivantes, nous détaillons ces trois étapes, avant d'expliquer comment elles s'enchaînent pour produire un cycle.

1.1.1 Capture d'informations par la rétine

La première étape nécessaire à la vision est l'acquisition d'une image par la rétine. Il s'agit de la perception, pour chaque point, d'une relation directe entre une position dans l'espace et une intensité lumineuse [Coh00].

La particularité des images perçues par la rétine est la variabilité de leur résolution selon la position du point d'intérêt. En effet, l'aire visuelle se découpe en deux zones : l'aire fovéale où l'information est à haute résolution, et l'aire périphérique contenant une information à plus faible résolution[1]. Cette possibilité de varier la résolution nous permet d'examiner chaque objet avec une précision adaptée dans la zone fovéale, tout en conservant une attention globale grâce à la vision périphérique.

1.1.2 Extraction de primitives

A partir de l'image captée par la rétine, des analyseurs du cortex extraient des propriétés bas niveau, présentes devant le point d'attention. Ces informations sont appelées *pré-attentives* puisqu'elles sont disponibles pour le cerveau avant toute forme d'interprétation ou de sélection.

[1]Selon les conventions habituellement utilisées, la résolution *haute* correspond à une vision très détaillée alors que la résolution *basse* ou *faible* permet uniquement une vision globale, avec peu de détails.

Wolfe [Wol05] recense une liste finie de caractéristiques ainsi extraites, parmi lesquelles la couleur, la forme, l'orientation... Ce sont ces caractéristiques qui sont ensuite organisées selon les modèles contenus dans la mémoire.

1.1.3 Modèle en mémoire

Dans cette troisième étape de la vision perceptive, l'utilisateur reconstruit des objets cohérents en combinant intelligemment les caractéristiques pré-attentives extraites à l'étape précédente.

Ces objets sont produits grâce à des modèles présents en mémoire, stockés dans le cortex. Noton [Not70] présente une théorie sur la mémorisation des modèles d'objets : étant donné que le cortex reçoit en entrée des caractéristiques présentes au niveau du point d'attention, mémoriser une forme c'est stocker ses caractéristiques et leurs positions relatives. Reconnaître une forme, c'est retrouver dans l'image l'ensemble des caractéristiques mémorisées pour cette forme.

Cette théorie sur la mémoire est reprise par Arathorn [Ara05] qui présente un simulateur d'attention guidé par la mémoire, capable de reconnaître une image 2D à partir d'un modèle mémorisé en 3D.

Rybak *et al.* [RGG+05] précisent que les informations mémorisées sont de deux type : le *quoi* qui correspond à la nature de l'objet à reconnaître, stocké dans la mémoire sensorielle, et le *où* qui correspond à la position de l'objet, stocké dans la mémoire motrice.

D'autre part, les auteurs [Not70] [DZ01] s'accordent à dire que, pour un même objet, on dispose de multiples niveaux de représentation interne, ce qui nous permet de reconnaître une forme soit en ayant une impression globale, soit à partir de détails.

Chun précise dans [Chu05] que des éléments du contexte sont également mémorisés. Il distingue trois types d'informations : le contexte spatial sur la position de l'objet, le contexte de voisinage qui prend en compte les objets qui peuvent être situés dans un environnement proche, et enfin le contexte temporel qui permet de prévoir à quel moment on pourra détecter un objet.

En fonction de la manière dont sont interprétées les caractéristiques, la mémoire peut avoir besoin de plus d'informations pour reconnaître un objet donné, ce qui nécessite d'acquérir une nouvelle image. C'est pourquoi le mécanisme de reconnaissance est cyclique.

1.1.4 Enchaînement en cycle

La représentation de la figure 1.1 traduit l'aspect cyclique de la prise d'information chez l'homme.

En effet, le processus d'analyse d'une image est itératif [Bar05]. En fonction de l'attention portée à certains objets et de l'interprétation qu'en fait la mémoire, la région d'attention évolue et la zone fovéale est modifiée [SAA01]. Une nouvelle itération du cycle peut donc démarrer dans ces nouvelles conditions.

Le point clé pour guider les conditions de parcours du cycle est l'attention visuelle pour certains objets, c'est ce que nous allons détailler dans la suite.

1.2 Mécanisme psychologique : l'attention visuelle

La vision et l'interprétation d'une image sont fortement guidées par une composante psychologique : l'attention visuelle. L'attention visuelle est le mécanisme qui permet de trier les informations visuelles disponibles, à la fois en filtrant les informations à étudier et en orientant le regard vers des points particuliers [IK01]. L'attention est donc au cœur du cycle perceptif, tel que complété sur la figure 1.2.

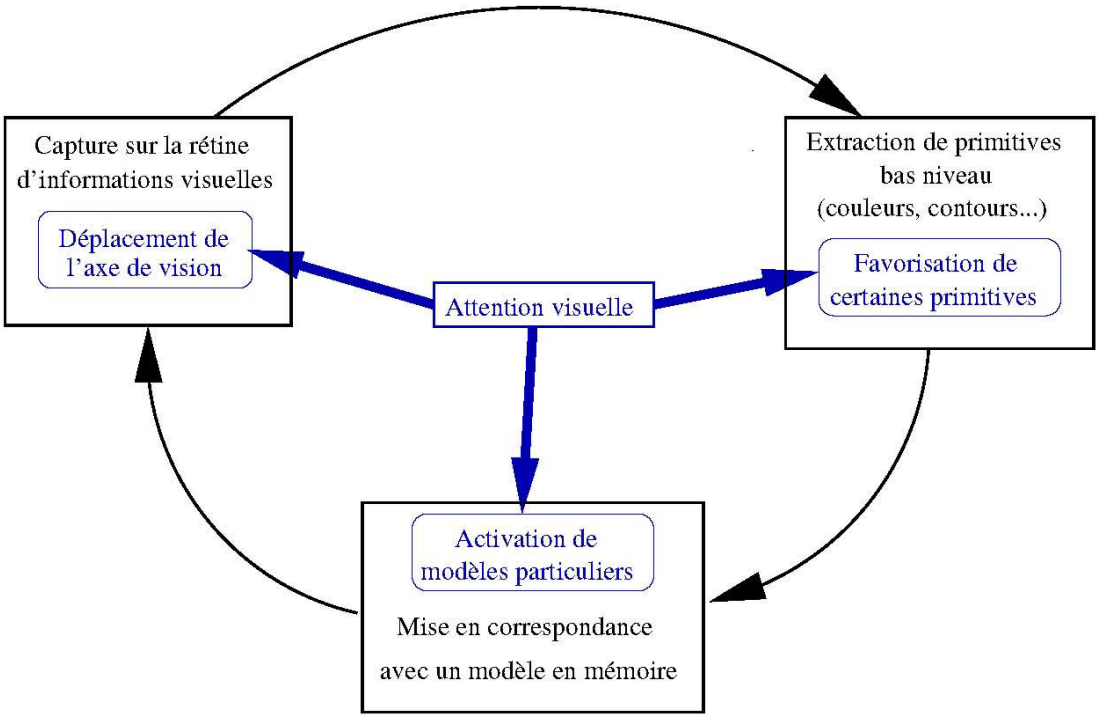


FIG. 1.2 – Le rôle de l'attention dans le cycle perceptif

Cette figure montre que l'attention visuelle est responsable de l'extraction de certaines caractéristiques plus que d'autres. Elle permet également de déplacer l'axe de vision en fonction des modèles activés en mémoire.

L'attention visuelle guide le cycle perceptif selon deux grands modes de fonctionnement : l'attraction par un objet prégnant et la recherche d'un but précis [IK01].

Lorsqu'on examine une scène, certains stimuli sont naturellement *prégnants* dans un contexte donné. Par exemple, lorsqu'on circule en voiture, la perception d'un feu tricolore attire naturellement notre attention. C'est le phénomène du *pop-out*. Dans ce cas, l'attention est guidée par ces objets aux caractéristiques prégnantes. Notre cerveau va ensuite chercher, à partir des éléments prégnants, à élargir le champ de vision pour

en comprendre le contexte : on parle d'analyse ascendante.

Un mécanisme opposé intervient lorsqu'on étudie une scène en étant guidé par la recherche d'un objectif précis. Par exemple, si on cherche sa voiture rouge sur un parking, notre attention va se focaliser uniquement sur les objets rouges, et occulter le reste de la scène. L'attention est alors dite *guidée par un but*. Les caractéristiques extraites correspondent uniquement à celles correspondant à la description du but, ici la couleur rouge. A partir d'une vision globale, notre cerveau émet des hypothèses sur la présence d'objets qui seront validées localement. Dans le cas de l'exemple précédent, l'œil va détailler chaque entité vérifiant le modèle de « voiture rouge » pour essayer de reconnaître l'unique véhicule cherché. On parle d'analyse descendante.

Le mécanisme d'attention guidé par un élément prégnant est automatique tandis que l'attention guidée par un objectif demande un effort visuel. La vision perceptive est composée de ces deux types d'attention, qui peuvent fonctionner en parallèle.

De nombreux travaux en neuropsychologie tentent d'expliquer davantage le fonctionnement de ces deux mécanismes, que nous détaillons dans les parties suivantes.

1.2.1 L'attention guidée par des éléments prégnants

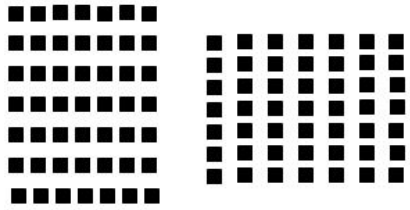
L'attention guidée par la prégnance est un phénomène automatique qui déplace le point d'attention sur un élément qui « saute aux yeux » dans l'image (phénomène de *pop-out*). Cet élément est repéré principalement par des caractéristiques qui diffèrent des objets voisins [Wol05]. A partir de ces objets prégnants, l'humain produit automatiquement une organisation des objets par un groupement ascendant : c'est ce qu'on appelle l'organisation perceptive.

L'organisation perceptive peut être définie comme la capacité à imposer une organisation structurelle à des données sensorielles, c'est à dire les caractéristiques issues de l'image perçue sur la rétine. Les règles de l'organisation perceptive sont décrites dans la théorie de la forme (*Gestalt*) [Kof35], proposée par des psychologues allemands au début du 20ème siècle. Cette théorie met en avant certains critères que l'homme utilise intuitivement pour regrouper des éléments entre eux. Citons parmi ces facteurs de regroupement :

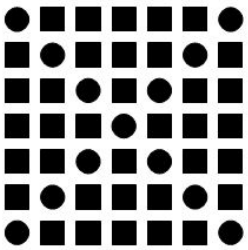
- la *proximité* entre éléments (figure 1.3(a)),
- la *similitude* entre éléments (figure 1.3(b)),
- la *continuité* qui permet de distinguer des éléments en inférant des traits absents (figure 1.3(c)),
- la *simplicité* qui favorise la reconnaissance d'éléments connus (figure 1.3(d)),
- la *fermeture* qui regroupe des éléments formant une entité complète (figure 1.3(e)).

Ces lois servent de base dans la plupart des travaux qui décrivent la vision perceptive dans le cas où aucune connaissance *a priori* n'est fournie sur les objets à retrouver. Sarkar et Boyer [SB93] présentent un état de l'art qui recense les modèles de l'attention ascendante, et de l'interprétation qu'on peut faire d'une scène à partir d'éléments prégnants et de leurs caractéristiques.

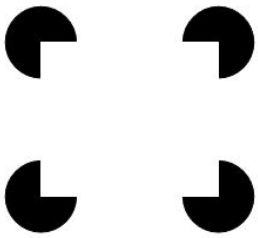
L'attention produite par des éléments prégnants et l'organisation perceptive associée conditionnent donc le parcours du cycle perceptif d'une manière ascendante et automa-



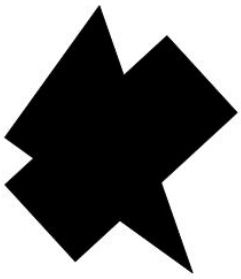
(a) Proximité : on regroupe les points à gauche en lignes horizontales, à droite en lignes verticales



(b) Similitude : on distingue une croix faite de ronds au milieu des carrés



(c) Continuité : on perçoit un carré blanc



(d) Simplicité : on perçoit un triangle et un rectangle

(e) Fermeture : on regroupe les traits en un tout : un visage

FIG. 1.3 – Les lois de l'organisation perceptive

tique, lorsque rien ne prédétermine le contenu de l'image à traiter. On oppose à ce type d'analyse l'attention guidée par la recherche d'un but précis, que nous détaillons dans la partie suivante.

1.2.2 L'attention guidée par un but

L'attention guidée par le but est un mécanisme qui demande un effort visuel. Cela consiste à localiser et segmenter uniquement les objets du champ visuel qui sont contextuellement valides par rapport à un modèle contenu dans la mémoire.

La scène est d'abord analysée avec une vue globale, à faible résolution, puis l'attention est focalisée successivement sur des zones plus précises jusqu'à l'identification complète de l'objet cherché [DZ01]. Le point clé de cette approche descendante est que l'information globale disponible à basse résolution sur les formes (orientation générale et proportions) est suffisamment typique pour activer un ensemble relativement petit de candidats [Bar05]. Ce mécanisme permet donc de limiter le nombre de modèles à prendre en considération.

D'autre part, Noton [Not70] montre que la vision perceptive permet de reconnaître des modèles dans des conditions défavorables. En effet, la vision globale permet de détecter les caractéristiques de base de l'objet cherché et d'émettre une hypothèse sur sa localisation. La focalisation d'attention permet alors de ne pas tenir compte du bruit ou de la confusion.

L'étape de reconnaissance d'un objet contenu en mémoire correspond selon Rybak [RGG$^+$05] à un véritable programme comportemental de reconnaissance créé lors de la première vision de l'objet. Ce programme contient des mouvements programmés par la mémoire motrice qui précise où aller chercher chacune des caractéristiques de l'objet à reconnaître, ainsi que des fragments d'images prédits par la mémoire sensorielle, qui doivent correspondre à la réalité.

Deco et Zihl [DZ01] présentent une modélisation neurodynamique basée sur des modèles hiérarchiques. Dans ces travaux, les auteurs cherchent à modéliser les interactions entre les différentes aires du cortex. Ils associent à chaque niveau de résolution un module capable d'interpréter l'image. Au début de l'analyse, seul le module de faible résolution est actif. L'image est ainsi analysée à faible résolution, et va prédire la position la plus probable de l'objet cible. On répond ainsi à la question *où*. Le focus d'attention va alors accroître la résolution dans cette zone définie, jusqu'à ce que l'objet soit identifié grâce à des caractéristiques, répondant ainsi au problème du *quoi*. Le point fort de ce mécanisme est que le contrôle de l'attention décide itérativement dans quelle région il faut accroître la résolution. Les auteurs expérimentent cette théorie dans le cas de l'attention guidée par un but : localiser un motif donné, compter le nombre d'occurrences d'un motif donné.

Ces travaux vont dans le sens de l'existence d'un cycle perceptif guidé par l'attention (figure 1.2) qui précise à chaque étape la nouvelle résolution et le nouveau point d'attention à étudier, le *où*, et quelles sont les caractéristiques à extraire, le *quoi*.

1.3 Intérêt de l'attention dans la vision perceptive

L'attention visuelle, qu'elle soit guidée par des éléments prégnants ou par un but précis, est un mécanisme de sélection, responsable de filtrer les informations requises pour l'analyse d'une scène. Cette sélection est nécessaire dans la mesure où une scène de la vie courante contient plus d'informations que ce qu'on ne peut en interpréter [IRT05]. Ainsi, la compréhension d'une scène est réduite à l'interprétation de petits problèmes locaux situés dans des zones visuelles très précises. Ceci est permis par la vision perceptive qui sélectionne une localisation d'intérêt avant de conforter la représentation d'un objet à cette position par une étude plus précise [IK01].

La vision perceptive guidée par une attention automatique sur des objets prégnants et l'organisation perceptive permettent d'interpréter des images sans avoir de connaissance *a priori* sur le contenu. A l'opposé, la vision perceptive guidée par un but permet de reconnaître un objet à partir de son modèle, dans un contexte difficile, en ignorant les éléments parasites.

Nous allons maintenant voir comment ces atouts sont utilisés pour la reconnaissance de documents.

Chapitre 2

Approches perceptives en analyse d'images

Les principes de la vision perceptive sont largement utilisés dans le domaine de la vision par ordinateur et le traitement d'images naturelles. De nombreuses approches sont proposées autour de l'analyse multirésolution des images et du traitement multi-échelle des données. En revanche, dans le domaine de l'analyse de documents, les approches perceptives n'ont été développées que plus récemment.

Nous proposons donc une présentation rapide de quelques utilisations de la vision perceptive pour le traitement d'images naturelles, avant de détailler les approches existantes en analyse de documents.

2.1 Traitement d'images naturelles

Une des premières utilisations de la vision perceptive, trouvée dans la littérature, est proposée par Bajcsy et Rosenthal [BR77] [BR80]. Ils observent qu'un spectateur ne regarde pas une scène complète avec la même intensité, mais se focalise visuellement sur les objets qui attirent son attention. Pour imiter ce mécanisme, ils ont mis en place un système basé sur une double hiérarchie : la hiérarchie visuelle et la hiérarchie conceptuelle. La hiérarchie visuelle correspond à une pyramide d'images de différentes tailles. La hiérarchie conceptuelle exprime le fait que la connaissance varie selon le niveau d'analyse. Dans ces travaux, le lien entre les deux hiérarchies est réalisé par une structure de contrôle qui guide les interactions entre les différents niveaux. Rosenthal enrichit ensuite ces travaux [Ros84] en proposant des règles de description pour chacun des éléments. Ces travaux posent des bases fondamentales pour l'approche perceptive mais ont été appliqués uniquement sur quelques images dédiées.

L'idée d'imiter la vision humaine a été reprise par Burt [Bur88] qui propose la *Pyramid Vision Machine*. Ce système est basé sur une pyramide Gaussienne d'images construites en sous-échantillonant récursivement l'image initiale. Ces images sont ensuite utilisées dans l'ordre inverse, de la plus grossière à la plus fine, afin de localiser rapidement les objets cherchés dans une scène. Ce système est contrôlé par un mé-

canisme haut niveau qui guide le regroupement des données au fur et à mesure de l'interprétation des informations visuelles. Cette méthode est dédiée à la reconnaissance d'objets dans des mécanismes de surveillance et d'alerte.

Silberberg [Sil88] propose une structure pour la vision multirésolution d'objets. Il stocke les données dans un tableau multirésolution qui contient les propriétés des pixels, les propriétés des objets et les relations spatiales entre objets à chaque résolution. Il propose une manière d'interpréter les images en fonction des données multirésolutions présentes : à chaque résolution, une hypothèse est émise sur la présence d'éléments. Les hypothèses sont confirmées lorsque ces objets sont détectés à différentes résolutions, selon un principe d'accumulation de preuves. Cette méthode est basée sur deux modules : une description symbolique de l'image et un mécanisme d'interprétation dont le but est d'essayer d'appliquer les modèles d'objets. La philosophie de ce travail est intéressante, mais il est appliqué uniquement sur deux images : une image sous-marine et une image prise par avion. L'implémentation semble donc très dédiée à ces images.

Jolion et Rosenfeld [JR94] présentent dans leur livre *A pyramid framework for early vision* un panorama de l'utilisation des structures pyramidales pour la combinaison des points de vue à différents niveaux de résolution. Cependant, ces travaux sont liés à des traitements de vision précoce et n'incluent pas de connaissance haut niveau.

Dyer [Dye87] synthétise les points clés de l'analyse d'images multi-échelle, que sont :

- la persistance de propriétés à des échelles différentes,
- la détection globale du contenu en utilisant des opérations à faible échelle,
- la possibilité de faire la recherche d'un objet donné en allant du plus grossier au plus précis,
- l'organisation hiérarchique des connaissances.

Il met également en avant les intérêts de la stratégie *coarse-to-fine*, allant d'une image grossière à une image plus fine. En effet, cela permet de :

- zoomer itérativement à la position exacte de l'objet, vers des résolutions de plus en plus fines,
- vérifier des hypothèses émises à des résolutions plus faibles,
- gagner en temps de calcul, en appliquant les traitements coûteux uniquement sur des zones spécifiques.

Ces éléments clés sont également repris dans les travaux de Bottoni *et al.* [BCLM98] qui exposent une méthode pour déterminer la résolution satisfaisante pour appliquer des traitements, sans être encombrés par trop de détails.

La vision perceptive a été appliquée à des problèmes variés : extraction de contours [PK89] [ML95], analyse d'images aériennes [CT90], détection de routes dans des images satellitaires [BSME97] [MLBS97], détection de caractères dans des images naturelles [AGF04].

Dans tous ces problèmes d'analyse d'images naturelles, l'intérêt d'imiter la vision humaine a été démontré. Cependant, dans ce contexte, les traitements bas niveaux sont souvent très liés au contexte applicatif. Nous allons donc maintenant montrer comment la vision perceptive est utilisée en analyse d'images de documents, avec des connaissances de plus haut niveau sur la structure à reconnaître.

2.2 Analyse de documents

Dans le contexte de l'analyse de documents, de nombreux travaux s'appuient sur l'étude d'une même image à différents niveaux de résolution ou de perception. Ce type d'analyse est particulièrement adapté pour les documents. En effet, la nature multi-échelle des constituants d'un document (caractères, mots, lignes, paragraphes) justifie l'utilisation d'une analyse à différents niveaux de perception [EDC97].

Eglin présente [Egl06] un état de l'art des approches cognitives et perceptives pour la reconnaissance des documents. Elle classe ces approches selon l'importance que prend la part d'imitation de la vision humaine : utilisation ponctuelle ou système global.

Il nous semble intéressant d'apporter un autre éclairage sur ces travaux en les classant selon la forme d'attention à laquelle ils réfèrent. En effet, comme nous l'avons montré dans la partie 1.2, la vision perceptive peut être guidée par deux types d'attentions visuelles : l'attention due à des informations particulièrement prégnantes lorsqu'on ne cherche rien de précis, ou l'attention guidée par un but, c'est à dire la recherche d'un motif précis. Cela se traduit, pour l'analyse de documents, par deux types de mécanismes de reconnaissance. Les modèles imitant l'attention guidée par des éléments prégnants sont liés à des analyses ascendantes du document : on procède à des regroupements successifs d'éléments, construits grâce à une vision multi-échelle. A l'opposé, lorsque la présence de connaissances sur le modèle vise à extraire une structure précise, on procède à des focalisations d'attention successives sur des points d'intérêt, depuis la globalité du document. On parle alors d'analyse descendante.

Nous présentons donc, dans les deux parties suivantes, les approches utilisant la vision perceptive et la combinaison d'analyse à plusieurs résolutions pour la reconnaissance de la structure de documents. Nous distinguons les approches ascendantes guidées par les données, des approches descendantes guidées par un but précis.

2.2.1 Approches ascendantes guidées par les données

Les approches ascendantes en vision perceptive utilisent des informations locales qui sont regroupées successivement, en tenant compte de la hiérarchie du document.

On distingue les analyses qui se basent sur une étude bas niveau de caractéristiques, de celles qui proposent de reproduire les mécanismes de l'organisation perceptive.

2.2.1.1 Analyses bas-niveau

Plusieurs travaux procèdent à des regroupements successifs des constituants d'un document, en se basant uniquement sur une analyse bas niveau des éléments, principalement le voisinage entre composantes connexes et l'analyse de la texture. Un des outils fréquemment utilisé pour l'analyse est la pyramide irrégulière qui permet une analyse multirésolution d'un document [Ber95], et une organisation multi-échelle de ses composants.

Ainsi, les travaux de Loo et Tan [LT01] [LT02] [LT03] se basent sur un graphe irrégulier, permettant une segmentation qui tient compte du voisinage entre les éléments.

Les composantes connexes du document forment les sommets du graphe et sont successivement regroupées selon leur voisinage. A la fin de l'analyse, les sommets du graphe sont les mots ; les relations de voisinage expriment les distances entre les mots. Ces travaux utilisent donc l'aspect hiérarchique du document pour construire récursivement sa structure.

Cette idée de construction de pyramide par agglomérations successives est reprise par Lee *et al.* [LR01] qui proposent d'utiliser une analyse des fréquences spatiales pour regrouper successivement les régions du document. On retrouve ici un des grands principes des approches basées sur les textures, qui utilisent de manière générale une combinaison d'images à plusieurs résolutions. Les travaux d'Etemad *et al.* [EDC97] se basent également sur une analyse multi-échelle de la texture. Leur approche vise à découper des documents en éléments de texte, images et graphique. Bloomberg [Blo91] propose aussi une approche multi-échelle pour détecter les formes et les textures. La spécificité de sa méthode vient de la manière de construire les différentes résolutions, à savoir appliquer des filtres basés sur la dilatation et l'érosion morphologique.

Les travaux de Rangoni et Belaïd [RB06] poussent plus loin la combinaison entre les niveaux de vision en modélisant des cycles perceptifs. Leur méthode est basée sur un réseau de neurones qui utilise l'observation de caractéristiques physiques pour regrouper au fur et à mesure les éléments et former une structure logique. Les phases de vision - interprétation sont réalisées successivement jusqu'à l'obtention d'une structure cohérente. Cette itération est qualifiée de cycle perceptif.

Tous ces travaux sont des exemples de la vision perceptive dans le sens où on combine les éléments à différents niveaux de vision. Ces combinaisons peuvent être successives et même cycliques. Afin de cadrer davantage la manière dont les éléments sont regroupés entre eux dans une approche ascendante, certains auteurs appliquent plus directement les lois de l'organisation perceptive. C'est ce que nous montrons dans la partie suivante.

2.2.1.2 Construction de structure selon les théories perceptives

Likforman et Faure proposent une utilisation de la vision perceptive pour la reconnaissance de ligne de texte dans des documents manuscrits [LSF94] [LSF95]. Leurs travaux sont basés sur le principe que l'œil humain peut percevoir des lignes de texte à distance, indépendamment de leur lecture proprement dite. Cette approche utilise des principes de la psychologie de la vision et des lois d'organisation perceptive de la théorie de la forme (*Gestalt*) [Kof35]. Pour construire les lignes de texte, un point d'ancrage sélectionné sert de base à l'agglomération des composantes connexes selon des principes de proximité, de continuité de direction, de similarité et d'alignement. Une fois ces regroupements locaux effectués, une analyse à un niveau plus global permet de résoudre des conflits éventuels entre lignes.

Les travaux de Wattenberg et Fisher [WF03] se basent également sur les lois de l'organisation perceptive. En effet, dans leurs travaux, ces lois permettent de mettre en correspondance des structures extraites à différentes échelles pour la reconnaissance de l'organisation de graphiques.

Cette idée est également utilisée par Sanchez *et al.* [SVL+04] qui proposent une

architecture générale pour la description est la reconnaissance de graphiques. Ces graphiques sont en effet décrits par des descripteurs basés sur des techniques de groupement perceptif, où les caractéristiques prégnantes sont extraites. Selon les auteurs, ces caractéristiques sont nécessaires, mais pas suffisantes pour déterminer l'existence d'un élément donné. Elles peuvent cependant être utilisées comme points de départ pour la recherche.

L'idée d'utiliser la théorie de la forme (*Gestalt*) est reprise dans une méthode plus globale proposée par Eglin [EBE99]. Son but est de produire une segmentation du document telle que la réaliserait un lecteur humain n'ayant pas de connaissance *a priori* sur le contenu du document. Elle considère que toutes les informations contenues dans le document n'ont pas le même poids visuel. L'œil humain étant attiré par certaines de ces informations, cela induit un ordre logique de parcours du document, qu'elle appelle *survol*. De plus, ce survol du document a lieu dans un contexte multirésolution puisque l'œil est capable d'acquérir des images à résolutions différentes en vision fovéale et périphérique.

Les travaux d'Eglin visent donc à reproduire cette notion de survol du document, tout en gardant une dimension de multirésolution. La mise en œuvre se base sur des images multirésolutions produites par imitation du rayon fovéal, dans lesquelles sont extraits les contours. La sélection des points d'attention lors du survol dépend davantage de l'organisation des caractéristiques visuelles, que des caractéristiques elles-mêmes. Ainsi, les critères de focalisation mis en œuvre sont la distribution des contours, la surface relative des formes, une mesure de symétrie, la compacité et la courbure. Ces travaux sont validés par une application à la segmentation de pages de magazines.

Le principal intérêt des approches guidées par l'organisation perceptive est qu'elles ne nécessitent pas de connaissance *a priori* sur le type de document étudié. Elles fonctionnent bien pour la séparation d'éléments identifiables : lignes de texte, paragraphes, gros titres, graphiques. Par contre, dans le cas de documents faiblement structurés, présentant peu d'éléments prégnants, ou au contraire pour les documents bruités où les éléments prégnants se retrouvent perdus dans la masse, il semble difficile d'appliquer cette approche perceptive ascendante.

2.2.2 Approches descendantes guidées par un but

Le principe des approches descendantes est de découper récursivement le document en fonction d'un modèle à reconnaître, appris statistiquement ou décrit selon des règles.

Nous séparons les méthodes statistiques et bas niveau, des méthodes de plus haut niveau dans lesquelles il est possible d'introduire des connaissances plus élaborées sur le type de document à étudier.

2.2.2.1 Méthodes statistiques et bas niveau

Nous présentons tout d'abord plusieurs travaux, axés sur la multirésolution, qui utilisent des méthodes statistiques. C'est le cas des travaux de Cheng *et al.* [CBA97] [CB98] [CB01]. Ils proposent une approche bayésienne pour la segmentation de pages de magazines en blocs, et leur classification. La méthode est basée sur la modélisation des

probabilités de transition entre des niveaux adjacents dans la structure multi-échelle. Cela permet de prendre en compte à la fois le contexte global et le contexte local. La principale limite de cette approche est que la segmentation doit être apprise sur des documents suffisamment homogènes. D'autre part, la méthode ne prévoit pas l'introduction de connaissances spécifiques pour des types de documents particuliers.

Une méthode de coopération bas niveau est proposée par Cantoni *et al.* [CCLM97] [CLM98] pour la segmentation de pages. Ils construisent une représentation pyramidale des images par des cartes de caractéristiques contenant la moyenne, la variance, un seuil et la médiane. Chaque carte de caractéristiques est construite à 4 résolutions. Trois processus précis sont ensuite appliqués, basés sur les cartes de caractéristiques : analyse du fond, analyse de graphiques et analyse de texte. Les éléments trouvés aux différentes résolutions sont ensuite comparés pour valider la segmentation, mais il n'y a pas de réelle coopération lors de l'analyse des résolutions.

Shi et Govindaraju proposent une autre méthode d'analyse bas niveau [SG05], en affichant la volonté d'imiter la multirésolution de l'œil humain. Les niveaux de résolutions sont construits en utilisant une carte dynamique de connectivité locale, avec un seuil de connectivité variable. Les blocs de données sont alors segmentés hiérarchiquement. Cette méthode a été validée pour la segmentation de colonnes et de graphiques dans des pages de journaux, de magazines et de livres. Cette méthode reste très bas niveau. Il semble donc difficile d'y inclure une connaissance spécifique qui serait requise par des documents plus complexes.

2.2.2.2 Méthodes contenant une connaissance plus complexe

La plupart des méthodes qui décrivent de manière multirésolution des objets complexes se basent sur le principe suivant : la vision globale permet de localiser grossièrement les zones d'intérêt dans lesquelles on peut localement appliquer un traitement approprié.

C'est le cas des travaux de Déforges [VGB91] [DB94] [DPVGB95], pour qui la coopération entre les différents niveaux de la pyramide multirésolution permet d'appliquer une binarisation adaptée à chacun des points d'intérêt. Dans le contexte de ses travaux sur la reconnaissance de blocs postaux, la vision perceptive permet donc de s'affranchir de la variabilité des caractères (taille, imprimé *vs* manuscrit). En revanche, les traitements effectués à faible résolution sont dédiés au problème des blocs postaux. De plus, la méthode ne prévoit pas la possibilité d'aller-retour entre les différents niveaux de résolution.

Cinque *et al.* [CFL$^+$99] proposent une application dédiée à la décomposition de pages de journaux. Ils proposent de faire collaborer deux niveaux de résolutions. La segmentation initiale de l'image est réalisée sur une image réduite par un facteur 16, simulant ainsi une vue globale. Une fois la décomposition en blocs effectuée à cette échelle, l'image est analysée à résolution haute, de manière distincte pour chacun des blocs. Les traitements aux deux résolutions sont donc bien disjoints et il n'est pas possible de remettre en cause la segmentation initiale. De plus, cette méthode est très liée aux difficultés liées à la décomposition de pages de journaux.

Les travaux de Ramel *et al.* [RVE98] utilisent également la vision globale dans le but de déterminer des zones locales dans lesquelles appliquer un traitement spécifique. En effet, ils traitent des documents composites, et disposent de *spécialistes*, capables de traiter des zones de texte, des formes pleines, des pointillés ou des courbes. La vision globale permet donc de déterminer à quel endroit il est nécessaire d'appliquer les extracteurs spécialistes. Deux limites apparaissent dans cette approche. D'une part, le choix des *spécialistes* dépend du type de document à étudier. L'étude d'un nouveau type de document demanderait donc potentiellement la création de nouveaux spécialistes (traitements bas niveau). D'autre part, et surtout, les résultats de la phase globale ne sont jamais remis en cause.

Ogier, Mullot *et al.* [OMLL95] [PAhM+97] proposent une méthode intéressante qui permet de remettre en cause les éléments reconnus avec une vision globale. En effet, ils présentent des travaux dont le but est de reconnaître des documents cadastraux, en combinant des hypothèses émises d'un point de vue global et vérifiées de manière locale. La particularité de leur approche est la mise en place d'un parcours cyclique entre ces deux points de vue, de manière à retraiter les informations au fur et à mesure, en respectant une cohérence globale. Cette idée nous semble particulièrement intéressante. La principale limite des travaux présentés est qu'ils font intervenir des spécificités liées aux documents cadastraux pour la réalisation des traitements bas niveau.

2.2.3 Bilan sur les approches de la littérature

Nous avons classé les approches perceptives pour l'analyse de documents selon deux axes, qui présentent tous les deux des intérêts.

Certains travaux se placent dans une analyse ascendante et imitent l'organisation perceptive basée sur l'attention pour des éléments prégnants. Ces méthodes ont l'avantage de nécessiter très peu de connaissance sur le document. Ce mécanisme fonctionne bien pour la séparation d'éléments identifiables : lignes de texte, titres, graphiques. Par contre, dans le cas de documents faiblement structurés, présentant peu d'éléments prégnants, ou au contraire pour les documents bruités dans lesquels les éléments prégnants perdent de l'importance, il semble difficile d'appliquer cette approche perceptive ascendante.

Dans le cas des documents bruités et lorsque l'on dispose de connaissances précises sur le type de document à reconnaître, une approche descendante semble particulièrement adaptée. En effet, elle permet d'émettre globalement une hypothèse sur la présence d'un objet, qui sera confirmée à plus haute résolution. Cependant, la plupart des systèmes de la littérature sont dédiés à un type de document précis, et la connaissance spécifique est imbriquée au niveau des traitements bas niveau, ce qui limite leur utilisation pour des documents de nature différente.
Nos travaux proposent de combiner les deux types d'approches perceptives, ce qui semble nouveau vis à vis des méthodes trouvées dans la littérature. En effet, il est intéressant de bénéficier des atouts des deux modes de l'attention visuelle.

Certains éléments structurels semblent particulièrement prégnants lorsqu'on analyse un document : lignes de textes et segments. Ces éléments de base peuvent donc être re-

connus indépendamment de toute connaissance sur le document. A l'opposé, la recherche de certains éléments structurels très précis dans un document requiert une connaissance sur le type de document. Dans ce cas, une description descendante semble bien adaptée. De plus, il est important de mettre en œuvre une coopération entre les résolutions, afin de pouvoir remettre en cause l'analyse réalisée à chacune des résolutions.

Il est également nécessaire de produire un système générique dans lequel la connaissance spécifique à un type de document donné pourra être introduite de manière indépendante aux traitements. En effet, la littérature propose des cas variés d'application de la vision perceptive, mais peu de méthodes génériques. Notre but est donc de produire un système unique, générique, permettant de créer des mécanismes de coopération perceptive adaptés à chaque type de problème étudié.

Dans cette optique, nous recensons dans le chapitre suivant un ensemble de cas concrets pour lesquels la vision perceptive semble pouvoir améliorer la reconnaissance de la structure de documents.

Chapitre 3

Intérêts de la vision perceptive pour la reconnaissance de documents

Le but de ce chapitre est de mettre en avant des exemples de problèmes pour lesquels la vision perceptive semble intuitivement faciliter et améliorer la reconnaissance.

Nous présentons d'abord les apports de la vision perceptive pour la reconnaissance d'objets structurels élémentaires, qui peuvent être considérés comme prégnants dans le document : les lignes de texte et les traits. En effet, ces objets peuvent être construits selon des règles de l'organisation perceptive (proximité, alignement), et peuvent être reconnus dans des documents variés, sans nécessiter de connaissance spécifique à un type de document.

Nous montrons ensuite les apports possibles de la vision perceptive pour reconnaître des documents ayant une structure plus complexe, grâce à la combinaison des différents niveaux de vision.

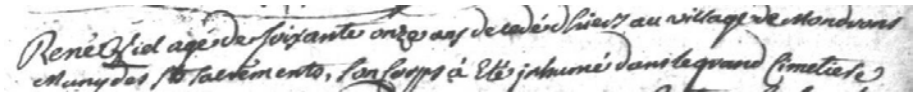
3.1 Analyse d'objets structurels élémentaires

Nous présentons deux types d'objets structurels : les lignes de texte et les traits. Ils sont dits élémentaires dans la mesure où, une fois extraits, ils pourront servir de base pour la reconnaissance de documents ayant une structure plus complexe.

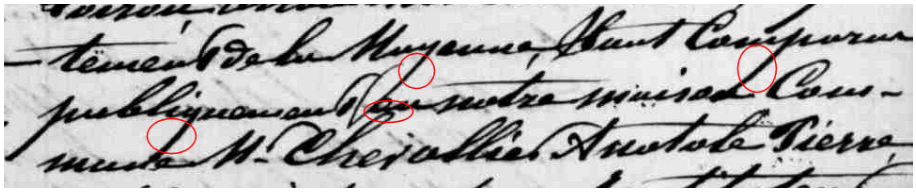
3.1.1 Lignes de texte

L'extraction de lignes de texte est considérée comme un problème résolu en ce qui concerne les documents imprimés [LZD06]. En revanche, c'est encore un domaine de recherche ouvert dans le cas de l'écriture manuscrite. En effet, l'analyse de documents manuscrits doit faire face à la variabilité du style du scripteur, à la courbure et au biais des lignes (figure 3.1(a)), au chevauchement des lignes entre elles (figure 3.1(b)).

Nous rappelons les différentes méthodes proposées dans la bibliographie avant de montrer intuitivement l'intérêt d'une approche par vision perceptive.



(a) Lignes incurvées et en biais



(b) Chevauchement des lignes entre elles (en rouge)

FIG. 3.1 – Difficultés rencontrées pour l'extraction de lignes de texte manuscrit

3.1.1.1 Approches de la littérature

Likforman-Sulem *et al.* proposent une étude détaillée des méthodes existantes pour la segmentation en lignes de texte dans les documents anciens [LSZT07]. Nous rappelons brièvement les grands axes de ces approches.

Une des approches utilisée classiquement pour l'extraction de lignes de texte imprimé est la projection. Cependant, cette méthode est très sensible au biais et à la courbure ; elle ne peut donc pas être directement appliquée au cas des documents manuscrits. De nombreux travaux s'en inspirent néanmoins et proposent des adaptations spécifiques, par exemple pour la détection du biais de la page [SGS93]. Différentes utilisations de la transformée de Hough ont également été proposées [LGPH08].

Afin de pouvoir traiter des lignes moins régulières, les auteurs ont proposé des méthodes basées sur des regroupements successifs de pixels ou de composantes connexes. Ainsi, Kise *et al.* [KIDM98] proposent une méthode d'agglutination dans un graphe de Voronoï, selon un seuil de distance. Cette méthode est capable de gérer du biais mais requiert un document suffisamment homogène pour la détermination du seuil de distance.

Plus récemment, des travaux dédiés au cas de l'écriture manuscrite ont été proposés. Par exemple, Nicolas *et al.* [NPH04] proposent d'apprendre des règles de production permettant de prendre des décisions locales sur le regroupement de composantes connexes en une seule ligne. Dans les travaux de Feldbach et Tönnies [FT01], on utilise le squelette de l'écriture pour détecter les lignes dans des registres paroissiaux. Li *et al.* [LZD06] proposent un regroupement ascendant par un principe d'extension morphologique des frontières. La difficulté dans ces approches ascendantes basées sur un regroupement local est de gérer le chevauchement entre lignes. En effet, certains alignements locaux peuvent mener à la construction de lignes globales qui ne sont pas pertinentes.

Afin de s'affranchir de la forte variabilité des textes, Déforges *et al.* [DPVGB95] proposent d'utiliser la vision perceptive, en utilisant une détection globale des objets linéiques afin d'adapter localement le seuil de binarisation au niveau de chaque ligne de texte. La perception visuelle est également utilisée par Likforman-Sulem *et al.* [LSF95] qui précisent qu'à une certaine distance du document, les lignes de texte apparaissent comme des segments. Ils s'inspirent donc de la théorie des formes (Gestalt) pour regrouper localement les pixels en un tout cohérent.

La principale difficulté rencontrée par les méthodes de la littérature est liée au chevauchement entre lignes. Pour tenter de répondre à ce problème, nous proposons d'utiliser une approche perceptive.

3.1.1.2 Approche perceptive proposée

Nous partons du même constat que Likforman-Sulem [LSF95] : à une certaine distance, les lignes de texte apparaissent comme des segments.

Si on s'intéresse à l'image avec un point de vue global, c'est à dire à faible résolution, l'extraction des segments nous donne des informations sur les lignes de texte. Cette vision globale permet ainsi d'émettre une hypothèse sur la position, la pente, la courbure et l'épaisseur des lignes de texte. Cette hypothèse permet de définir une zone d'intérêt précise, dans laquelle on peut se focaliser à haute résolution pour valider la présence de la ligne, et en détailler sa composition comme un ensemble de composantes connexes, en tenant compte du contexte. Ce mécanisme est une application du principe de prédiction/vérification.

L'utilisation de la vision perceptive pour la reconnaissance de lignes de texte permet donc de s'affranchir des problèmes locaux de courbure, de bruits parasites et de chevauchement entre les lignes. Ceci est permis par un changement de primitive visuelle : les lignes de texte ne sont plus perçues comme des lettres successives mais comme des segments.

3.1.2 Traits

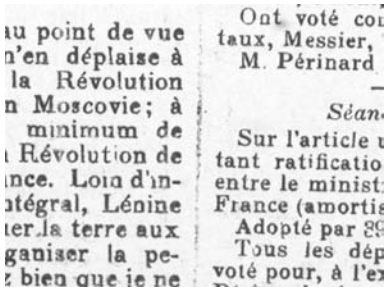
Les traits ou filets sont des éléments structurels qui peuvent servir de base pour la reconnaissance de documents fortement structurés, tels que les formulaires, les tableaux ou les pages de journaux. Ainsi, dans le concours de reconnaissance de journaux proposé lors d'ICDAR'01 [GMA01], deux des trois méthodes basent leur approche sur la reconnaissance des lignes horizontales et verticales.

Cependant, la détection des filets est particulièrement complexe dans le cas des documents anciens ou abîmés. Des exemples de cas difficiles sont présentés sur la figure 3.2. Plus généralement, de mauvaises techniques d'impression peuvent produire des lignes avec des bavures ou partiellement effacées ; de mauvaises conditions de conservation du document font apparaître des déchirures ou des tâches liées à des pliures du papier ; enfin, l'étape de numérisation introduit parfois du biais ou de la courbure dans le document.

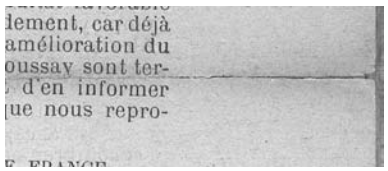
(a) Ligne épaisse mouchetée de blanc

(b) Lignes doubles qui se recouvrent

(c) Ligne discontinue



(d) Ligne fine légèrement effacée



(e) Ligne due à une déchirure du papier

FIG. 3.2 – Exemple de traits difficiles à détecter, ou à ignorer, dans des documents anciens

Nous rappelons quelques méthodes de la littérature avant d'aborder les intérêts d'une approche perceptive.

3.1.2.1 Méthodes de la littérature

Les méthodes classiques basées sur la projection ne sont pas très appropriées aux difficultés rencontrées dans les documents anciens. En effet, elles utilisent uniquement une analyse globale de l'image, et sont très sensibles au biais et à la courbure.

Par conséquent, d'autres méthodes ont été proposées. Ainsi, Gatos *et al.* proposent dans [GMC+99] une méthode spécifique basée sur l'assignation d'un poids à chaque pixel appartenant à une ligne. Un algorithme de combinaison permet ensuite de regrouper ces pixels. Cependant, cette méthode requiert des informations *a priori* sur la longueur et l'épaisseur des lignes contenues dans le document.

Hajdar *et al.* [HHI01] détectent des lignes discontinues en se basant sur un regroupement de composantes connexes. Là encore, ils ont besoin de connaître une distance maximale entre deux composantes appartenant à une même ligne.

Les travaux de Liu *et al.* [LLHY01] sont aussi basés sur un seuil de distance.

Dans toutes ces méthodes, les auteurs utilisent à bas niveau une forte connaissance *a priori* sur l'épaisseur et la longueur des lignes étudiées.

Xi *et al.* [XL05] proposent une méthode basée sur les ondelettes qui permet de faire coopérer différentes résolutions de l'image. Cependant, leur méthode n'est pas adaptée pour traiter des lignes courbes ou avec un biais de plus de 2 degrés, cas fréquent dans les documents d'archives.

Une approche qui s'apparente à la vision perceptive est proposée par Hori et Doermann [HD95], dans le contexte de l'analyse de structures de formulaires. En effet, ces auteurs sont confrontés au problème des traits cassés de manière aléatoire. Pour y remédier, ils proposent un algorithme pour créer une image réduite (sorte de sous-résolution), dans laquelle les lignes cassées sont ré-échantillonnées et forment des lignes solides. La coopération entre les deux niveaux de perception permet alors de définir les différentes cases des formulaires étudiés. Ainsi, pour le cas des filets cassés, la coopération entre plusieurs niveaux de vision semble bénéfique. Nous proposons donc d'explorer cette piste pour résoudre plus largement le problème de la reconnaissance des filets.

3.1.2.2 Approche perceptive proposée

Nous proposons d'appliquer le mécanisme de la vision perceptive pour la reconnaissance des traits. En effet, leur perception peut varier selon la résolution étudiée.

En regardant un document avec une vision globale, les traits qui ressortent sont les lignes épaisses, même si elles sont dégradées (figure 3.2(a)), et les lignes multiples (figure 3.2(b)) qui apparaissent comme un seul segment. Les filets fins sont trop peu contrastés pour être perçus à ce niveau de résolution.

En s'intéressant à ce même document à une distance intermédiaire, les éléments perçus comme des segments sont des morceaux de filets multiples (figure 3.2(b)), des

morceaux de filets fins (figures 3.2(c) et 3.2(d)), des morceaux de filets épais (figure 3.2(a)).

Enfin, en regardant le document à haute résolution, on peut percevoir des morceaux de filets multiples (figure 3.2(b)) et des morceaux de filets simples (figures 3.2(c) et 3.2(d)). Les segments épais peuvent ne pas être perçus si le bruit est trop important, par exemple dans le cas de mouchetage blanc (figure 3.2(a)).

Ces constatations nous montrent l'intérêt de combiner les visions à différentes résolutions pour reconnaître un trait avec certitude. Le principe de base est que l'étude de l'image à la résolution inférieure permet d'émettre une hypothèse sur l'existence et la nature d'un trait (position, épaisseur, courbure). Ces caractéristiques guident l'analyse à une résolution supérieure, et la présence de segments dans cette nouvelle résolution permet de confirmer l'hypothèse de la présence d'un trait. De plus, cette analyse peut être réalisée sans introduction de connaissance sur l'épaisseur ou la longueur des traits à reconnaître.

3.1.3 Bilan

Nous avons montré que, intuitivement, la vision perceptive semble apporter une aide pour la reconnaissance d'entités structurelles élémentaires telles que les lignes de texte et les traits. En effet, la vision perceptive permet un mécanisme de prédiction/vérification : des hypothèses sur la nature et la position des objets sont émises à basse résolution et confirmées à plus haute résolution. Cette approche permet notamment de gérer plus facilement les difficultés liées au bruit.

Les lignes de texte et les traits peuvent être décrits de manière indépendante du type de document étudié. Ils peuvent donc être considérés comme des éléments prégnants qui pourront servir de base à des descriptions structurelles plus complexes.

3.2 Recherche d'éléments structurels complexes

Au delà des éléments structurels simples que sont les lignes de texte et les traits, nous voulons montrer que la vision perceptive peut avoir des intérêts pour la reconnaissance de document à structure plus complexe. Nous présentons deux cas pour lesquels la vision perceptive semble intuitivement faciliter la reconnaissance.

3.2.1 Information dense : sélection

Nous étudions le cas des documents bruités puis des documents à structure complexe avant de synthétiser leur point commun.

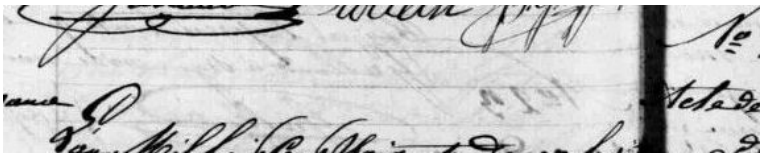
3.2.1.1 Documents bruités

Dans le contexte de l'analyse de documents anciens, manuscrits, abîmés, on rencontre de nombreux éléments qui ne sont pas prévus dans la description de la structure :

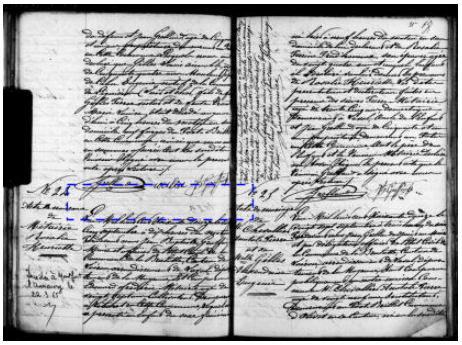
c'est ce qu'on appelle le *bruit*, que nous avons présenté en introduction de ce manuscrit. On peut citer comme exemple de bruit : les pliures du papier, les taches d'encres ou les salissures, l'encre qui transparaît depuis l'autre face du papier ... Le bruit est généralement trop variable au sein d'une collection de documents pour qu'on puisse en faire une description exhaustive ou en prévoir un traitement spécifique. Il peut donc perturber localement la reconnaissance de la structure de documents.

L'utilisation de la vision globale permet de diminuer l'influence du bruit pour repérer globalement certains indices.

Par exemple, la tache et les lignes présentes sur la figure 3.3(a) perdent de l'importance lorsqu'on s'intéresse à la vue globale du document 3.3(b).



(a) Les lignes fines et l'écriture du verso sont des exemples de bruit pouvant perturber l'analyse locale



(b) Avec une vision globale du document, le bruit (lignes fines, écriture du verso) a moins d'importance

FIG. 3.3 – Diminution de l'importance du bruit dans une vision globale

3.2.1.2 Documents à structure complexe

On nomme documents à structure complexe les documents pour lesquels de nombreuses informations structurelles sont présentes, parfois en excès par rapport à la tâche de structuration demandée.

Par exemple, si on s'intéresse au découpage de pages de journaux en colonnes, certains éléments sont superflus : le détail des lettres du titre, les variations de polices, les

détails des images (figure 3.4(a)). Par contre, si on s'intéresse à une vision globale du document, les éléments structurels sont simplifiés ce qui facilite la reconnaissance de la structure globale (figure 3.4(b)). Une fois cette structure globale extraite, on peut alors rentrer dans les détails à une vision à plus haute résolution. Pour l'exemple des pages de journaux, avoir une vision globale permet d'extraire les colonnes sans tenir compte du texte ni des images qu'elles contiennent. Ces éléments pourront être détaillés dans une phase d'analyse plus précise.

Cette idée est utilisée dans [BCLM98] où l'utilisation la multirésolution vise à réduire le coût d'analyse en diminuant les détails de l'image.

(a) Détails structurels locaux superflus pour un découpage en colonnes : alignements de pixels dans les lettres majuscules

(b) Vision globale du document : les colonnes ressortent mieux

FIG. 3.4 – Élagage des détails structurels dans la vision globale

3.2.1.3 Point commun

Dans le cas des documents bruités, comme des documents à structure complexe, l'image initiale présente une information structurelle dense : bruit ou informations superflues vis à vis de la structure à extraire.

L'utilisation de la vision perceptive permet la combinaison de deux points de vue, global et local. La vision globale permet de sélectionner plus facilement les éléments structurels pertinents. L'hypothèse émise globalement sur la structure est ensuite validée ou complétée à plus haut niveau de résolution. Cet exemple est encore une application du principe de prédiction/vérification.

3.2.2 Information diffuse : reconstitution

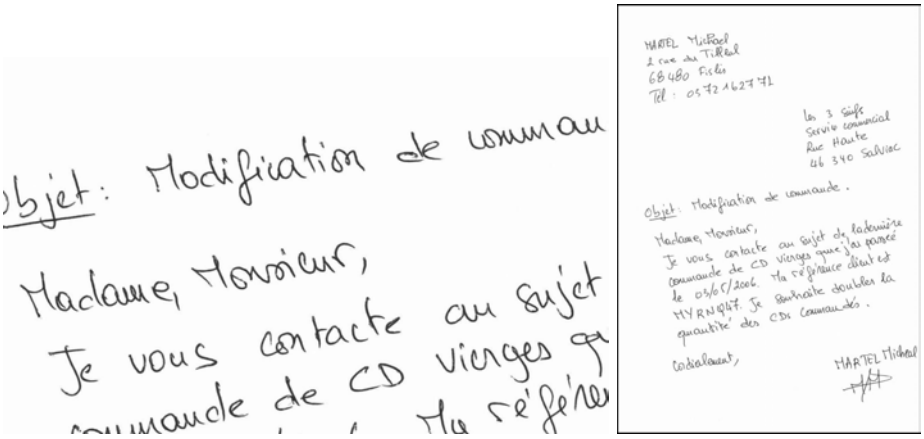
Nous étudions le cas des documents faiblement structurés puis du positionnement d'éléments structurels avant de synthétiser leur point commun.

3.2.2.1 Documents faiblement structurés

On appelle documents faiblement structurés les documents pour lesquels la structure n'est matérialisée par aucun trait ou aucun bloc de texte régulier. Si on se contente d'analyser le document à très haute résolution, il est difficile d'extraire une structure logique sans élément structurant.

Cependant, la vision humaine est capable de détecter instantanément l'organisation complète d'un tel document, à partir d'une vision globale.

Ainsi, sur la figure 3.5(a), l'extraction locale des blocs est complexe, alors qu'elle semble plus simple à réaliser globalement sur la figure 3.5(b).



(a) Vision locale

(b) Vision globale du document

FIG. 3.5 – La vision globale facilite l'extraction des blocs

Nous proposons donc d'utiliser la vision perceptive et de montrer comment la coopération entre une vision globale et une analyse plus fine des détails permet de simplifier la détection de la structure.

3.2.2.2 Positionnement d'éléments structurels

On appelle positionnement d'éléments structurels une analyse dont le but est de trouver plus précisément la position d'un élément de structuration n'ayant pas d'existence physique dans l'image initiale.

Par exemple, pour la reconnaissance d'écriture manuscrite, il est parfois plus simple de s'appuyer sur la détection de la ligne de base. Cependant, dans une image à haute résolution, cette ligne de base n'a pas d'existence physique directe et n'apparaît pas toujours de manière très claire, *a fortiori* dans l'écriture manuscrite.

Cependant, une vision globale du document permet d'extraire des caractéristiques globales sur les lignes de texte, et donc sur les lignes de base des mots. En tenant compte de cette vision globale, une analyse locale permettra de positionner de manière très précise la ligne de base repérée de loin.

La vision perceptive permet donc, en combinant les visions à différents niveaux, de positionner précisément un élément structurel n'ayant pas d'existence physique dans l'image.

3.2.2.3 Point commun

Dans le cas des documents faiblement structurés, comme pour le cas du positionnement d'éléments structurels, les informations structurelles ayant une existence physique sont diffuses.

Dans ces cas, l'utilisation de la vision perceptive permet de reconstituer des éléments structurels. En effet, l'utilisation de la vision globale permet de faire ressortir une structure, dont le positionnement peut être ensuite précisé en utilisant une vision plus locale.

Conclusion de la première partie

La vision perceptive est un mécanisme formé d'une composante physiologique, le cycle perceptif, guidé par un aspect psychologique, l'attention visuelle. L'attention visuelle peut prendre deux formes :

- elle est entraînée par des éléments prégnants qui « sautent aux yeux »,
- ou bien guidée par la recherche d'un objet précis.

Ces deux formes d'attention cohabitent dans la vision humaine.

Dans le domaine de l'analyse de documents, plusieurs approches de la littérature visent à imiter la perception humaine. Les mécanismes basés sur une attention prégnante et les théories de l'organisation perceptive sont caractérisés par le peu de connaissances *a priori* requises pour analyser un document. D'autres travaux basés sur l'attention guidée par un but permettent de reconnaître des modèles complexes, décrits par une connaissance spécifique, dans des environnements bruités.

Nous proposons de combiner ces deux approches liées aux deux formes d'attention. En effet, l'extraction d'entités structurelles élémentaires, telles que les lignes de texte ou les traits, peut se faire intuitivement, sans connaissance particulière sur le type de document. A l'opposé, pour certains documents anciens, abîmés ou bruités, il est nécessaire d'introduire des connaissances liées à un type de documents. De plus, nous proposons de séparer la connaissance spécifique de l'ensemble des traitements du système, afin d'obtenir un système générique pouvant s'adapter à des problèmes variés.

Nous avons mis en avant, de manière intuitive, les intérêts de la vision perceptive pour la reconnaissance de documents complexes. Elle semble, d'une part, faciliter la sélection des informations présentes dans un document très dense, afin d'en extraire une structure cohérente, et d'autre part permettre de reconstituer la structure à partir d'informations physiques diffuses.

Dans la suite du document, nous allons démontrer ces intuitions en proposant une méthode générique, permettant de décrire des mécanismes de coopération perceptive adaptés à chaque type de problèmes. Cette méthode sera ensuite validée dans la troisième partie par l'application à des problèmes variés, qui permettront également de quantifier de manière plus précise les apports de la vision perceptive pour la reconnaissance de documents.

Deuxième partie

Méthode perceptive DMOS-P

Introduction

La première partie de ce document a montré l'intérêt de simuler la vision perceptive dans le cas de l'analyse de la structure de documents. Nous présentons maintenant la manière dont nous avons implémenté une méthode générique s'inspirant de la vision perceptive pour la reconnaissance de documents.

Cette partie présente ainsi le cœur de notre travail et expose les concepts que nous avons mis en œuvre pour créer une méthode complète de reconnaissance de documents, à la fois souple et générique, s'appuyant sur le cycle perceptif et l'attention visuelle.

Dans le chapitre 4, nous listons les éléments nécessaires à l'implémentation de la vision perceptive, au vu de la définition présentée dans le chapitre 1. Pour répondre à ces besoins, nous proposons d'utiliser le contexte d'une méthode existante : DMOS (Description et MOdification de la Segmentation), que nous présentons avec un éclairage perceptif dans le chapitre 5. En gardant la philosophie de cette méthode, nous en créons une nouvelle version, basée sur l'introduction de nouveaux outils et formalismes liés à la multirésolution (chapitre 6). Nous abordons ensuite la gestion des objets structurels prégnants (chapitre 7).

La combinaison de tous ces apports mène à l'obtention d'une nouvelle méthode, DMOS-P, qui permet de spécifier simplement des mécanismes de coopération perceptive, intégrant une modélisation de connaissances décrivant n'importe quel type de documents, et adaptés à chaque type de problème.

Les concepts présentés dans cette partie s'appuient sur quelques exemples illustratifs. Nous présenterons dans la troisième partie des exemples d'applications réelles qui permettront de valider les choix d'implémentation réalisés, ainsi que d'évaluer, de manière quantitative, les apports de la vision perceptive pour la reconnaissance de documents.

Chapitre 4

Éléments requis pour un système de vision perceptive

Dans ce chapitre, nous identifions les éléments clés qui sont requis pour réaliser un mécanisme analogue à la vision perceptive. Nous exposons ensuite la solution que nous proposons afin d'implémenter tous ces éléments.

4.1 Éléments requis

Chercher à implémenter un système de vision perceptive revient à imiter d'une part le cycle perceptif, et d'autre part les deux formes d'attention visuelle qui guident ce cycle. Pour chacun de ces deux aspects, nous mettons donc en avant les différents besoins. Enfin, nous insistons sur l'aspect générique nécessaire pour notre méthode.

4.1.1 Imiter le cycle perceptif

Le cycle perceptif de la vision humaine, déjà présenté dans le chapitre 1, est rappelé sur la figure 4.1. Sur la figure 4.2, nous replaçons le cycle perceptif de la vision dans le cadre de la reconnaissance de documents, en adaptant le vocabulaire utilisé.

Dans la suite du document nous définissons un *point de vue* comme l'association
- d'un niveau de perception (une résolution de l'image, par exemple),
- et de la localisation spatiale d'une zone d'intérêt dans ce niveau de perception.

Dans le cycle perceptif, la capture d'informations sur la rétine se traduit donc par la sélection d'un point de vue dans l'image de document.

La mise en correspondance des primitives avec un modèle en mémoire correspond, dans le cas de l'analyse de documents, à une adéquation des primitives avec un processus de reconnaissance du document.

Nous complétons le schéma 4.2 afin de mettre en avant les différents éléments à implémenter pour simuler ce cycle perceptif (figure 4.3). Ces points sont détaillés dans les paragraphes suivants.

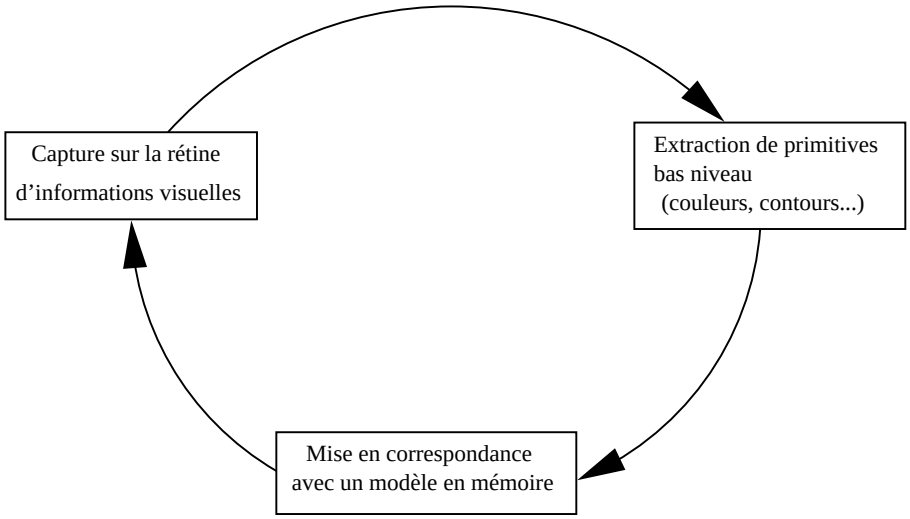


FIG. 4.1 – Le cycle perceptif de la vision humaine

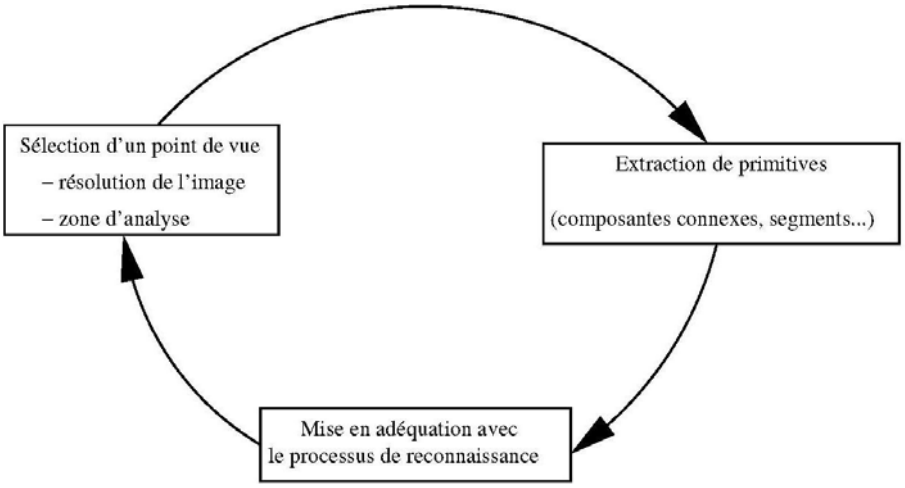


FIG. 4.2 – Le cycle perceptif appliqué à la reconnaissance de documents

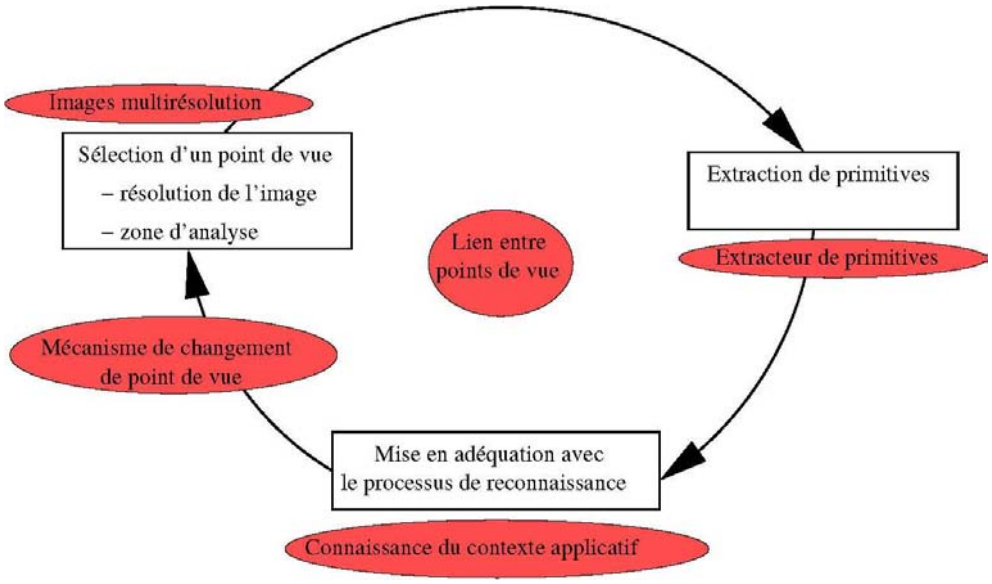


FIG. 4.3 – Éléments requis pour l'implémentation du cycle perceptif

4.1.1.1 Images multirésolutions

Le point fort de la vision perceptive est la possibilité de combiner une vision à différents niveaux de détail d'une même image. Nous proposons de simuler ceci en utilisant des images construites à plusieurs résolutions, à partir de l'image à résolution initiale. La construction d'images à des résolutions plus faibles doit se faire en imitant au mieux la sensation que l'œil humain aurait en augmentant sa distance à l'image étudiée. Les différentes résolutions étant construites à partir d'une seule image initiale, il existe un lien direct entre la position des pixels appartenant aux différentes résolutions.

Selon le type de documents étudié, les résolutions nécessaires pourront différer. L'implémentation retenue devra donc se baser sur une pyramide d'images multirésolutions, adaptable à chaque problématique.

4.1.1.2 Extraction de primitives

Lorsqu'une image est extraite au niveau de la rétine, le cerveau humain est capable d'en extraire automatiquement des primitives. Nous choisissons de nous en intéresser à deux types :

- les composantes connexes : les ensembles de pixels noirs qui se touchent,
- les segments : les ensembles de pixels noirs formant un alignement long et fin.

En effet, pour l'analyse de documents, ces deux types de primitives permettent de représenter la plupart des informations à analyser. Notre implémentation devra donc permettre d'extraire, pour chacune des résolutions étudiées, les composantes connexes et les segments.

4.1.1.3 Connaissance du contexte applicatif

La vision perceptive est guidée par des modèles appris, présents en mémoire, qui mettent en relation les différentes primitives perçues pour en former un objet interprété.

Nous avons vu dans la partie 1.1.3 que les informations stockées en mémoire sont de deux types : le *quoi* qui correspond à la nature de l'objet à reconnaître, et le *où* qui correspond au contexte de l'objet. Ce contexte est défini à la fois par une localisation spatiale dans une image, et par les notions de voisinage avec d'autres entités. D'autre part, pour un même objet, on dispose de multiples représentations internes, variant selon le niveau de détail.

Si on se réfère aux travaux de Rybak [RGG+05], décrits dans la partie 1.2.2, on mémorise aussi, pour chaque objet, un processus d'analyse qui décrit les mouvements de l'œil à effectuer pour reconnaître le dit objet. On stocke ainsi les déplacements successifs du point d'attention, et la focalisation selon le contexte sur des zones spécifiques, à la recherche de caractéristiques précises.

Dans le cas de l'analyse de structure de documents, il va donc falloir fournir, pour chaque type d'application, une description de la structure à reconnaître. Cette description devra être constituée des caractéristiques physiques des éléments de la structure, mais aussi de leurs positions respectives dans le document, et du contexte avec d'éventuels objets voisins pouvant aider à localiser cette structure et à lever les ambiguïtés. Enfin, il faudra élaborer une stratégie d'analyse propre à chaque type de documents : déplacements entre zones d'intérêt et focalisation d'attention selon les éléments reconnus successivement.

Une approche grammaticale semble par exemple bien adaptée pour décrire à la fois le contenu, la localisation spatiale et l'interaction avec le voisinage des modèles à reconnaître.

4.1.1.4 Changement de point de vue

Lorsqu'un modèle est partiellement reconnu en mémoire, l'aspect cyclique de la vision perceptive permet de modifier le point de vue de l'analyse pour en extraire de nouvelles informations. Ce changement de point de vue correspond à une adaptation de la résolution étudiée ainsi que du contexte spatial. Cela doit se traduire au niveau de l'implémentation par :

- la possibilité de changer de résolution au cours de l'analyse, pour simuler le changement de niveau de perception ;
- le transfert du contexte spatial dans la nouvelle résolution pour conserver le choix d'une zone d'intérêt particulière.

Il faut donc prévoir de poursuivre l'analyse à une autre résolution, tout en gérant un transfert de zones de recherche.

4.1.1.5 Transfert d'information

L'intérêt de la vision perceptive est de percevoir les choses différemment selon la résolution de l'image étudiée. Fréquemment, la vision à une résolution `A` sert à émettre une hypothèse sur la présence d'éléments, qui est confirmée ou infirmée selon le résultat de l'analyse d'une résolution `B`. Il faut donc être capable de transmettre des hypothèses sur la localisation d'éléments d'une résolution à une autre. Cela peut se traduire par le calcul de seuils, de longueurs, de zones de recherches ...

Mais de manière plus forte, il est parfois nécessaire de mettre en correspondance des éléments vus dans une résolution `A` avec leur équivalent dans une résolution `B`. Par exemple, nous avons vu dans la partie 3.1.1 qu'une ligne de texte peut être vue comme un segment à résolution faible. Il faut parfois retrouver, lors du passage vers une résolution supérieure, les pixels de la ligne de texte qui ont servi à sa création. Il est donc nécessaire de prévoir un opérateur qui réalisera cette mise en correspondance.

Plus généralement, il faut prévoir la possibilité d'utiliser des représentations des segments et des composantes connexes indépendamment de leurs résolutions d'origine.

4.1.2 Imiter l'attention visuelle

Nous avons mis en évidence (partie 1.2) l'intérêt d'utiliser conjointement les deux formes d'attention visuelle qui guident le fonctionnement du cycle perceptif. En effet, l'attention guidée par des éléments prégnants permet d'extraire des éléments tels que les lignes de textes, sans connaissance *a priori* sur le document, en appliquant simplement les lois de proximités de l'organisation perceptive. A l'opposé, l'attention guidée par un but permet de rechercher un objet complexe dans un environnement bruité.

Pour faire collaborer ces deux formes d'attention, nous souhaitons avoir la possibilité d'extraire, dans une zone donnée, un ensemble d'objets structurels prégnants, qui pourront alors servir dans un second temps à la recherche d'objets complexes décrits par une connaissance spécifique. De cette manière, on calcule localement dans un premier temps les objets qui « sautent aux yeux », et qui ne demandent pas de connaissance spécifique au type de document, avant de les utiliser en étant guidés par un but précis.

4.1.3 Respecter la généricité

Nous avons montré dans le chapitre 3 que la vision perceptive pouvait répondre à plusieurs types de problèmes, eux même associés à plusieurs types de documents. Notre implémentation devra donc être générique pour laisser la possibilité d'utiliser l'approche multirésolution pour des types variés de documents. Pour cela, la connaissance associée à chaque type de document devra être séparée du système propre à la vision perceptive.

4.1.4 Bilan

Nous synthétisons les différents éléments nécessaires pour l'implémentation de la vision perceptive :

- pour imiter le cycle perceptif :
  - une analyse basée sur des images multirésolutions,
  - l'extraction de primitives,
  - la connaissance d'un contexte applicatif pour simuler la représentation du processus de reconnaissance ; c'est le cœur du mécanisme perceptif,
  - le mécanisme de changement de point de vue, comprenant le choix
    - du niveau de perception
    - du contexte spatial
  - la possibilité de mettre en correspondance des éléments issus de différentes résolutions ;
- pour imiter l'attention visuelle :
  - la possibilité d'extraire des éléments prégnants,
  - la possibilité de décrire des éléments complexes en étant guidé par un but ;
- pour respecter la généricité :
  - un cadre générique pour permettre l'application à de nombreux problèmes.

4.2 Solution proposée

Afin de répondre aux contraintes posées pour créer un système analogue à la vision perceptive, nous proposons de créer une nouvelle version d'une méthode existante, en l'enrichissant selon deux aspects : la multirésolution et l'attention prégnante. Nous détaillons le choix d'implémentation qui est récapitulé dans le tableau 4.1.

4.2.1 Méthode DMOS

Nous proposons de travailler dans le contexte de la méthode DMOS (Description et MOdification de la Segmentation), développée au sein de l'équipe IMADOC par Coüasnon [Coü01]. Cette méthode générique permet la reconnaissance de documents structurés. Elle est basée sur le langage grammatical EPF (Enhanced Position Formalism) qui permet d'effectuer une description bidimensionnelle de la position des éléments structurels présents dans un document. Ces éléments structurels sont des segments et des composantes connexes. Une fois la description réalisée dans le langage EPF, pour un type de document donné, l'analyseur associé est produit automatiquement par compilation.

Bien que n'ayant pas été initialement créée dans une optique perceptive, la méthode DMOS permet de répondre à plusieurs des critères listés précédemment, pour l'implémentation de la vision perceptive. En effet, le langage EPF est un outil qui permet d'exprimer simplement la description de chaque type de documents, et donc d'introduire une connaissance du contexte applicatif et des stratégies mises en œuvre pour la recherche de la structure. Cette analyse est basée sur l'extraction de primitives que sont

Besoins	Existant	Apports	
	Méthode DMOS	Multirésolution	Objets prégnants
Imiter le cycle perceptif			
Images multirésolutions		x	
Extraction de primitives	x		
Contexte applicatif	x		
Changement de point de vue - niveau de perception - contexte spatial	x	x	
Transfert d'informations		x	
Imiter l'attention visuelle			
Attention guidée par un but	x		
Attention prégnante			x
Respecter la généricité			
Aspect générique	x		

TAB. 4.1 – Réponse aux besoins de la vision perceptive par l'utilisation conjointe de la méthode DMOS, et de nouveaux principes pour la gestion de la multirésolution et des objets prégnants : l'ensemble forme la méthode DMOS-P

les composantes connexes et les segments. Les primitives sont positionnées les unes par rapport aux autres grâce à la définition d'un contexte spatial.

De plus, la description d'une structure précise à reconnaître correspond à l'imitation de l'attention guidée par un but précis : on ne décrit que les parties du document qui servent à reconnaître cet objectif.

Enfin, un des avantages majeurs de la méthode DMOS est que la connaissance associée à un type de document est séparée du système, ce qui assure sa généricité. Ainsi cette méthode a été appliquée à des types de documents très variés : tableaux, formulaires, partitions musicales, documents d'archives. En utilisant la méthode DMOS, nous répondons donc au critère de généricité requis pour l'implémentation de la vision perceptive.

Nous décrirons la méthode DMOS plus en détails dans le chapitre 5, en lui donnant l'éclairage perceptif nécessaire pour notre approche.

4.2.2 Gestion de la multirésolution

Notre système perceptif doit combiner différents niveaux de perception. L'analyse doit donc pouvoir être basée sur des images multirésolutions, avec la possibilité de changer de niveau de perception selon les besoins, et de transférer des informations d'une résolution à une autre.

Pour répondre à ces besoins, nous proposons donc d'introduire dans la méthode DMOS et le langage EPF de nouveaux outils et formalismes qui seront décrits plus en détail dans le chapitre 6.

4.2.3 Gestion des objets prégnants

Nous souhaitons utiliser avec notre méthode les deux formes d'attention visuelle. La possibilité de reconnaître un objet en étant guidé par un but étant déjà offerte par la méthode DMOS, il faut prévoir la gestion d'objets reconnus par une attention guidée par la prégnance.

Nous proposons donc d'introduire également dans la méthode DMOS, une architecture spécifique pour la construction et l'exploitation d'objets prégnants. Dans ce cadre, nous proposons dans un premier temps deux types d'objets prégnants : les lignes de texte et les traits, dont la construction ne nécessite pas de connaissance spécifique à un type de documents, et qui peuvent être réutilisés pour la reconnaissance de structures plus complexes. L'architecture gérant ces objets sera décrite en détail dans le chapitre 7.

4.2.4 Bilan

En résumé, notre objectif est de conserver les bonnes propriétés fournies par la méthode DMOS, et de créer une nouvelle version enrichie par l'introduction de la multirésolution et d'objets prégnants. Ceci permet de produire un nouveau système complet et générique de vision perceptive, DMOS-P, qui respecte les différents besoins récapitulés dans le tableau 4.1.

Chapitre 5

Méthode DMOS existante

La méthode DMOS (Description et MOdification de la Segmentation) est une méthode générique de reconnaissance de documents structurés, développée dans l'équipe Imadoc par Coüasnon [Coü01] [Coü06].

Nous rappelons ici le principe de fonctionnement de cette méthode, en orientant notre description selon les besoins pour la réalisation d'un système perceptif.

5.1 Architecture globale

L'architecture globale de la méthode DMOS est présentée sur la figure 5.1.

La méthode est basée sur le formalisme grammatical EPF (Enhanced Position Formalism) qui permet d'effectuer une description graphique, syntaxique et sémantique d'un type de documents. Cette description forme le niveau symbolique de la méthode, et contient la connaissance spécifique à chaque type de documents.

La description symbolique est basée sur un niveau numérique. En effet, la grammaire définie en EPF utilise pour terminaux des primitives extraites directement dans l'image : les segments et les composantes connexes. Une fois la description réalisée dans le langage EPF, l'analyseur associé est produit automatiquement par une étape de compilation.

La particularité de cette méthode est donc de séparer la connaissance liée à chaque type de document, du noyau. La généricité de cette méthode a été validée sur de nombreux types de documents : partitions musicales, tableaux, formulaires, documents d'archives, et à grande échelle, sur plus de 500 000 documents.

Nous présentons plus en détail l'extraction des terminaux de la grammaire, le langage EPF et les propriétés de l'analyseur.

5.2 Extraction des terminaux

Pour chaque image analysée, il est possible d'extraire une liste de composantes connexes, une liste de segments horizontaux et une liste de segments verticaux. Ces primitives sont obtenues grâce à des extracteurs spécifiques qui se basent sur une image

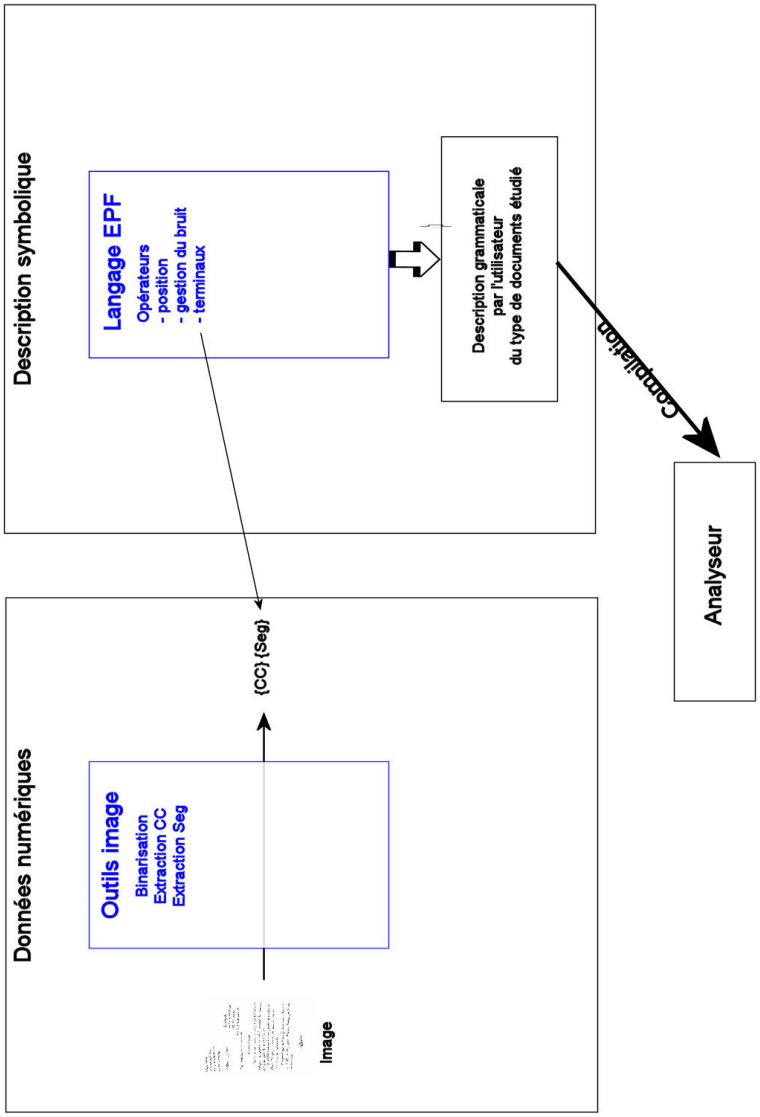
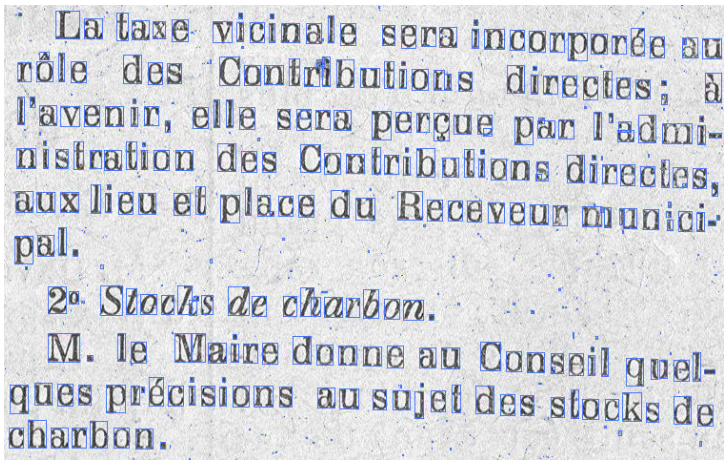


FIG. 5.1 – Méthode DMOS initiale

en niveau de gris binarisée[1].

5.2.1 Composantes connexes

Les composantes connexes sont représentées par les coordonnées de leur rectangle englobant. La figure 5.2 présente ainsi des exemples de composantes connexes extraites d'une page de journal.


La taxe vicinale sera incorporée au rôle des Contributions directes; à l'avenir, elle sera perçue par l'administration des Contributions directes, aux lieu et place du Receveur municipal.

2e. *Stocks de charbon.*

M. le Maire donne au Conseil quelques précisions au sujet des stocks de charbon.

FIG. 5.2 – Exemples de composantes connexes extraites dans une image de page de journal

5.2.2 Segments

Les segments sont représentés par les coordonnées de leurs extrémités. La détection des segments est réalisée grâce à un extracteur développé dans l'équipe Imadoc, présenté dans [LCQ95], et basé sur un filtre de Kalman. Cette méthode est décrite en détails dans l'annexe A.

Un *segment* est défini comme un alignement long et fin de pixels. On nomme *empan* un ensemble de pixels noirs qui se touchent, dans une même colonne, c'est-à-dire dans la direction orthogonale à celle du segment. Un segment idéal peut être défini comme une succession continue d'empans ayant la même épaisseur, et dont les points milieux sont alignés (figure 5.3).

Dans les cas réels, les segments ne sont pas toujours aussi nettement marqués. Notre méthode présente donc plusieurs propriétés permettant de faire face aux problèmes rencontrés. Ces propriétés sont les suivantes :

1. existence possible de discontinuités : il est utile de permettre localement une absence de points, due à la qualité ou à la nature des objets extraits (ligne pointillée, défaut de binarisation, bruit) (figure 5.4(a)) ;

[1] L'étape de binarisation est réalisée grâce à une méthode existant dans l'équipe, qui ne sera pas décrite ici.

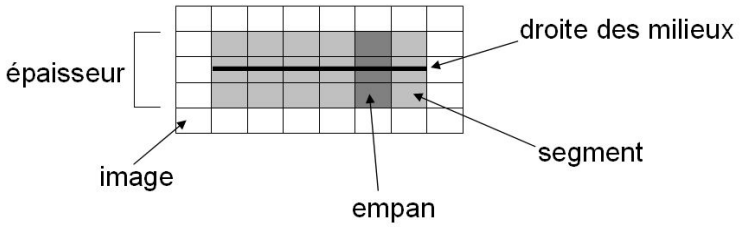
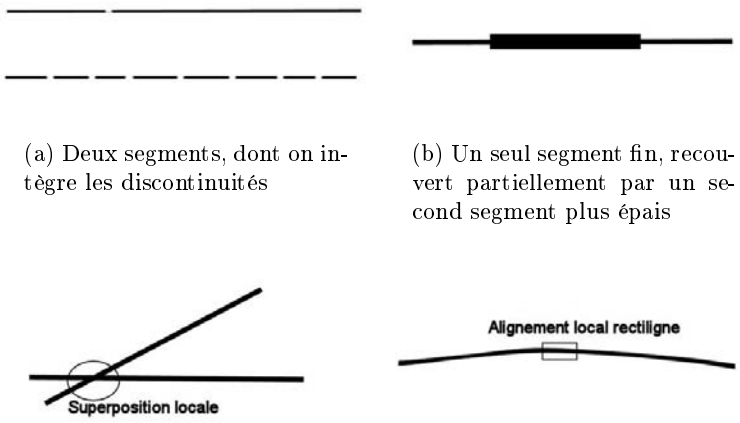


FIG. 5.3 – Exemple de segment et des empans qui le constituent

2. prise en compte de l'épaisseur pour chaque point représentatif (figure 5.4(b)) ;
3. gestion de la taille variable des segments, allant de quelques pixels à plusieurs centaines ;
4. prise en compte de segments qui se croisent (figure 5.4(c)) ;
5. prise en compte de la courbure dans un segment (figure 5.4(d)) ;
6. prise en compte du biais.



(a) Deux segments, dont on intègre les discontinuités

(b) Un seul segment fin, recouvert partiellement par un second segment plus épais

(c) Deux segments qui se croisent

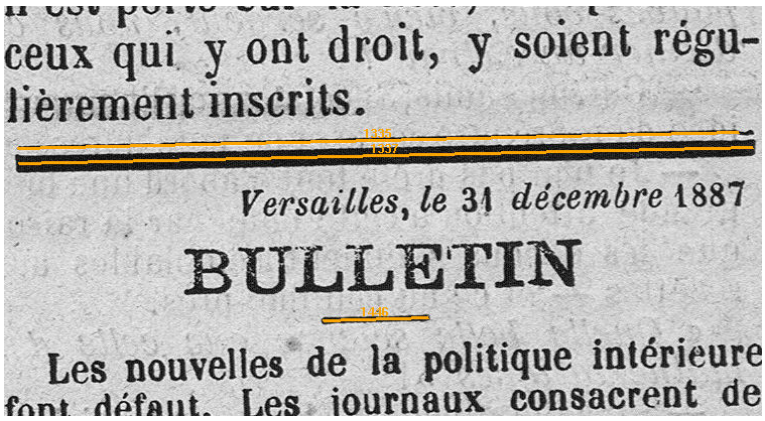
(d) Un segment globalement courbe

FIG. 5.4 – Exemples d'interprétations réalisées par l'extracteur de segments utilisé

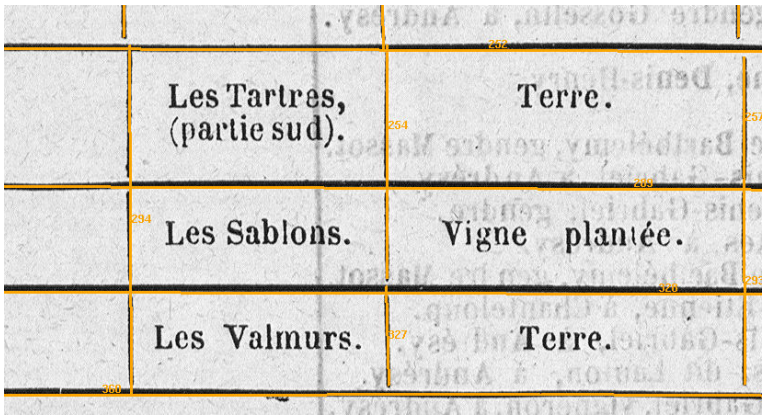
Des exemples de segments reconnus sont présentés sur la figure 5.5.

5.3 Langage de description EPF

Le langage EPF est un langage grammatical bidimensionnel permettant d'exprimer une description graphique, syntaxique et sémantique d'un document.



(a) Epaisseur et longueur variables



(b) Gestion des croisements et de la légère courbure

FIG. 5.5 – Exemples de segments reconnus sur des images de journaux

Les composantes connexes et les segments extraits précédemment servent de terminaux pour la description grammaticale dans le langage EPF. Les éléments du langage EPF sont présentés en détail dans [Coü06]. Nous en rappelons ici les points principaux.

Nous présentons tout d'abord un exemple de description dans le langage EPF. Supposons que l'on s'intéresse à la description de la structure de courriers manuscrits comme celui présenté sur la figure 5.6.

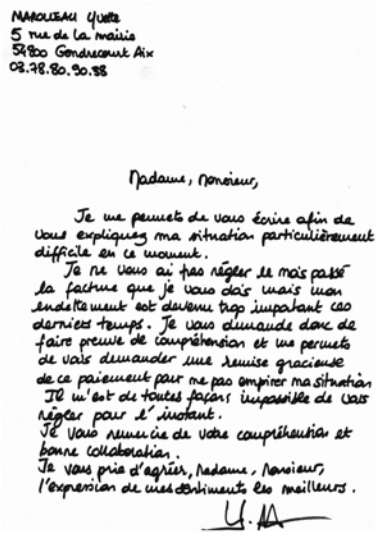


FIG. 5.6 – Exemple de document dont on effectue la description en EPF

Il faut pouvoir exprimer qu'une page de courrier est constituée de quatre éléments : des coordonnées expéditeur, une ouverture, un corps de texte et une signature, qui sont disposés à des positions relatives particulières. La description d'un tel document est donc la suivante :

```
pageDeCourrier ::
      AT_ABS(hautGauchePage)&&
      coordonneesExpediteur &&
      AT_ABS(milieuPage) &&
      ouverture Ouv &&
      AT(sousOuverture Ouv) &&
      corpsDeTexte C &&
      AT(sousCorpsDeTexte C) &&
      signature.
```

Cette règle permet de reconnaître une `pageDeCourrier` comme étant constituée de `coordonneesExpediteur`, d'une `ouverture`, d'un `corpsDeTexte` et d'une `signature`. Ces éléments sont organisés selon un positionnement précis, décrit grâce à l'opérateur `AT`, dont nous détaillerons la syntaxe ci-dessous. Le symbole `&&` est l'opérateur de conca-

ténation. Les attributs utilisés pour les règles commencent par une majuscule et peuvent être, selon les cas, synthétisés ou hérités.

Chaque non terminal tel que `coordonneesExpediteur` ou `ouverture` est détaillé dans une sous-règle spécifique. Le plus bas niveau de description est constitué de la reconnaissance des terminaux : composantes connexes et segments.

Nous détaillons maintenant les principaux opérateurs du formalisme EPF.

5.3.1 Opérateurs de position

Les opérateurs de position sont à la base de la méthode d'analyse. En effet, habituellement, lors d'une analyse grammaticale, on ne se pose pas la question du prochain terminal à reconnaître : les terminaux s'enchaînent sous la tête de lecture (cas d'une chaîne de caractères, par exemple). Dans le cas d'un document en deux dimensions, il faut savoir à quel endroit aller chercher le prochain terminal à reconnaître. Les opérateurs de positions ont donc pour rôle d'indiquer où trouver le prochain symbole.

La syntaxe de ces opérateurs de position est la suivante :

```
AT(position)
AT_ABS(position)
```

L'opérateur `AT` est utilisé pour un positionnement relatif, tandis que `AT_ABS` est utilisé pour un positionnement absolu. L'exemple présenté ci-dessus montre un cas d'utilisation :

```
ouverture Ouv &&
AT(sousOuverture Ouv) &&
corpsDeTexte C &&
```

qui se traduit par : partant de l'ouverture `Ouv`, le non-terminal de référence synthétisé par `ouverture`, on va chercher le corps de texte `C` dans la *zone* définie par `sousOuverture`, en fonction de `Ouv`. Cette zone est un polygone, dans lequel un point d'ancrage fixe l'ordre de parcours des éléments. Les éléments `Ouv` et `C` peuvent être des terminaux ou des non-terminaux. On peut définir dans la grammaire autant de positionnements (tels que `sousOuverture`) que nécessaire.

5.3.2 Détection des terminaux

Les terminaux d'un document sont extraits tel que décrit dans la partie 5.2, et peuvent être de deux types :

- composantes connexes, représentées par leur rectangle englobant,
- segments à tendance horizontale ou verticale, représentés par leurs extrémités.

Les opérateurs `TERM_CMP` et `TERM_SEG` sont utilisés pour la reconnaissance, respectivement, des composantes connexes et des segments, dans la zone de recherche fixée. Leur syntaxe est la suivante :

```
TERM_CMP PreCondition PostCondition Etiquette ComposanteReconnue
TERM_SEG PreCondition PostCondition Etiquette SegmentReconnu
```

Lors de la recherche de la composante dans une zone, on peut vouloir ne pas prendre la première trouvée, mais celle respectant une certaine condition. Par exemple, on peut rechercher une composante connexe qui soit assez grande, ou un segment qui soit horizontal. Ce souhait est exprimé grâce à `PreCondition`. Lorsque la composante est trouvée, la `PostCondition` permet de vérifier un critère d'acceptation. Ces conditions permettent de gérer d'éventuels éléments de bruit, c'est-à-dire non prévus par la description, en n'en tenant pas compte s'ils ne vérifient pas les conditions.

L'`Etiquette` est le nom qui sera donné à l'élément reconnu.

5.3.3 Opérateur IN DO

Cet opérateur permet de réduire la zone d'application d'une règle de la grammaire. Il est formulé de la manière suivante :

```
IN(zone) DO(regle)
```

Cela signifie que la règle `regle` ne doit s'appliquer que dans la zone `zone`. Cet opérateur est très utile notamment pour faire des définitions récursives et limiter les recherches dans une sous-partie du document : une case d'un tableau, par exemple.

5.3.4 Opérateur FIND

Dans une analyse classique, l'analyseur essaie d'appliquer chaque règle sur l'élément courant, jusqu'à ce qu'une des règles s'applique. L'opérateur `FIND` permet de gérer les problèmes de bruit en modifiant l'ordre d'analyse : on essaie de faire réussir une règle donnée sur tous les éléments de la structure, jusqu'à ce qu'un élément convienne, où jusqu'à une condition d'arrêt.

La syntaxe de cet opérateur est la suivante :

```
FIND(regle) UNTIL(conditionArret)
```

Cet opérateur essaie de faire réussir le (non-)terminal `regle`, sur tous les éléments contenus dans la zone d'analyse, tant que la condition `conditionArret` n'est pas vérifiée.

5.4 Propriétés de l'analyseur

Le langage EPF permet d'exprimer une description de la structure d'un type de documents. A partir de cette description, une étape de compilation produit automatiquement l'analyseur associé au type de document. Cet analyseur est notamment capable de modifier la structure analysée en cours d'analyse, mais aussi de gérer le bruit.

Nous détaillons deux points de son fonctionnement interne, utiles pour la suite : le mécanisme de la structure analysée et la gestion de la combinatoire.

5.4.1 Structure analysée

Le fonctionnement interne de l'analyseur se base sur l'analyse d'un couple (structure en entrée, curseur) qui représentent respectivement l'image et la position dans l'image.

5.4.1.1 Structure

La structure en entrée de l'analyseur permet de faire le lien entre les données numériques extraites de l'image et la description grammaticale. Elle contient donc :

- un pointeur vers l'image source,
- la liste des composantes connexes, extraites dans l'image, servant de terminaux pour la grammaire,
- la liste des segments horizontaux, extraits dans l'image, servant de terminaux pour la grammaire,
- la liste des segments verticaux, extraits dans l'image, servant de terminaux pour la grammaire.

Ces trois listes sont initialisées selon les besoins de l'analyse, et peuvent être vides. Au fur et à mesure que les terminaux sont reconnus dans l'analyse grammaticale, ils sont retirés de la structure.

5.4.1.2 Curseur

Le curseur modélise la position de la tête de lecture pour l'analyseur grammatical. Le curseur est donc composé de :

- la position courante dans l'image, ou point d'ancrage, définie en coordonnées pixel,
- l'élément en cours d'analyse (segment ou composante connexe),
- une zone de recherche polygonale, mise à jour par les opérateurs de position (présentés dans la partie 5.3.1).

5.4.2 Gestion de la combinatoire

L'implémentation du noyau de la méthode DMOS est réalisée en λProlog, et compilé grâce au compilateur PM (Prolog Mali) développé à l'IRISA [Bri92].

L'utilisation de la programmation logique permet d'assurer la gestion de la combinatoire et du retour arrière. En effet, durant l'analyse, les différentes règles vont pouvoir être essayées successivement. Les mécanismes de coupure permettent de gérer les essais successifs.

Ainsi, l'analyseur LL(k) donne une solution seulement s'il est possible de trouver une solution globale.

5.5 Bilan

La méthode DMOS fournit un contexte favorable pour la mise en place d'un système perceptif, selon les besoins définis dans le chapitre précédent.

En effet, cette méthode fournit les éléments du cycle perceptif suivants :

- l'extraction de primitives : les composantes connexes et les segments, qui servent de base à la reconnaissance puisqu'ils forment les terminaux de la grammaire
- la description d'un contexte applicatif grâce au langage EPF, qui est apte à piloter un mécanisme cyclique,

– la définition d'un contexte spatial à chaque étape de l'analyse, grâce à l'opérateur `AT`.

De plus, cette méthode imite déjà l'attention visuelle guidée par un but. En effet, les grammaires décrivent une structuration précise à reconnaître dans chaque type de documents, et l'analyseur permet de prendre en compte le bruit rencontré au fur et à mesure de l'analyse, par exemple grâce à l'opérateur `FIND`.

Enfin, travailler dans le contexte de cette méthode nous permet d'assurer la généricité de notre approche qui pourra être appliquée à des types de documents variés.

Chapitre 6

Gestion de la multirésolution

Pour créer la méthode perceptive DMOS-P, nous avons introduit la multirésolution dans la méthode DMOS. Nous avons ainsi créé les éléments suivants :

- une représentation multirésolution des images et la gestion des primitives associées, grâce au formalisme de calque perceptif,
- un opérateur de changement de niveau de perception,
- des outils de mise en correspondance des éléments entre niveaux de perception.

6.1 Images et données multirésolutions

Nous avons montré que l'analyse de la méthode DMOS se base sur une image dans laquelle on extrait des ensembles de composantes connexes et de segments. Afin de pouvoir simuler les différents niveaux de vision de l'œil humain, nous introduisons la possibilité de baser l'analyse sur `n` images à des résolutions différentes.

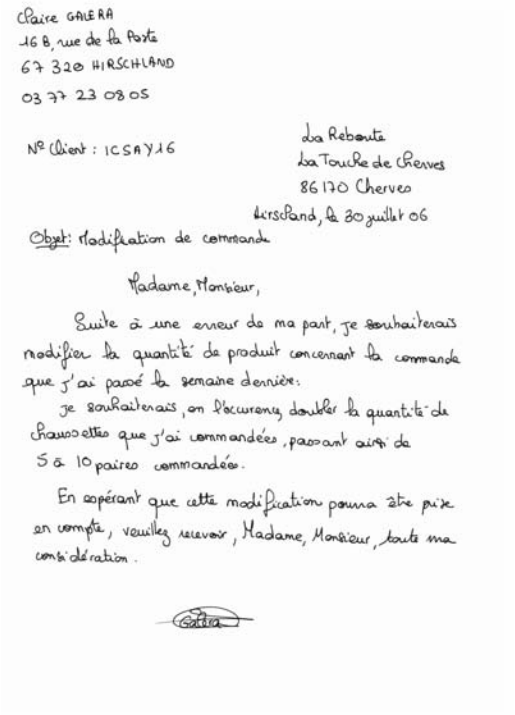
Nous présentons la construction de la pyramide d'images multirésolutions, avant d'exposer le principe d'organisation des données extraites aux différentes résolutions.

6.1.1 Pyramide d'images

La pyramide d'images est construite par analogie à la vision humaine qui permet différents niveaux de détail. Nous construisons dans ce but une pyramide d'images sous-échantillonnées. Afin de respecter le théorème de Shannon, le sous-échantillonnage est réalisé après un filtre passe-bas (de dimension 3x3).

Chaque image issue de l'application du filtre a des dimensions deux fois plus petites que l'image initiale. En appliquant le filtre de manière récursive, on parvient à la production d'une pyramide, telle que celle présentée sur la figure 6.1. On nomme résolution `-n` l'image dont les dimensions ont été divisées par `n`.

Ces résolutions ne sont pas toutes utiles pour un type de document donné. Pour chaque problème étudié, il est nécessaire de sélectionner les niveaux de résolution appropriés, en fonction des éléments perçus à chaque résolution. En revanche, une analyse empirique montre qu'au sein d'une même collection de documents, la nature des éléments perçus (composantes connexes, segments) à une résolution donnée est stable.



(a) Résolution 1 (initiale)

(b) Résolution -2

(c) Résolution -4

(d) Résolution -8

(e) Résolution -16

FIG. 6.1 – Pyramide d'images multirésolutions construite par application récursive d'un filtre passe-bas à partir de l'image initiale.

De plus, nous avons remarqué expérimentalement qu'un pas de 4 entre les résolutions semble généralement approprié : il permet de percevoir les choses différemment tout en conservant un lien suffisant entre les résolutions. Ainsi, plusieurs de nos applications présentées dans la troisième partie se basent sur trois résolutions d'images : `1`, `-4` et `-16`. Toutefois, nous insistons sur le fait que le choix des résolutions utilisées est laissé au concepteur de chaque grammaire.

Afin de faciliter la synthèse visuelle des informations, nous utilisons un outil de visualisation des images qui permet de choisir la résolution d'affichage, indépendamment des résolutions utilisées pour les traitements. Les données extraites ou calculées dans les différentes résolutions sont transférées depuis leur résolution d'origine pour être visualisées de manière cohérente. Dans la suite du document, les images proposées pourront ainsi présenter dans un même référentiel des traitements effectués à des résolutions différentes.

6.1.2 Données multirésolutions

Pour chaque image de la pyramide, il est possible d'appliquer les extracteurs de segments et de composantes connexes présentés dans la partie 5.2. On dispose donc, pour chaque résolution, des primitives associées, exprimées dans le repère de coordonnées de l'image dont elles ont été extraites.

On souhaite pouvoir utiliser toutes ces primitives de manière homogène, comme terminaux de la grammaire EPF. De plus, il est important de conserver la connaissance de la résolution d'origine de chaque ensemble de primitives.

Afin d'organiser toutes les données issues des différents niveaux de résolutions, nous introduisons le concept de *calque perceptif*. Nous en donnons une définition avant de détailler son utilisation puis son implémentation dans la méthode DMOS.

6.1.2.1 Formalisme du calque perceptif

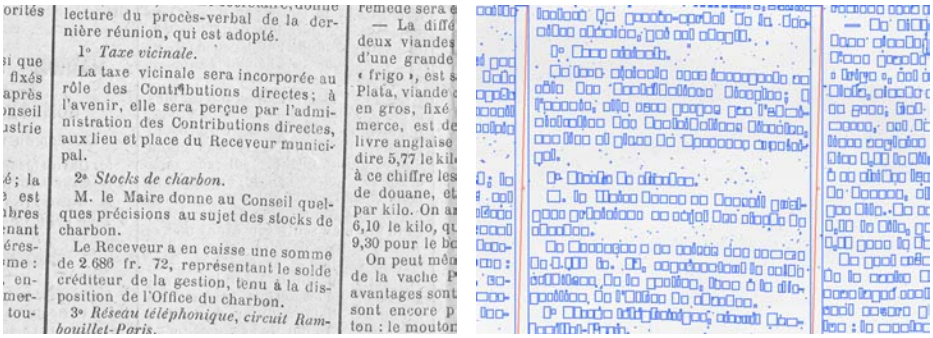
Les calques perceptifs ont pour but d'homogénéiser le traitement des données extraites aux différentes résolutions, tout en conservant le lien avec l'image d'origine de ces données.

Nous définissons le calque perceptif comme ci-dessous :

Calque perceptif Localisation spatiale d'un ensemble d'objets perçus dans l'image, qui sont les terminaux du langage EPF. Chaque objet localisé dans le calque a un type associé.

Chaque calque correspond donc à un niveau de perception particulier de l'image et contient la représentation des éléments perçus dans l'image, tels que des composantes connexes représentées par leurs rectangles englobants, des segments représentés par leurs extrémités. La figure 6.2 représente un exemple de calque perceptif contenant les composantes connexes et les segments extraits dans l'image associée.

Nous définissons deux types de calques perceptifs.



(a) Image initiale

(b) Calque perceptif contenant les composantes connexes (en bleu) et les segments (en rouge) extraits dans l'image

FIG. 6.2 – Exemple de calque perceptif direct

Calque perceptif direct Calque perceptif dont les données sont produites directement par application d'un traitement bas niveau sur les pixels contenus dans une image existante.

Calque perceptif induit Calque perceptif dont les données sont construites par une fusion de données présentes dans d'autres calques perceptifs, selon des règles d'organisation perceptive.

Dans cette partie, nous nous intéressons uniquement aux calques perceptifs *directs*. Les calques de type *induit* seront détaillés dans le chapitre 7.

6.1.2.2 Utilisation des calques perceptifs

Nous rappelons que le rôle des calques perceptifs est d'organiser les données extraites à chacune des résolutions afin de faciliter leur utilisation de manière homogène comme terminaux de la grammaire.

Dans la version initiale de la méthode DMOS, on extrait, dans l'image à analyser, la liste des composantes connexes et des segments horizontaux et verticaux. Initialement, l'analyse grammaticale se base donc sur un calque perceptif direct contenant l'ensemble des segments et des composantes connexes extraits dans l'image (figure 6.3).

Dans la version multirésolution de DMOS, les composantes connexes et les segments horizontaux et verticaux sont extraits à chaque résolution. On peut donc associer un calque perceptif direct à chaque résolution(figure 6.4).

Tous les objets, composantes connexes et segments, contenus dans ces calques perceptifs représentent les terminaux de la grammaire. La description grammaticale peut donc se baser sur des données issues de chacune des résolutions étudiées, tout en ayant connaissance de la résolution de provenance de chaque élément.

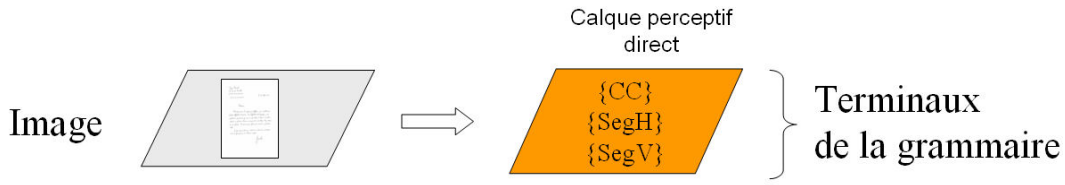


FIG. 6.3 – Version initiale de DMOS : les terminaux sont extraits dans un calque perceptif direct

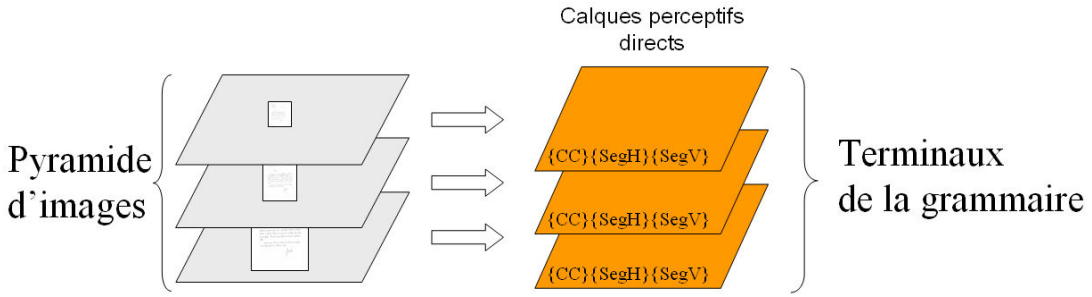


FIG. 6.4 – A chaque niveau de résolution est associé un calque perceptif direct, contenant les composantes connexes et les segments horizontaux et verticaux, qui serviront de terminaux pour la grammaire

Afin d'homogénéiser l'utilisation des terminaux de la grammaire, les coordonnées de tous les éléments contenus dans les calques sont exprimées dans le repère de l'image initiale. Ainsi, les positions d'éléments issus de différentes résolutions sont compatibles entre elles.

La construction du contenu des calques dépend des besoins de l'application. Par exemple, la figure 6.5 montre le contenu des calques nécessaire pour la reconnaissance des lignes de texte. Les lignes de texte sont vues de loin comme des segments, on a donc besoin de travailler avec les segments extraits à résolution basse. D'autre part, les lignes de texte peuvent être décrites dans l'image de résolution initiale (normale) comme des successions de composantes connexes.

Il n'est donc pas nécessaire d'effectuer l'extraction de toutes les primitives dans tous les niveaux : la construction du contenu des calques est adaptable à chaque type de documents.

6.1.2.3 Mise en œuvre dans DMOS

Nous détaillons maintenant l'aspect technique de la mise en œuvre des calques perceptifs dans la méthode DMOS.

Nous avons montré dans la partie 5.4.1.1 que les données extraites dans les images sont stockées dans la méthode DMOS sous la forme d'une *structure analysée*. Dans la méthode classique, nous rappelons que la structure analysée contient un pointeur vers

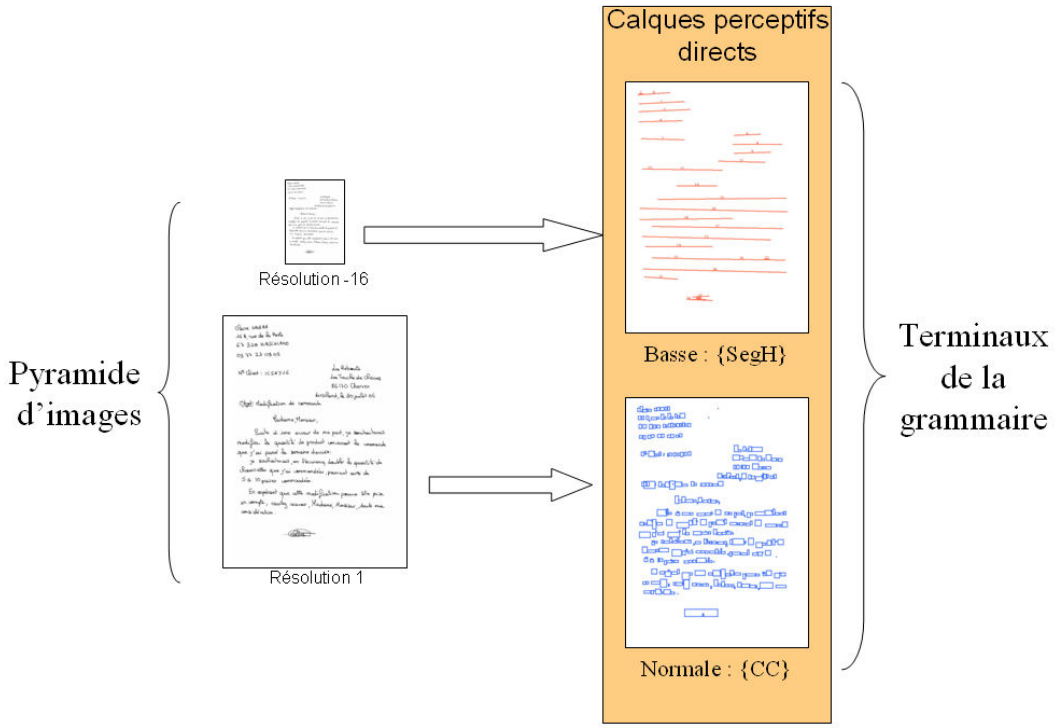


FIG. 6.5 – Exemple de calques requis pour la reconnaissance des lignes de texte : les segments à basse résolution (-16) et les composantes connexes à résolution normale (1)

l'image de référence, et des listes de composantes connexes, de segments horizontaux et verticaux qui proviennent du résultat de l'extraction des primitives dans cette image. La structure analysée a donc une expression simplifiée de la forme :

```
Structure = {RefImage, ListeCC, ListeSegHoriz, ListeSegVerti}.
```

Nous utilisons cette structure comme base d'un calque perceptif, puisqu'elle contient à la fois les données extraites et un lien vers l'image d'origine ayant servi à produire ces données. Afin de distinguer et de manipuler les calques perceptifs, nous ajoutons un nom à ce calque. Ainsi, un calque perceptif aura pour forme :

```
Calque = {"Nom", Structure}
```

où la `Structure` est celle définie ci-dessus.

Dans la nouvelle version, l'analyse doit pouvoir se baser sur un ensemble de calques, c'est à dire sur un ensemble de structures repérées par leurs noms. L'ensemble des calques disponibles est stocké dans un objet manipulé : `EnsembleCalques`.

Par exemple, prenons le cas de lignes de texte vues de loin comme des segments et de près comme des composantes connexes. On construit une pyramide faite de deux images : `ImageNormale` est l'image initiale et `ImageBasse` est l'image 16 fois plus petite. Dans l'image `ImageBasse`, on extrait les segments. Dans l'image `ImageNormale`, on extrait les composantes connexes. Pour l'analyse, nous utilisons deux calques, un associé à chacune des résolutions, `Normale` et `Basse`. :

```
EnsembleCalques = [{"Normale", StructureNormale},
                   {"Basse", StructureBasse}].
```

Les structures associées à chacune des résolutions sont les suivantes :

```
StructureNormale = {ImageNormale, ListeCCNormale, [],[]}.
StructureBasse = {ImageBasse, [], ListeSegHBasse, ListeSegVBasse}.
```

`[]` représente la liste vide. En effet, on n'utilise ni les segments à résolution normale, ni les composantes connexes à résolution basse. Les listes `ListeCCNormale`, `ListeSegHBasse` et `ListeSegVBasse` sont construites grâces aux extracteurs de terminaux présentés dans la partie 5.2.

De manière plus générale, l'analyseur a désormais à sa disposition autant de structures à analyser que de calques permettant de décrire le document. Il faut donc pouvoir choisir, à chaque étape de l'analyse d'un document, quelle est la structure à étudier. Ceci est permis par l'opérateur de changement de niveau de perception.

6.2 Outils de changement de niveau de perception

Nous présentons l'opérateur `USE_LAYER`[1] dont le rôle est de changer le niveau de perception en cours d'analyse, ce qui revient à spécifier le calque perceptif étudié à chaque étape de l'analyse.

[1] L'opérateur USE_LAYER correspond à l'opérateur FOCUSING présenté dans plusieurs publications précédentes [3] [5]. Ce changement de nom est dû à l'ambiguïté qui existait pour le terme FOCUSING.

Cet opérateur est défini par la syntaxe suivante :

```
USE_LAYER(nomResolution) FOR(regle)
```

Cet opérateur permet de spécifier qu'on utilise les terminaux contenus dans le calque perceptif `nomResolution` pour la reconnaissance du (non-)terminal `regle`.

D'un point de vue interne, cet opérateur a pour effet de chercher dans la liste des calques disponibles le calque de nom `nomResolution`, puis de charger la structure associée comme structure à analyser. L'analyse se poursuit alors en tenant compte des terminaux présents dans le calque `nomResolution`, c'est à dire extraits dans l'image de résolution `nomResolution`. Cet opérateur est entièrement compatible avec la version initiale de DMOS et permet de continuer à utiliser de manière transparente des systèmes écrits antérieurement.

Nous présentons un exemple de l'utilisation de l'opérateur `USE_LAYER` dans le cas de la reconnaissance des lignes de texte. Les résolutions et les calques perceptifs disponibles sont rappelés sur la figure 6.5. Voici un exemple de règle simplifiée, décrivant une ligne de texte perçue globalement comme un segment puis localement comme un ensemble de composantes connexes :

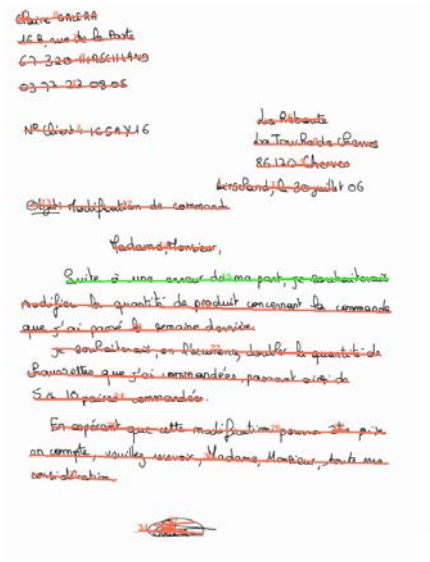
```
ligneDeTexte ::=
      TERM_SEG condSegHoriz noCondS etiq S &&
      AT(zoneSegment S) &&
      USE_LAYER("Normale") FOR(ensCompConnexes).
```

L'analyse d'une ligne de texte commence avec une vision de loin, c'est-à-dire dans le calque correspondant à la résolution `Basse`. Dans ce calque, on commence par reconnaître un segment horizontal `S` grâce à l'opérateur d'extraction de segments `TERM_SEG` (figure 6.6(a)). Ce segment `S` nous permet de définir une zone d'intérêt `zoneSegment` (figure 6.6(b)), c'est à dire un polygone centré sur `S`. Cette zone définie permet de spécifier un contexte spatial dans lequel on souhaite s'intéresser à l'image dans sa résolution normale. L'appel à l'opérateur `USE_LAYER` permet de prendre en compte le contenu du calque associé à la résolution `Normale` pour reconnaître un ensemble de composantes connexes `ensCompConnexes` (figure 6.6(c)).

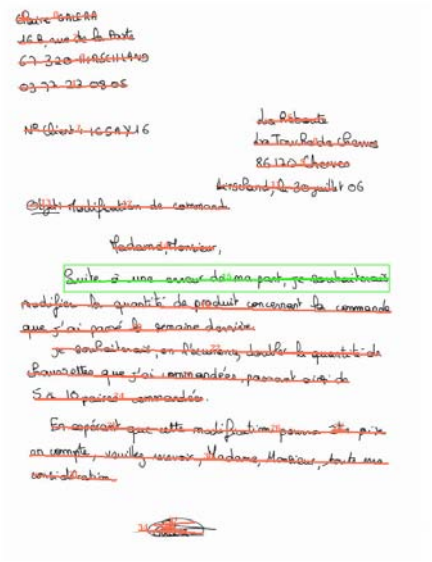
L'opérateur `USE_LAYER` que nous avons introduit permet ainsi de changer le niveau de perception en cours d'analyse. Il est donc maintenant possible de spécifier un nouveau point de vue par la combinaison de deux opérateurs :

- `USE_LAYER` spécifie le niveau de perception,
- `AT` précise le contexte spatial.

Grâce à ces deux opérateurs, le changement de point de vue est commandé depuis la description symbolique dans le langage EPF. Il s'agit donc d'une analogie avec le cerveau humain puisque l'activation de modèles spécifiques dans la mémoire de l'homme modifie le point d'attention de la rétine. Ce changement de point de vue permet d'itérer dans le cycle perceptif.



(a) Choix d'un segment horizontal (en vert)



(b) Définition d'une zone d'intérêt autour du segment

(c) Focalisation d'attention dans cette zone pour reconnaître un ensemble de composantes connexes

FIG. 6.6 – Principe de reconnaissance d'une ligne de texte

6.3 Outils de transfert d'information

Nous avons mis en évidence la nécessité de faire coopérer les éléments reconnus à différents niveaux de résolution.

Un premier moyen de coopération est fourni par la notion de calque perceptif, dans lequel les éléments extraits aux différentes résolutions sont décrits dans un même repère de coordonnées. La cohérence de ces coordonnées permet notamment de transférer des zones de contexte spatial lors du passage d'une résolution à une autre.

Cependant, deux difficultés se présentent lorsqu'on veut faire coopérer de manière plus précise des éléments extraits de résolutions différentes. Il s'agit de :

- la précision de la localisation,
- le changement de nature des primitives étudiées.

La précision de la localisation est un problème lié au bruit de quantification, qui apparaît lors du changement de résolution, d'une résolution basse vers une résolution haute. En effet, la différence d'échelle entre les résolutions entraîne, lors de la conversion, d'une part une approximation de ces coordonnées, ayant pour ordre de grandeur le facteur de zoom, et d'autre part une perte de précision sur la forme.

Par exemple, la transposition d'un segment extrait à résolution -16 (figure 6.7(a)) dans une image de résolution 1 (figure 6.7(b)) est réalisée avec une approximation à 16 pixels près. Ceci peut provoquer un décalage par rapport aux pixels effectivement présents dans l'image de résolution 1. Ce décalage est dû d'une part au manque de précision des coordonnées, et d'autre part à la perception de la pente ou de la courbure (c'est à dire la forme) qui varie selon la résolution.

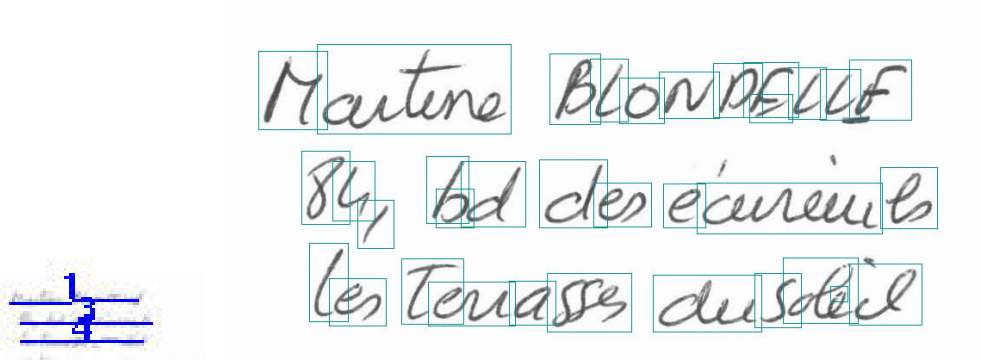
(a) Segment extrait à résolution -16

(b) Le même segment transféré à résolution 1 : localisation imprécise par rapport au trait existant dans l'image

FIG. 6.7 – Perte de précision lors du transfert d'un segment

Le changement de nature des primitives étudiées est la deuxième difficulté rencontrée lors du transfert d'informations. Par exemple, une ligne de texte est visible de loin comme un segment (figure 6.8(a)), mais de près comme un ensemble de composantes connexes (figure 6.8(b)). Il faut donc pouvoir établir un lien entre le segment et les pixels formant les composantes connexes.

Pour répondre aux deux difficultés rencontrées pour le transfert des informations, nous devons utiliser des données :



(a) Vision de loin : lignes de texte perçues comme des segments

(b) Vision de près : lignes de texte vues comme des ensembles de composantes connexes

FIG. 6.8 – Changement de nature des primitives selon la résolution

– présentant un niveau de détail suffisant pour pouvoir exprimer une localisation précise,
– n'ayant pas de lien direct avec une résolution donnée de l'image, pour s'abstraire de la représentation par des primitives différentes.

Ainsi, nous proposons deux entités : la *ligne abstraite* et le *rectangle abstrait*, abstractions inter-résolution du segment et de la composante connexe.

6.3.1 Ligne abstraite

6.3.1.1 Définition

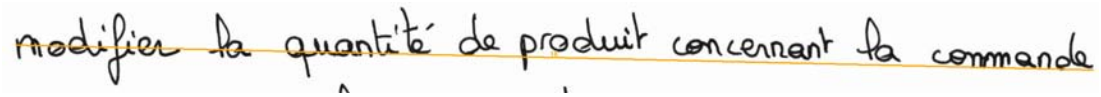
Nous définissons la ligne abstraite comme ci-dessous :

Ligne abstraite Objet des données numériques, constitué d'un ensemble de pixels globalement rectilignes, caractérisé par une épaisseur, n'ayant pas d'existence physique directe dans une image donnée.

Le concept de *ligne abstraite* permet de répondre à deux besoins :
– manipuler des données précises au niveau numérique,
– manipuler des données indépendamment de la résolution.

Les lignes abstraites permettent de manipuler des données plus précises que les segments. En effet, au niveau de DMOS, les segments sont représentés uniquement par les coordonnées de leurs extrémités. Cependant, l'extracteur basé sur le filtre de Kalman produit davantage d'informations, qui sont stockées au niveau numérique : position des pixels, épaisseur moyenne. Ces informations sont disponibles pour les lignes abstraites. Un exemple de cas où ces informations sont utiles est celui des lignes courbes pour

lesquelles un segment constitué de ses extrémités ne représente pas la ligne de manière assez précise (figure 6.9(a)). Dans ce cas, l'utilisation de l'entité de ligne abstraite apporte plus de précision sur la position (figure 6.9(b)).



(a) Segment représenté par ses extrémités

(b) Ligne abstraite permettant de manipuler des données plus précises

FIG. 6.9 – Intérêt de la ligne abstraite pour le cas des lignes courbes

Par ailleurs, la ligne abstraite est indépendante du niveau de résolution : elle n'est pas liée à une résolution donnée, et sa position peut être ajustée en fonction d'informations présentes à différentes résolutions. Ces informations peuvent être de nature hétérogène, permettant ainsi de traiter des changements de primitives entre résolutions.

6.3.1.2 Fonctionnalités

Nous avons créé plusieurs fonctionnalités facilitant le transfert d'informations grâce aux lignes abstraites.

Création des lignes abstraites Il est existe trois moyens principaux pour définir une ligne abstraite :

- à partir des coordonnées de deux points qui formeront les extrémités de la ligne abstraite,
- par approximation polygonale d'un segment existant dans une image,
- par concaténation de deux lignes abstraites déjà existantes.

Modification des lignes abstraites Trois outils permettent de modifier des lignes abstraites :

- l'extension d'une ligne par extrapolation des extrémités,
- le découpage d'une ligne abstraite pour en sélectionner une sous-partie,
- le recalage sur des pixels : cet opérateur est le principal, nous le détaillons dans la partie suivante.

6.3.1.3 Opérateur de recalage

Le but de l'opérateur de recalage est de répondre aux deux difficultés liées au transfert d'informations :

- préciser la localisation spatiale en fonction des pixels présents,
- faire correspondre des primitives de nature différentes.

Ajustement de la localisation spatiale L'outil de recalage permet de répondre au problème de positionnement présenté sur la figure 6.7.

Lors du transfert d'une ligne abstraite d'une résolution basse vers une résolution haute, la ligne abstraite initiale donne des indices globaux tels que la position, la courbure, l'épaisseur. Ces indices sont alors utilisés pour définir un contexte de recherche dans l'image de destination. Dans ce contexte, on s'intéresse aux pixels noirs présents dans l'image pour ajuster le positionnement de la ligne abstraite, et ainsi limiter les erreurs de quantification.

Un exemple est présenté sur la figure 6.10. Un segment est détecté à faible résolution (visualisé à résolution haute sur la figure 6.10(a)). Ce segment sert de base pour la création d'une ligne abstraite, par approximation polygonale. Cette ligne permet de définir un contexte de recherche associé dans l'image de résolution supérieure (figure 6.10(b)). Dans cette zone de recherche, on recense tous les pixels noirs (figure 6.10(c)). Ces pixels noirs sont ensuite utilisés pour définir localement une position moyenne de la ligne abstraite finale (figure 6.10(d)).

Cet outil permet donc d'utiliser un contexte défini de manière symbolique pour faire appel aux informations numériques, à savoir les pixels noirs de l'image, afin de produire un positionnement plus précis lors du transfert de données d'une résolution à une autre.

Notons que le recalage peut être conditionné par la présence d'un segment dans la résolution supérieure. Ce mécanisme sera utilisé plus en détail dans la section 7.1.2.

Gestion du changement de primitive Dans l'exemple précédent, le trait présent dans l'image est perçu aux deux résolutions comme un segment. Cependant, notre outil de recalage est capable de mettre en correspondance des primitives de natures différentes. En effet, il est possible de recaler une ligne abstraite sur des pixels appartenant à des composantes connexes, selon le principe de recalage évoqué ci-dessus.

Ainsi, la figure 6.11 présente un exemple de recalage de lignes de texte, perçues comme des segments, sur les pixels des composantes connexes associées. Comme dans le cas précédent, les segments vus à faible résolution permettent de définir une zone de recherche dans laquelle on sélectionne les pixels concernés par le recalage. Un calcul de la position moyenne de ces pixels permet de produire le recalage final.

Dans cet exemple, on ajuste la position de la ligne abstraite sur le milieu de la ligne de texte, ce qui ne présente qu'un faible déplacement. Cependant, il est également possible de produire un recalage sur les extrema locaux, hauts ou bas, des pixels des composantes connexes. Ceci permet, par exemple, de trouver le positionnement de lignes de base de l'écriture manuscrite.

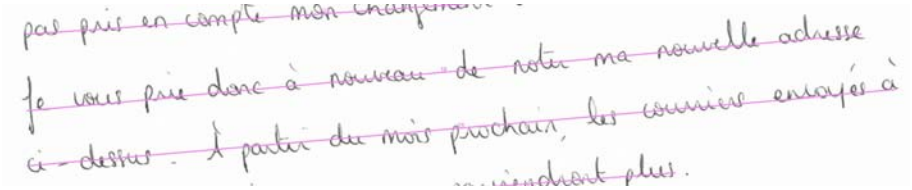
(a) Transposition du segment perçu à faible résolution dans une image à haute résolution

(b) Zone de recherche (en bleu)

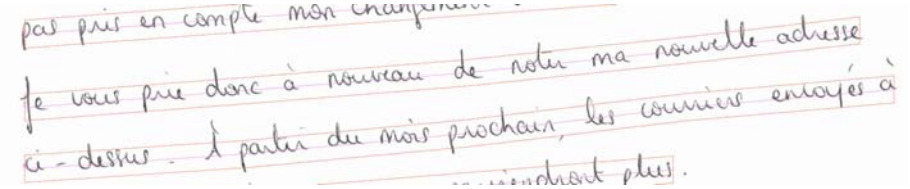
(c) Pixels utilisés pour le recalage à haute résolution (en bleu)

(d) Recalage final à haute résolution

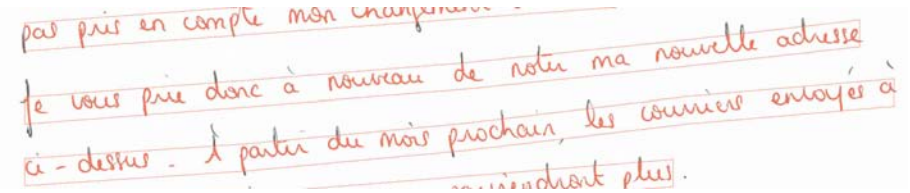
FIG. 6.10 – Processus de recalage, dans un contexte de changement de résolution, pour ajuster plus précisément le positionnement



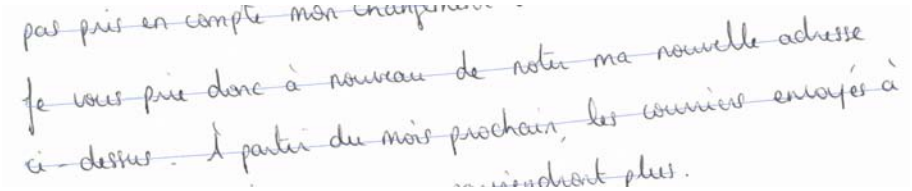
(a) Segments vus à faible résolution



(b) Zone de recherche (en orange)



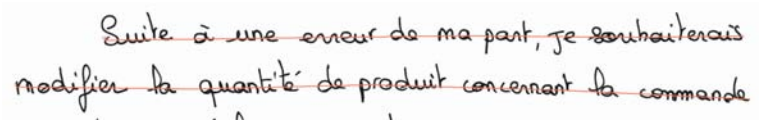
(c) Pixels utilisés pour le recalage à haute résolution (en rouge)



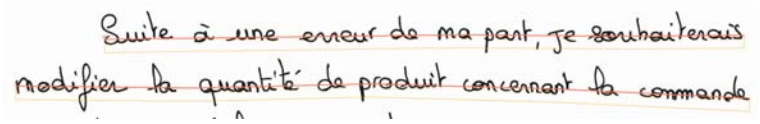
(d) Recalage final à haute résolution

FIG. 6.11 – Processus de recalage, dans un contexte de changement de résolution, avec gestion de changement de primitive : un segment est lié à des pixels de composantes connexes

La figure 6.12 présente un exemple de recalage sur les extrema bas de l'écriture. Ainsi, à partir des lignes abstraites initiales (figure 6.12(a)), on définit un contexte de recherche dans la partie inférieure (figure 6.12(b)), dans lequel sont détectés les minima locaux des zones de pixels noirs (figure 6.12(c)). Ces minima locaux sont alors utilisés pour positionner la ligne abstraite sur le bas de l'écriture (figure 6.12(d)). On note que les jambages de l'écriture ne perturbent pas le positionnement.

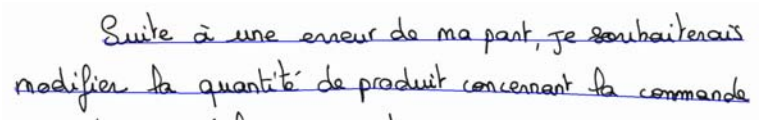


(a) Visualisation des lignes abstraites initiales



(b) Zone de recherche en dessous de la ligne initiale (en orange)

(c) Détail des pixels composant les extrema



(d) Lignes finales recalées sur le bas de l'écriture

FIG. 6.12 – Exemple de recalage de lignes abstraites sur le bas de l'écriture

La figure 6.13 présente un exemple de résultat du recalage de lignes abstraites sur les extrema hauts de l'écriture.

En conclusion, le concept de ligne abstraite permet de répondre, pour des éléments de type ligne, aux problèmes d'ajustement de la localisation spatiale et de gestion du changement de primitive dans la mise en correspondance entre résolutions.

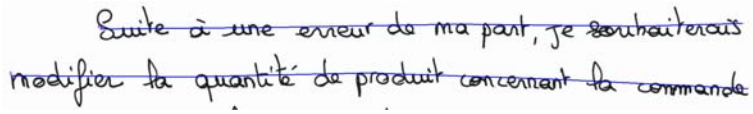


FIG. 6.13 – Exemple de recalage de lignes abstraites sur le haut de l'écriture

6.3.2 Rectangle abstrait

Le rectangle abstrait est une abstraction inter-résolution de la notion de composante connexe, défini ci-dessous :

Rectangle abstrait Objet de niveau numérique, décrit par les coordonnées de ses deux points extrêmes, n'ayant pas d'existence physique directe dans une image donnée.

Les rectangles abstraits sont créés de manière similaire aux lignes abstraites, afin de pouvoir manipuler des objets de type composantes connexes, indépendamment de la résolution.

Dans une première version de nos travaux, nous n'avons pas d'utilisation directe de ces rectangles abstraits. On peut imaginer, comme exemple d'application, la vision multirésolution des images ou des photos contenues dans un document. Il serait nécessaire pour leur reconnaissance de pouvoir décrire le rectangle englobant indépendamment de la résolution initiale.

6.4 Bilan

Nous avons introduit, dans la méthode DMOS existante, des nouveaux outils et formalismes liés à la multirésolution.

Grâce à nos travaux, il est désormais possible de baser l'analyse sur des images à plusieurs résolutions. Nous avons également introduit la notion de calque perceptif qui permet de gérer, de manière transparente, des données issues de plusieurs résolutions.

Dans ce contexte, il est maintenant possible de décrire des documents en confrontant des points de vue issus de plusieurs résolutions. Ceci est permis principalement par le nouvel opérateur que nous avons introduit dans le langage EPF : `USE_LAYER`. Cet opérateur, combiné avec l'opérateur de position `AT`, permet en effet de commander, depuis le niveau symbolique, les changements de points de vue successifs, à chaque étape de l'analyse.

Notre concept de calques perceptifs permet de faciliter la coopération entre les données issues de différents niveaux. Afin d'augmenter cette coopération, nous avons créé les concepts de ligne abstraite et de rectangle abstrait qui permettent de stocker des informations indépendamment d'une résolution donnée. Des outils tels que le recalage des lignes permettent d'augmenter la mise en correspondance des données entre résolutions, notamment en précisant la localisation et en gérant les changements de nature de primitives, ce qui est un cas fréquent.

Nos apports dans la méthode DMOS sont regroupés sur la figure 6.14, qui peut être comparée avec figure 5.1 présentant la méthode DMOS initiale. Comme le montre cette figure, les apports dans la méthode DMOS ont été réalisés à plusieurs niveaux et ont donc nécessité une bonne prise en main du fonctionnement interne de la méthode existante. L'intégration des nouvelles fonctionnalités a été réalisée en respectant la philosophie et le fonctionnement de la méthode DMOS initiale, afin de conserver une compatibilité avec les anciennes grammaires écrites pour DMOS. Nous avons pour cela abordé nos travaux par une étape de conception logicielle dans laquelle les interactions de DMOS avec la multirésolution ont été spécifiées au maximum.

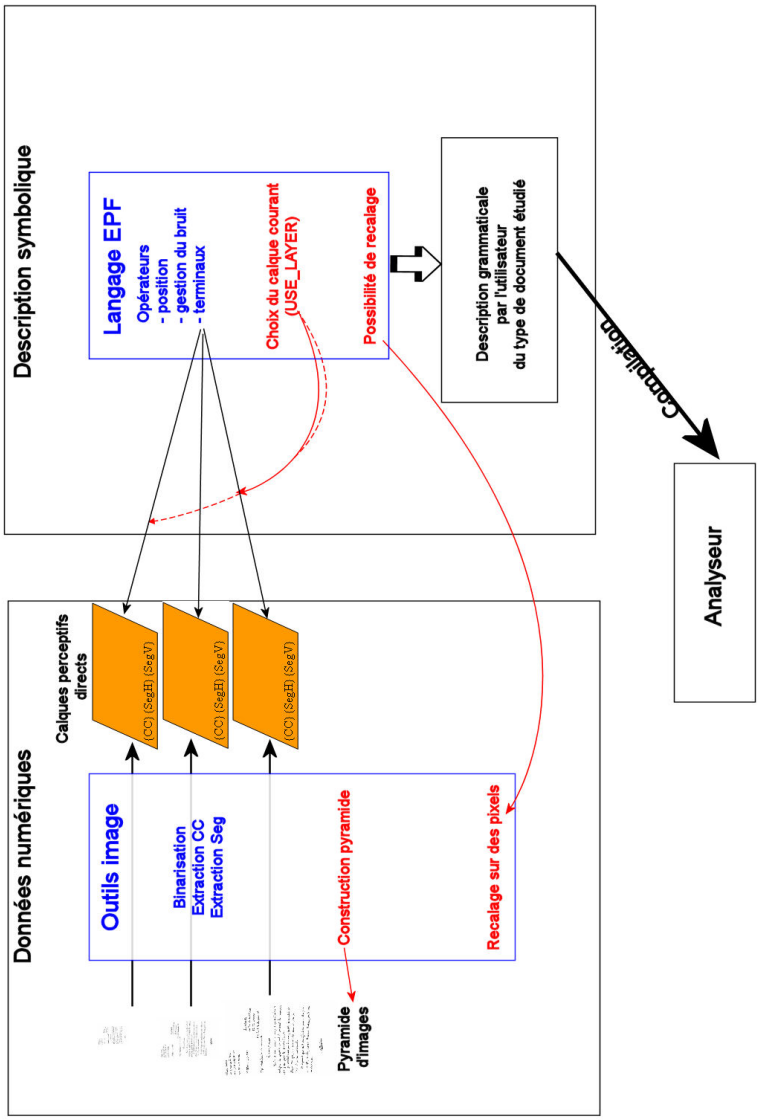


FIG. 6.14 – Méthode DMOS enrichie des outils de multirésolution

Vis à vis des besoins évoqués dans le chapitre 4, les outils de multirésolution présentés ici permettent de fournir les éléments suivants, nécessaires au cycle perceptif :

- des images et données multirésolutions,
- l'opérateur de changement de niveau de perception
- le transfert d'information entre résolutions.

Nous allons maintenant montrer comment la nouvelle version de la méthode DMOS ainsi créée sert de support pour la gestion d'objets structurels prégnants, qui complètent notre système DMOS-P.

Chapitre 7

Gestion d'objets structurels prégnants

Dans le chapitre 4, nous avons mis en évidence le besoin d'imiter les deux types d'attention visuelle. Nous avons montré dans le chapitre 5 que l'attention visuelle guidée par un but est déjà simulée dans la méthode DMOS par le principe de la description grammaticale. Nous proposons donc d'introduire dans l'architecture de DMOS-P de nouveaux mécanismes pour simuler l'attention guidée par des éléments prégnants.

Nous avons montré, dans le chapitre 2, l'intérêt de reconnaître des objets structurels prégnants. En effet, la reconnaissance de ces objets peut être réalisée par application des lois de l'organisation perceptive, sans avoir besoin de connaissance spécifique sur le type de document étudié, mais uniquement des connaissances sur les objets prégnants.

L'architecture proposée dans le chapitre précédent permet de décrire des éléments prégnants. Nous illustrons cette possibilité en étudiant deux types d'éléments structurels prégnants existant dans les images de documents :

- les lignes de texte,
- les traits.

Il est important de noter que les lignes de texte et les traits sont deux exemples d'éléments prégnants que nous avons décrits dans notre méthode, mais notre approche est suffisamment générique pour envisager la description d'autres éléments prégnants.

Pour chacun des éléments étudiés, nous allons produire une description grammaticale perceptive dans le langage EPF, qui pourra être appliquée sur tout type de documents, pour produire, selon le cas, une liste de lignes de textes ou de traits. Nous présentons dans un premier temps ces mécanismes de construction des éléments prégnants.

Nous souhaitons ensuite pouvoir réutiliser ces éléments prégnants comme une base pour la reconnaissance de documents plus complexes. Nous présentons dans la seconde partie la manière dont nous avons homogénéisé l'utilisation des éléments prégnants avec celle des terminaux usuels de la grammaire. Grâce à cette architecture, il est alors possible d'intégrer l'utilisation des éléments prégnants pour la description de mécanismes perceptifs plus complexes.

7.1 Construction

Pour chacun des deux types d'objets étudiés, lignes de texte et traits, nous avons produit une grammaire dans le langage EPF, en se basant sur la méthode DMOS gérant la multirésolution, présentée dans le chapitre précédent.

Nous présentons donc successivement le mécanisme de reconnaissance des lignes de texte puis des traits.

7.1.1 Lignes de texte

Nous avons présenté dans la partie 3.1.1 les principaux enjeux de la reconnaissance de lignes de texte et l'intérêt de la vision perceptive pour ce type de problème. Nous détaillons le mécanisme de reconnaissance, présentons la grammaire déduite, puis mettons en avant, au travers d'exemples d'applications variés, les intérêts de notre méthode.

7.1.1.1 Stratégie d'analyse

La description perceptive des lignes de texte consiste à combiner la perception des lignes, vues de loin comme des segments, et de près comme des composantes connexes. Nous utilisons donc deux résolutions de l'image : l'image dans sa résolution initiale, dite `Normale`, et l'image dont les dimensions sont divisées par 16, dite `Basse`. Les calques perceptifs contenant les terminaux de la grammaire sont présentés sur la figure 7.1.

Les étapes d'analyse sont les suivantes :

1. Analyse des segments à faible résolution
   (a) Sélectionner un segment de base S1, en déduire la ligne L1 (figure 7.2(a)).
   (b) Rechercher des segments S2 alignés avec S1, dans une zone tenant compte de l'épaisseur de S1 (figure 7.2(b)) ; en déduire les lignes L2.
   (c) Construire une ligne abstraite L à partir des lignes L1 et L2 (figure 7.2(c)).
   (d) Transférer L à résolution `Normale`, en la recalant sur le milieu des pixels noirs présents (figure 7.2(d)).
2. Focalisation sur pour obtenir le détail des composantes connexes à résolution normale :
   (a) Délimiter une bande de recherche sur toute la largeur de la page, dans l'axe de L (figure 7.2(e)).
   (b) Extraire un alignement de composantes connexes (figure 7.2(f)).
   (c) Mettre à jour L (figure 7.2(g)) :
   – ajuster la longueur en fonction des composantes connexes trouvées,
   – ajuster la position par recalage en tenant compte de la hauteur moyenne des composantes connexes.

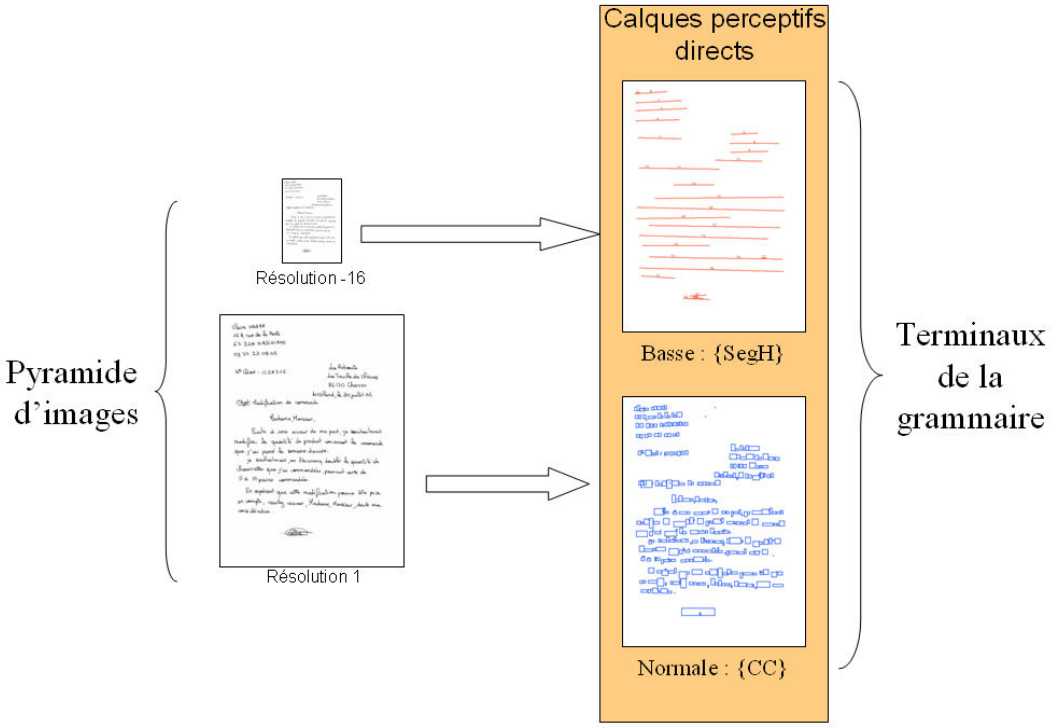
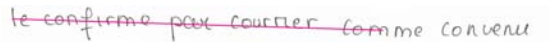
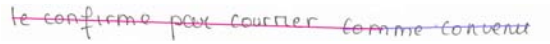


FIG. 7.1 – Calques perceptifs utilisés pour la reconnaissance des lignes de texte

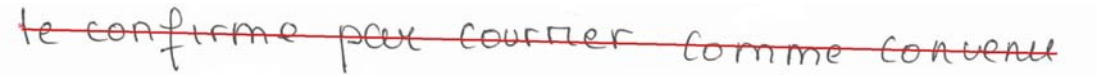


(a) Sélection d'un segment S1, en déduire la ligne L1

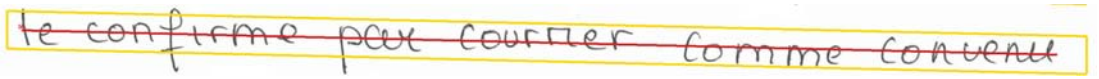


(b) Sélection d'un segment aligné S2(en bleu), en déduire la ligne L2

(c) Construction d'une ligne abstraite L à partir de L1 et L2



(d) Transfert à haute résolution et recalage de L



(e) Délimitation de la zone de recherche autour de L, déduite de l'épaisseur de L

(f) Extraction des composantes connexes

(g) Ligne finale

FIG. 7.2 – Mécanisme de reconnaissance d'une ligne de texte

7.1.1.2 Grammaire EPF

Cette stratégie d'analyse perceptive a été traduite dans le langage EPF pour former une grammaire d'extraction des lignes de texte. Nous en présentons ici une version simplifiée, dans laquelle certains attributs ont été enlevés pour faciliter la lecture.

La règle `ligneTexte` permet de générer une ligne de texte. Elle est appelée de manière récursive afin d'extraire toutes les lignes de texte du document. L'analyse commence à la résolution `Basse` grâce à un appel préliminaire à `USE_LAYER`.

On reconnaît d'abord le segment le plus long non encore analysé dans le document, `SegBase`. Ce segment permet de définir une zone de recherche, `autourSeg`, dans laquelle on recherche des segments `EnsSeg` alignés à `SegBase`. On construit ensuite la ligne abstraite `L` à partir de tous les segments, grâce à `consLigne`. La dernière étape consiste à se focaliser à la résolution `Normale` pour détailler les composantes connexes contenues dans la ligne.

```
ligneTexte L2 ::=
     TERM_SEG (condGlobaleS condSegPlusLong) noCondS segBase SegBase &&
     AT(autourSeg SegBase) &&
     autresSegs EnsSeg &&
     consLigne [SegBase|EnsSeg] L &&
     USE_LAYER(nomResol "Normale") FOR(decomposeLigne L L2).
```

La règle `consLigne` a pour but de construire une ligne abstraite `LH1` à partir des segments passés en paramètre, puis de recaler cette ligne à résolution normale, en tenant compte de son épaisseur `Ep`, pour produire la ligne `LH2`.

```
consLigne ListeSegs LH2 :-
     consLigneConcSegs ListeSegs LH1,
     epaisseurLigne LH1 Ep,
     recaleLigne LH1 Normale -Ep Ep LH2.
```

La règle `decomposeLigne` encapsule une utilisation de l'opérateur `IN DO` qui permet de limiter la zone de recherche des composantes connexes à la zone `zoneLigne` illustrée sur la figure 7.2(e). Elle appelle `decomposeLigneZone` dont le but est de rechercher les composantes connexes de la ligne, situées de part et d'autre d'une composante de base `CCbase`, puis de mettre à jour la ligne en fonction des composantes connexes trouvées.

```
decomposeLigne L L2::=
     IN(zoneLigne L) DO(decomposeLigneZone L L2).
```

```
decomposeLigneZone L L2 ::=
       AT(posDebutLigne L) &&
       TERM_CMP condPasTropPetiteCC noCondC ccbase CCbase &&
       AT(justeAGauche CCbase) &&
       debutAlignementCC CCsDeb &&
       AT(justeADroite CCbase) &&
       finAlignementCC CCsfin &&
       majCoorLigne L1 [CCbase|CCsDeb|CCsFin] L2.
```

La règle `majCoorLigne` permet de recaler la position de `L1` en fonction de la hauteur moyenne `Hcc` des composantes connexes trouvées `Ccs`, afin de produire la ligne finale `L2`. Grâce à ce recalage, la ligne abstraite finale est placée précisément et de manière indépendante à la résolution étudiée.

```
majCoorLigne L1 Ccs L2 :-
     hauteurMoyenneCCs Ccs Hcc,
     recaleLigne L1 Normale -Hcc Hcc L2.
```

Cette grammaire écrite en langage EPF génère par compilation un analyseur capable d'extraire les lignes de texte, sans connaissance *a priori* sur le type de document étudié, ce qui justifie le caractère prégnant des lignes de texte.

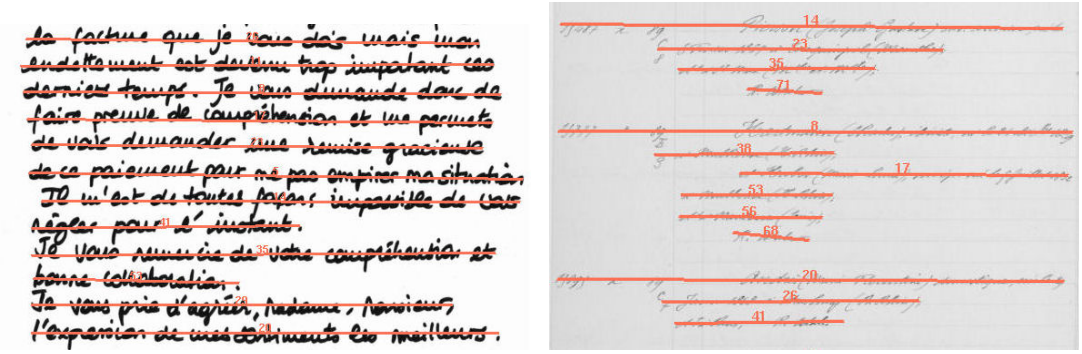
7.1.1.3 Application

Nous appliquons notre extracteur de lignes de texte sur plusieurs types de documents, sans avoir besoin d'adapter, ni la description, ni les paramètres, au type de document. Ceci permet de mettre en avant plusieurs points forts de la méthode.

Indépendance du style d'écriture L'utilisation combinée de deux points de vue local et global permet d'intégrer des variations dans le style d'écriture. Ainsi les lignes sont aussi bien reconnues sur de l'écriture manuscrite épaisse (figure 7.3(a)) , de l'écriture manuscrite fine (figure 7.3(b)), ou des documents imprimés (figure 7.3(c)).

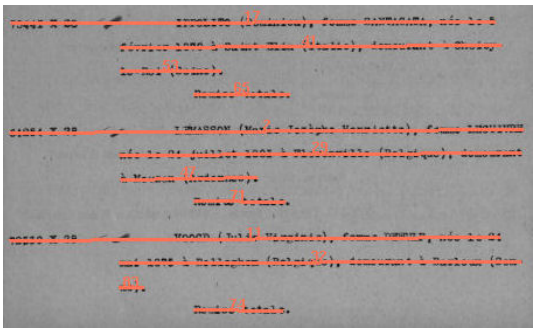
Indépendance de l'alphabet Notre méthode permet de reconnaître les lignes de texte indépendamment de la langue utilisée mais aussi des caractères utilisés. Par exemple, la figure 7.4 montre les lignes extraites dans un document écrit dans un dialecte indien : le Bangla.

Gestion du biais Notre méthode est basée sur les segments extraits avec le filtre de Kalman (présenté dans l'annexe B), capable d'extraire des éléments en biais. Cette gestion de la pente et du biais se répercute sur l'extraction des lignes de texte qui peuvent être inclinées (figure 7.5), et même avoir des pentes variables au sein d'un même document (figure 7.6).



(a) Écriture manuscrite épaisse

(b) Écriture manuscrite fine



(c) Écriture imprimée

FIG. 7.3 – Indépendance du style d'écriture

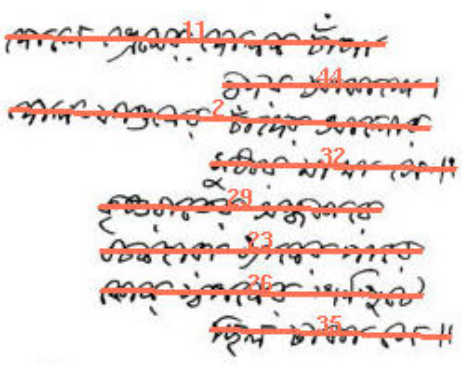


FIG. 7.4 – Reconnaissance sur un alphabet non latin

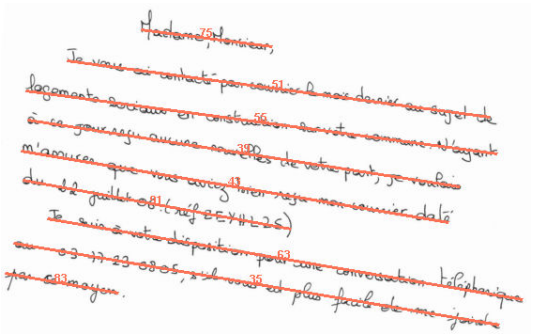


FIG. 7.5 – Document en biais (environ 10 degrés)

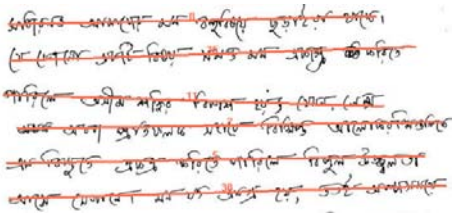


FIG. 7.6 – Lignes de texte ayant des pentes variables

Gestion de la courbure L'utilisation des lignes abstraites et de l'outil de recalage permet d'utiliser des informations sur les pixels qui guident la position de la ligne de texte. Ainsi, cela permet de trouver précisément la position des lignes, même lorsqu'elles sont marquées par une courbure (figure 7.7).

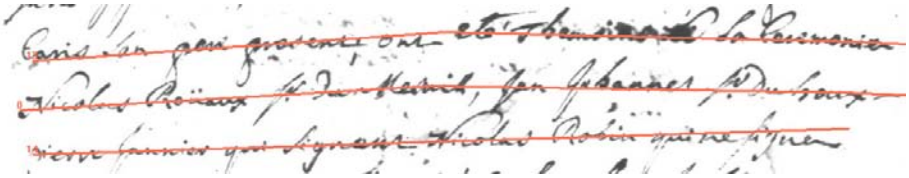


FIG. 7.7 – Lignes de texte courbes

Gestion des chevauchements L'utilisation d'une vision globale permet de distinguer les lignes de texte même si certaines lettres se chevauchent d'une ligne à l'autre (figure 7.8). Notre but est ici de détecter les endroits problématiques qui nécessiteraient une segmentation des composantes connexes, tout en fournissant un contexte. Ces informations pourront être utilisées dans une étape ultérieure de reconnaissance de l'écriture.

Notre méthode d'extraction des lignes de texte permet donc de traiter les différents problèmes habituellement rencontrés dans les documents d'archives (présentés dans la partie 3.1.1), sans connaissance *a priori* sur le type de documents étudiés, ce qui justifie

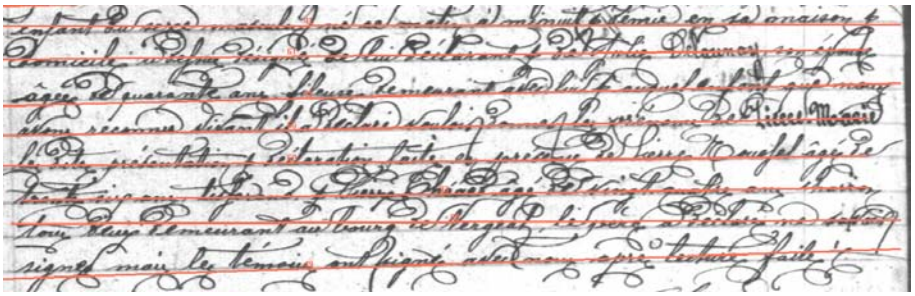


FIG. 7.8 – Lignes de texte avec chevauchement

l'aspect prégnant. Ces lignes de texte pourront servir de base pour la reconnaissance de structures plus complexes comme nous le présenterons dans la partie 7.2.

7.1.2 Traits

Nous présentons maintenant le processus de reconnaissance perceptive d'un second type d'objets prégnants : les traits, appelés aussi filets.

Nous avons mis en évidence dans la partie 3.1.2 les différents types de traits que l'on souhaite reconnaître ainsi que les difficultés pour les reconnaître dans les documents anciens : bruits, mouchetage, discontinuités .... Nous regroupons les traits à reconnaître sous trois catégories (figure 7.9) :

- les traits épais,
- les traits doubles,
- les traits fins.

Nous avons également mis en avant l'intérêt d'utiliser conjointement plusieurs points de vue pour la reconnaissance de traits dans des documents.

Dans les parties suivantes, nous présentons la stratégie d'analyse permettant de reconnaître différents types de traits, avant de présenter son implémentation sous forme d'une grammaire EPF. Nous terminons en montrant un exemple d'application de notre méthode s'appuyant sur l'architecture perceptive proposée.

7.1.2.1 Stratégie d'analyse

Pour la détection de traits, nous proposons de baser la perception sur trois niveaux de résolution.

En effet, nous avons constaté expérimentalement qu'en utilisant trois résolutions, les différences de perception entre points de vue sont significatives, sans être trop importantes. Notre stratégie se base donc sur les points de vue suivants :

- la vision *globale* correspond à une image à basse résolution ; elle est construite à partir de l'image initiale dont on a divisé les dimensions par 16 ;
- la vision *intermédiaire* est construite à partir de l'image initiale dont on a divisé les dimensions par 4 ;
- la vision *locale* correspond à l'image initiale, en haute résolution.

(a) Trait épais

(b) Trait double

(c) Trait fin discontinu

FIG. 7.9 – Les trois types de traits à reconnaître

Les calques perceptifs correspondant aux terminaux de la grammaire sont présentés sur la figure 7.10.

Pour chacun de ces points de vue, la perception des segments est différente (figure 7.11). C'est cette variation de perception qui permet d'élaborer notre mécanisme de reconnaissance des traits.

Vision globale En regardant un document à basse résolution (figure 7.11(b)), les éléments perçus comme des segments sont :
- les lignes épaisses, même si elles sont dégradées (figure 7.9(a)),
- les lignes multiples (figure 7.9(b)) qui apparaissent comme un seul segment,
- certaines lignes de texte en caractères gras, ou plus foncées.

Les filets fins ne sont pas assez contrastés pour être perçus à ce niveau de résolution.

Vision intermédiaire En regardant ce document à résolution moyenne (figure 7.11(c)), les éléments perçus comme des segments sont :
- des morceaux de filets multiples (figure 7.9(b)),
- des morceaux de filets fins (figures 7.9(c)),
- des morceaux de filets épais (figure 7.9(a)),
- des parties rectilignes de lettres majuscules.

Vision locale En regardant ce document à résolution haute (figure 7.11(d)), on peut percevoir les segments suivants :
- des morceaux de filets multiples (figure 7.9(b)),
- des morceaux de filets simples (figures 7.9(c)),

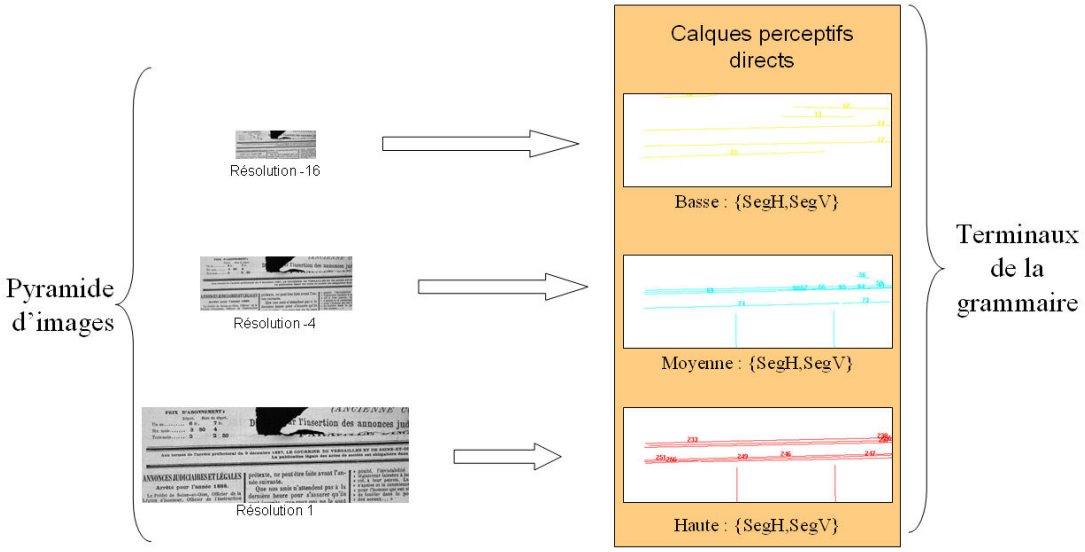


FIG. 7.10 – Calques perceptifs utilisés pour la reconnaissance des traits

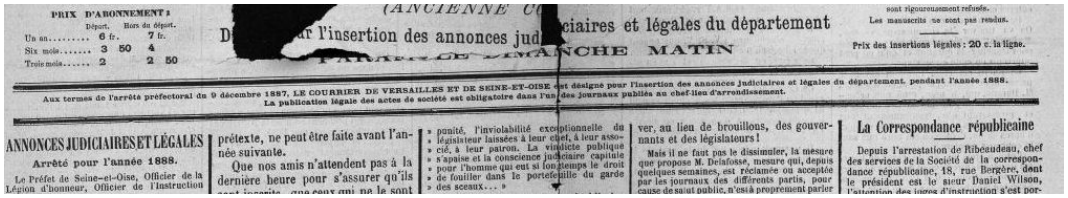
– des éléments causés par du bruit.
A cette résolution, les segments épais peuvent ne pas être perçus si le bruit est trop important, par exemple dans le cas de mouchetage blanc (figure 7.9(a)).

Ces constatations sont regroupées dans le tableau 7.1. On distingue les traits que l'on souhaite reconnaître, c'est à dire les traits épais, fins et multiples, des « faux » traits qui sont du bruit dans l'analyse (lignes de texte, caractères, bruits liés à la mauvaise qualité du document).

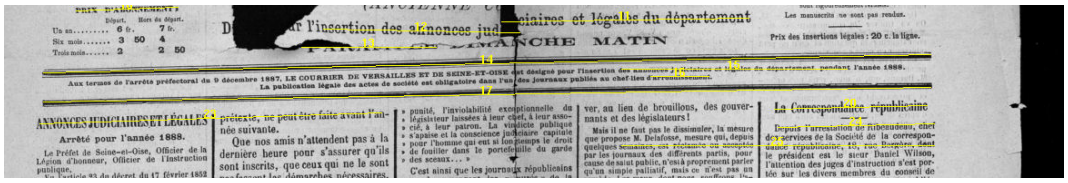
Vision	Globale	Intermédiaire	Locale
Traits fins	Non	Oui	Oui
Traits épais	Oui	Oui	Non
Traits multiples	Oui	Oui, 2 lignes	Oui, 2 lignes
Lignes de texte	Oui	Non	Non
Lettres	Non	Oui	Non
Bruit	Non	Oui	Non

TAB. 7.1 – Pour chaque type de trait, point de vue dans lequel ils sont généralement perçus comme des segments

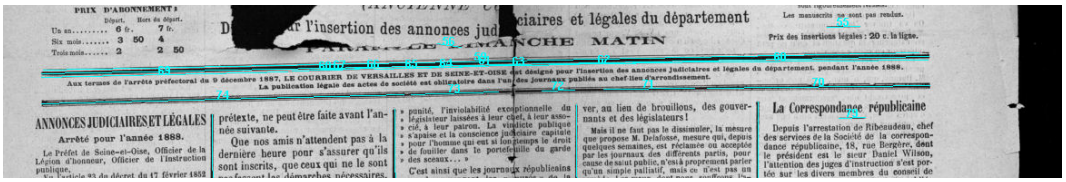
Le tableau 7.1 montre que la perception qu'on peut avoir d'un élément aux diverses résolutions permet de déterminer le type de trait étudié. Nous en déduisons donc une stratégie d'analyse qui combine les visions aux différentes résolutions pour construire un trait résultat. Le principe de base est que la vision à la résolution inférieure permet d'émettre une hypothèse sur la présence d'un trait. Cette hypothèse peut être confirmée



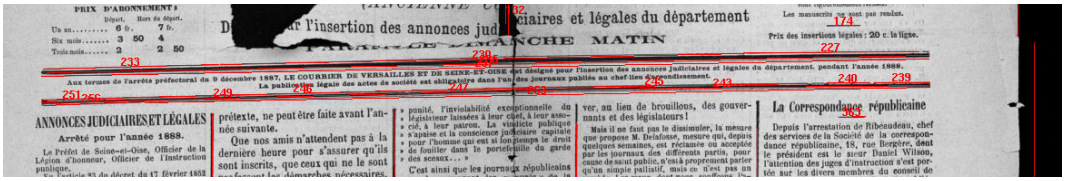
(a) Image initiale



(b) Présentation des segments perçus avec une vision globale (en jaune)



(c) Présentation des segments perçus avec une vision intermédiaire (en bleu)



(d) Présentation des segments perçus avec une vision locale (en rouge)

FIG. 7.11 – Segments perçus aux différents points de vue (reportés dans un même référentiel pour plus de lisibilité ; les segments sont numérotés pour pouvoir être identifiés)

par la présence de segments aux résolutions supérieures. La stratégie globale suit donc un principe de prédiction/vérification. Elle est décrite sur la figure 7.12.

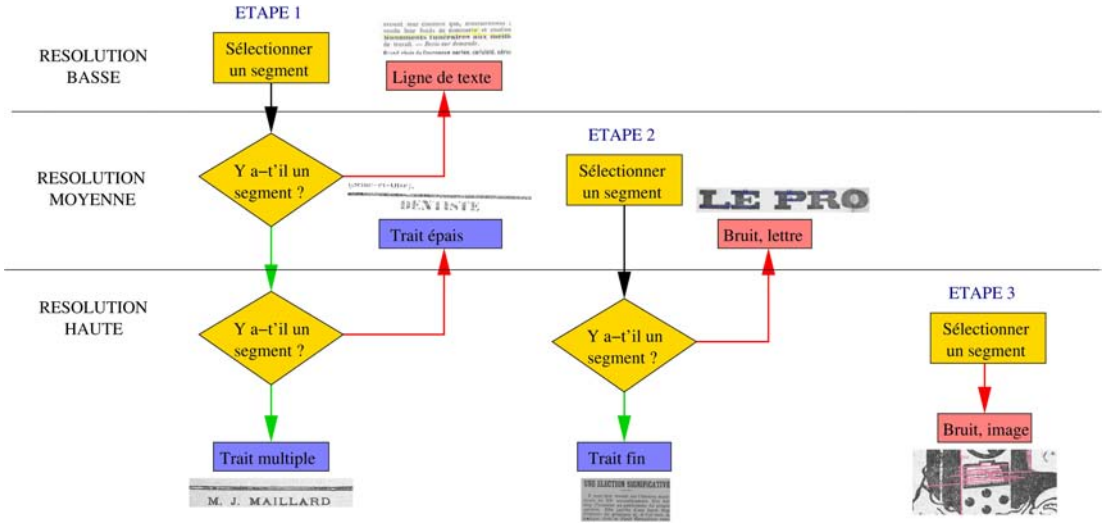


FIG. 7.12 – Stratégie de combinaison des segments détectés aux trois résolutions

L'analyse est réalisée en deux étapes. On recherche d'abord les segments à basse résolution et leur correspondance à moyenne et haute résolution. Puis, on s'intéresse aux segments restants à moyenne résolution et à leur correspondance à haute résolution.

La première étape consiste à sélectionner un segment à basse résolution. A partir de ce segment, on teste la présence de segments associés à moyenne résolution. Si aucun segment n'est trouvé, il doit s'agir d'une ligne de texte. Dans le cas contraire, la présence à haute résolution de segments associés permet de construire, selon le cas un trait multiple ou épais.

Le même mécanisme est proposé pour reconnaître les lignes fines, perceptibles aux résolutions moyenne et haute.

On considère que les segments perceptibles uniquement à haute résolution sont des éléments de bruit, tels que des segments contenus dans des images [1].

Il est important de noter que c'est la position du segment vu à basse résolution qui va déterminer la zone de recherche pour les résolutions supérieures. De plus, la vision à faible résolution fournit des indices sur la nature de la ligne : biais, courbure, épaisseur, longueur, qui servent à définir le contexte d'agglutination pour la construction de la ligne finale.

[1] Une évaluation sur 4967 traits montre que la prise en compte des segments vus uniquement à haute résolution entraîne 498 fausses reconnaissances supplémentaires (10.0%) en ne permettant de reconnaître que 18 traits supplémentaires (0.36%)

7.1.2.2 Grammaire EPF

Cette stratégie d'analyse a été traduite dans le langage EPF pour former une grammaire d'extraction des traits. Nous présentons ici une version simplifiée de la grammaire, dans laquelle certains attributs ont été enlevés pour faciliter la lecture.

Reconnaître les traits dans une image, avec une `Direction` donnée (tendance horizontale ou verticale) c'est produire deux listes de traits issues des deux premières étapes de la stratégie : à partir de la vision de loin et à partir de la vision intermédiaire. On ignore en effet les segments qui seraient produits par la troisième étape.

```
lesTraits Direction L::=
     USE_LAYER(nomResol "Basse")
         FOR(listeDeTraitsLoin Direction L1) &&
     USE_LAYER(nomResol "Moyenne")
         FOR(listeDeTraitsMoy Direction L2 ) &&
     append L1 L2 L.
```

Nous détaillons ici uniquement la première partie de la stratégie, à partir des segments vus de loin. La détection de la `listeDeTraitsLoin` passe par un appel récursif à la règle `chercheTraitLoin`. Cette règle a pour but d'extraire un segment `S1` puis de tenter de le recaler à résolution moyenne pour produire le trait `Trait`. [2]

```
chercheTraitLoin Direction Trait ::=
     %Chercher le segment de base, d'abord à faible résolution
     TERM_SEG Direction noCondS segBase S1 &&
     tenteRecalageMoy S1 Trait .
```

Pour recaler le segment, nous le transformons en une ligne abstraite, puis nous utilisons l'opérateur de recalage avec l'option où le recalage de la ligne est conditionné par la présence de segment (présentée dans la partie 6.3.1.3). Cette option de l'outil de recalage permet de répondre à la question « Y a-t-il un segment ? » à la résolution moyenne. Dans la première clause de `tenteRecalageMoy`, le recalage réussit, c'est donc qu'un segment est présent à résolution moyenne. Dans ce cas, on détaille les segments présents et on poursuit l'analyse à résolution haute. Dans la deuxième clause, le recalage échoue : le segment vu de loin n'a pas de correspondance à la résolution moyenne, on renvoie un trait vide `noTrait`.

```
%Cas où le recalage réussit
tenteRecalageMoy S1 Trait ::=
     recaleSegBasSurSegMoy S1 LRec  &&
     USE_LAYER(nomResol "Moyenne")
         FOR(detailleSeg LRec (condSegAssezPresEpais LRec) Ens2) &&
     tenteRecalageHaut LRec Trait.
%Cas où le recalage échoue
tenteRecalageMoy S1 noTrait.
```

[2] En EPF, le symbole % permet de préfixer les commentaires.

La règle `tenteRecalageHaut` permet d'essayer de recaler la ligne trouvée à la résolution moyenne en fonction de segments qui seraient présents à résolution haute. Là encore, nous utilisons l'outil de recalage avec l'option permettant de vérifier la présence de segments.

Si le recalage réussit, cela signifie que l'on a affaire à une ligne multiple. L'épaisseur et la position de la ligne à résolution moyenne nous permettent de définir une zone de recherche dans laquelle on peut détailler les morceaux de segments présents à haute résolution, qui appartiennent à la ligne multiple.

```
tenteRecalageHaut LMoy Trait ::=
     recaleLigneMoySurSegHaut LMoy LRec &&
     USE_LAYER(nomResol "Normale")
         FOR(detailleSeg LRec (condSegAssezPresEpais LRec) Ens2) &&
   %Construire le trait final avec tous les segments
     consTrait Ens2 LRec Trait.
```

Si le recalage échoue, cela signifie qu'il n'y a pas de segments perceptibles à résolution haute. Il s'agit donc probablement d'une ligne épaisse. Dans ce cas, on utilise le recalage sur le milieu des pixels pour positionner de manière précise la ligne `LMoy` à résolution haute. On en déduit le trait final `Trait`.

```
tenteRecalageHaut LMoy Trait ::=
     recaleLigneSurPixHaut LMoy LRec &&
     consTrait Ens Trait.
```

Cette grammaire écrite dans le langage EPF permet de produire automatiquement un analyseur capable d'extraire des traits de types variés, sans connaissance spécifique liée au document, ce qui montre la prégnance des traits.

7.1.2.3 Application

Nous présentons un exemple de résultats obtenus grâce à notre méthode. L'image présentée sur la figure 7.13 regroupe de nombreuses difficultés liées à la reconnaissance de traits. Ces difficultés sont regroupées dans le tableau 7.2 et identifiées par leur numéro dans la colonne Item. La figure 7.14 montre les éléments reconnus par notre méthode.

Le tableau montre que notre méthode est capable de reconnaître les lignes attendues, tout en ignorant les lignes superflues, qui sont perçues comme des segments à certaines résolutions. Ceci est permis par la combinaison des résolutions qui valide la présence d'un trait uniquement si celui-ci est perçu comme un segment aux résolutions attendues.

Notre méthode permet donc de reconnaître les différentes catégories de traits, au sein d'un même document, sans connaissance *a priori* sur la nature de ces traits. Ces traits vont pouvoir servir de base pour la reconnaissance de structures plus complexes. C'est ce que nous présentons dans la partie suivante.

Item	Difficulté	Résolutions de perception	Avec notre méthode
1	Traits fins et en biais	Moyenne, haute	Reconnus
2	Traits épais	Basse, moyenne	Reconnus
3	Filets multiples	Toutes	Reconnus
4	Filets doubles	Toutes	Reconnus
5	Pliure du papier	Moyenne	Ignorée
6	Caractères épais	Moyenne	Ignorés
7	Lignes de texte	Basse	Ignorées

TAB. 7.2 – Difficultés rencontrées pour la reconnaissance de traits (les items font référence aux numéros de la figure 7.13 ; les résultats de notre méthode sont présentés sur la figure 7.14)

FIG. 7.13 – Difficultés rencontrées pour la reconnaissance de traits : en bleu les numéros des traits à reconnaître, en rouge les éléments à ignorer, détaillés dans le tableau 7.2

FIG. 7.14 – Traits reconnus par notre méthode : lignes épaisses en rose, lignes multiples en turquoise, lignes fines en bleu

7.2 Exploitation

Dans la première partie de ce chapitre, nous avons montré deux utilisations de notre méthode perceptive, pour produire les descriptions grammaticales des lignes de texte et des traits.

Ces deux grammaires ont été réalisées par analogie avec le principe des objets prégnants, c'est à dire que les éléments constituant les lignes de texte et les traits ont été regroupés selon des critères de proximité et d'alignement, qui font partie des lois de l'organisation perceptive. Par conséquent, ces grammaires ont la propriété de ne pas contenir de connaissances liées à un document en particulier ; elles peuvent donc s'appliquer sur tous types de documents.

Par ailleurs, les éléments prégnants que sont les lignes de texte et les traits peuvent servir d'entités élémentaires pour la reconnaissance de structures plus complexes. Nous souhaitons donc pouvoir les réutiliser comme support dans des grammaires décrivant d'autres types de documents. Ainsi, il sera possible de construire des mécanismes perceptifs plus complexes, tout en ayant une spécification plus simple.

7.2.1 Définition de nouveaux terminaux

L'idée que nous avons souhaité mettre en place est de pouvoir utiliser le résultat de l'extraction des lignes de texte ou de la reconnaissance des traits, comme terminaux d'autres grammaires en EPF. Nous avons montré dans la partie 6.1.2.1 que les terminaux de la grammaire sont désormais stockés dans des calques perceptifs.

Pour prendre en compte de nouveaux terminaux, les lignes de texte et les traits, nous devons donc les utiliser via un calque perceptif. Cependant, ces terminaux ne sont pas liés directement avec une image à une résolution donnée, puisqu'ils sont construits de manière multirésolution. Nous avons donc étendu l'architecture du système de vision perceptive en créant un nouveau type de calque perceptif : le calque perceptif induit.

Nous en rappelons la définition, évoquée dans la partie 6.1.2.1 :

Calque perceptif induit Calque perceptif dont les données sont construites par une fusion de données présentes dans d'autres calques perceptifs, selon des règles d'organisation perceptive.

Ces calques perceptifs permettent donc de prendre en compte de nouveaux terminaux pour la grammaire, les lignes de texte et les traits, qui ne sont pas extraits directement dans une image, mais construit par l'application d'une description multirésolution. Les données stockées dans les calques perceptifs induits sont indépendantes de la résolution. Elles sont donc stockées sous formes de lignes ou de rectangle abstraits, en l'occurrence de lignes abstraites pour le cas des traits et des lignes de texte. Les données des calques perceptifs sont utilisées comme terminaux pour la grammaire (figure 7.15).

7.2.2 Utilisation des nouveaux terminaux

L'utilisation de ces nouveaux terminaux dans la grammaire se fait de manière homogène à l'utilisation faite des calques perceptifs directs, grâce aux opérateurs de change-

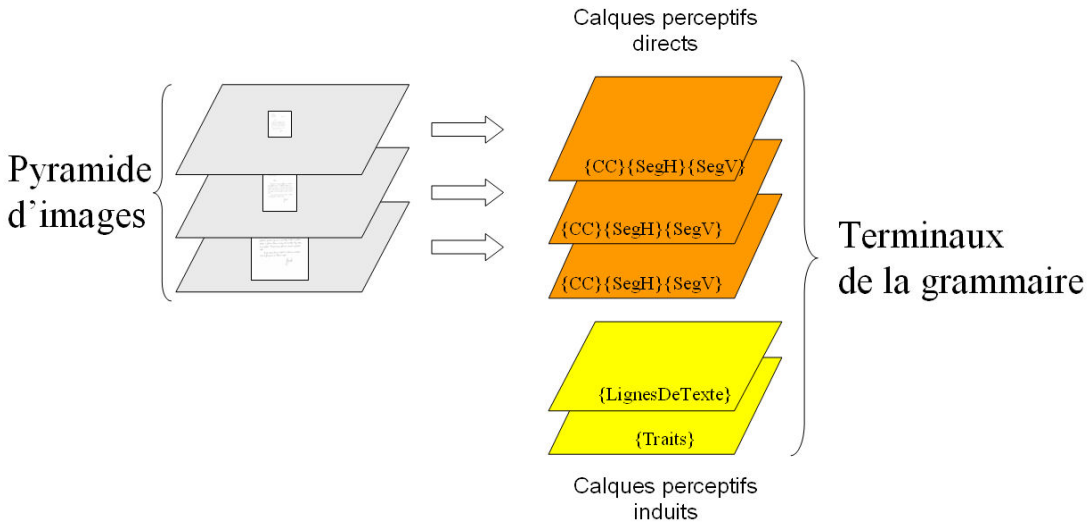


FIG. 7.15 – Utilisation des éléments prégnants comme terminaux de la grammaire, grâce au formalisme des calques perceptifs

ment de niveau de perception et de détection des terminaux. Cette homogénéité d'utilisation permet de mélanger les informations perceptives directes et induites (issues de la perception prégnante).

7.2.2.1 Changement de niveau de perception

Il est possible d'utiliser l'opérateur de changement de niveau de perception `USE_LAYER` pour accéder au contenu des calques perceptifs induits. En effet, le but de l'opérateur `USE_LAYER` est d'utiliser comme structure à analyser les éléments contenus dans un calque perceptif donné. Ainsi, l'expression

```
USE_LAYER("Traits") FOR(regle)
```

permet d'appliquer la règle `regle` en tenant compte des terminaux contenus dans le calque perceptif induit contenant les traits.

7.2.2.2 Reconnaissance des terminaux

Les données stockées dans les calques perceptifs des lignes de texte et des traits sont des lignes abstraites représentées par leurs extrémités. Ce sont donc des éléments de nature semblable aux segments contenus dans les calques perceptifs directs.

Il est donc possible d'utiliser l'opérateur d'extraction des terminaux de type segments, `TERM_SEG`, pour extraire les lignes de texte et les traits contenus dans les calques perceptifs induits. Par extension, il est donc possible d'appliquer des conditions pour la reconnaissance de ces éléments, de la même manière que l'on pouvait le faire pour les segments usuels.

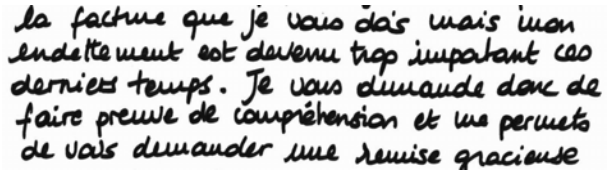
Par exemple, l'expression suivante,

```
TERM_SEG condAssezEpais noCondS etiq Trait
```

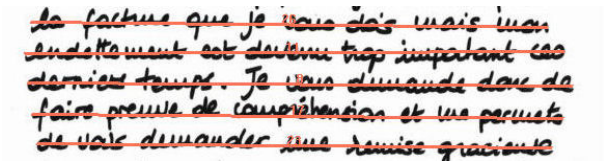
appliquée dans le calque `"Traits"` permet de reconnaître le trait `Trait`, dont l'épaisseur doit être suffisante.

7.2.3 Exemple d'utilisation

Nous présentons un exemple simple d'utilisation des lignes de texte pour la reconnaissance d'un paragraphe, tel que celui présenté que la figure 7.16(a). Les lignes de texte ont été reconnues lors d'une première phase d'analyse (figure 7.16(b)) et sont maintenant disponibles comme des terminaux, stockés dans le calque perceptif induit `"lignesDeTexte"`.



(a) Paragraphe



(b) Lignes de texte reconnues

FIG. 7.16 – Exemple de paragraphe dont on produit la description

Reconnaître un paragraphe, c'est reconnaître une succession de lignes de texte les unes en dessous des autres. Le premier prédicat consiste à se placer dans le point de vue du calque `"lignesDeTexte"` :

```
paragraphe ::=
     USE_LAYER("lignesDeTexte") FOR(detailParagraphe).
```

Le détail du paragraphe consiste alors à un appel récursif à la reconnaissance de lignes, situées les unes en dessous des autres :

```
detailParagraphe [LigneDeTexte|AutresLignes] ::=
     TERM_SEG noCondS noCondS ligne LigneDeTexte &&
     AT(sousLigne LigneDeTexte) &&
     detailParagraphe AutresLignes.
```

```
%Condition d'arrêt
detailParagraphe [].
```

7.3 Intérêt des éléments prégnants

Du point de vue de la grammaire, les terminaux habituellement proposés (composantes connexes et segments) ne sont pas parfaits, de par la nature accidentée des documents : recouvrement inopportun de composantes connexes, segments interrompus. A l'opposé, les éléments prégnants représentent des terminaux de meilleure qualité, construits grâce à une fusion des données contenues dans les calques perceptifs.

L'intérêt majeur de cette approche est donc la simplification apportée à l'utilisateur qui peut désormais décrire des documents dans le langage EPF, en utilisant directement les traits et les lignes de texte en plus des autres terminaux (composantes connexes et segments) qui restent disponibles. Les mécanismes perceptifs utilisés pour construire les lignes de texte et les traits en combinant les différentes résolutions sont totalement occultés pour les utilisateurs. En effet, la construction des éléments prégnants est totalement indépendante de son utilisation pour la description d'entités plus complexes.

L'architecture mise en place permet de gérer autant de type d'éléments prégnants que souhaité. En effet, il est possible d'ajouter de nouveaux calques perceptifs induits, traduisant de nouveaux niveaux de perception. Dans ce chapitre, nous avons étudié plus spécifiquement les lignes de texte et les traits. Ces deux types d'éléments prégnants ont été choisis parce qu'ils sont fréquemment présents dans de nombreux documents.

L'introduction des éléments prégnants finalise l'architecture de notre système perceptif DMOS-P, dont la version finale est présentée sur la figure 7.17.

Comme la méthode DMOS (figure 5.1), DMOS-P se base sur l'analyse de données numériques, guidée par une description symbolique contenant la connaissance. La synthèse d'un analyseur adapté est réalisée par une étape de compilation.

Le niveau numérique est désormais basé sur une pyramide d'images multirésolutions. Le formalisme des calques perceptifs permet d'organiser les données numériques, à la fois extraites directement des images, et induites par fusion intelligente de données. L'introduction de nouvelles structures de contrôles symboliques (`USE_LAYER`, recalage) permet de tirer pleinement profit de l'ensemble des données numériques disponibles.

L'utilisateur de DMOS-P a donc désormais à sa disposition une architecture complète et cohérente pour la description simplifiée de mécanismes perceptifs complexes, adaptés à chaque type de documents étudiés.

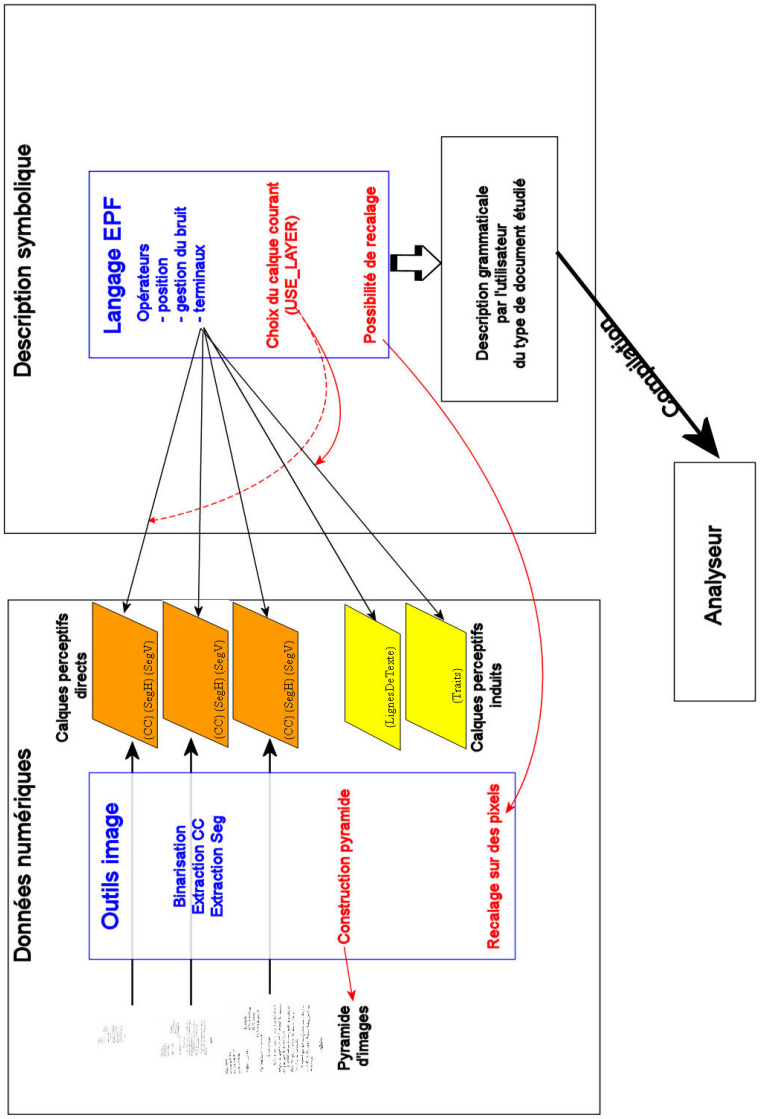


FIG. 7.17 – Notre méthode finale : DMOS-P

Conclusion de la seconde partie

Dans la seconde partie de la thèse, nous avons montré comment nous avons enrichi la méthode existante DMOS, pour produire notre méthode perceptive DMOS-P. Nous synthétisons maintenant la manière dont DMOS-P imite les mécanismes de la vision perceptive humaine, c'est à dire un cycle perceptif guidé par l'attention visuelle.

Mise en évidence des cycles perceptifs

Notre implémentation permet d'imiter le cycle perceptif, à la fois en respectant ses différents composants, mais aussi par l'aspect cyclique de l'enchaînement des étapes.

Composants du cycle perceptif

Nous rappelons les différentes étapes du cycle perceptif sur la figure 7.18, en noir. Il s'agit de :

- la définition d'un point de vue : niveau de perception et contexte spatial ;
- l'extraction de primitives dans cette image ;
- la mise en adéquation des primitives avec les modèles contenus en mémoire.

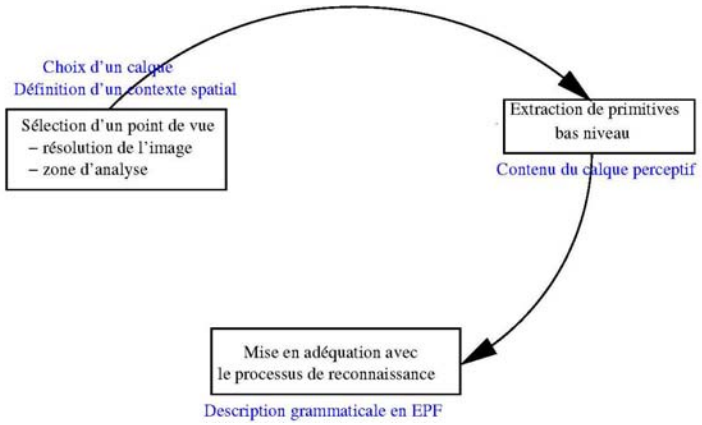


FIG. 7.18 – Composants du cycle perceptif imités dans DMOS-P

L'étape de sélection d'un point de vue est réalisée par le choix du niveau de perception (calque) à étudier, à tout moment de l'analyse, grâce à l'opérateur `USE_LAYER`.

La définition d'un contexte spatial dans la grammaire, avec l'opérateur `AT` du langage EPF, permet de préciser la zone d'intérêt.

Le résultat de l'extraction des primitives dans l'image se trouve contenu dans le calque perceptif lui-même. Dans notre approche, les primitives étudiées sont les composantes connexes, les segments horizontaux et verticaux, et des objets construits plus complexes tels que les lignes de texte ou les traits.

L'interprétation des primitives présentes est réalisée par la grammaire dans le langage EPF qui décrit l'agencement multirésolution des primitives les unes par rapport aux autres.

Aspect cyclique

Selon les aspects présentés dans le chapitre 1, l'aspect cyclique de la vision perceptive apparaît :

- lorsque la mise en correspondance avec le modèle en mémoire nécessite plus de précisions et un détail à un autre point de vue ;
- ou lorsque la mise en correspondance avec des modèles en mémoire échoue, ce qui nécessite une remise en cause des éléments observés.

Dans ces deux cas, la vision perceptive forme un cycle puisqu'une nouvelle acquisition de l'image est réalisée (figure 7.19). Nous montrons comment cet aspect cyclique est effectivement réalisé au sein de notre méthode.

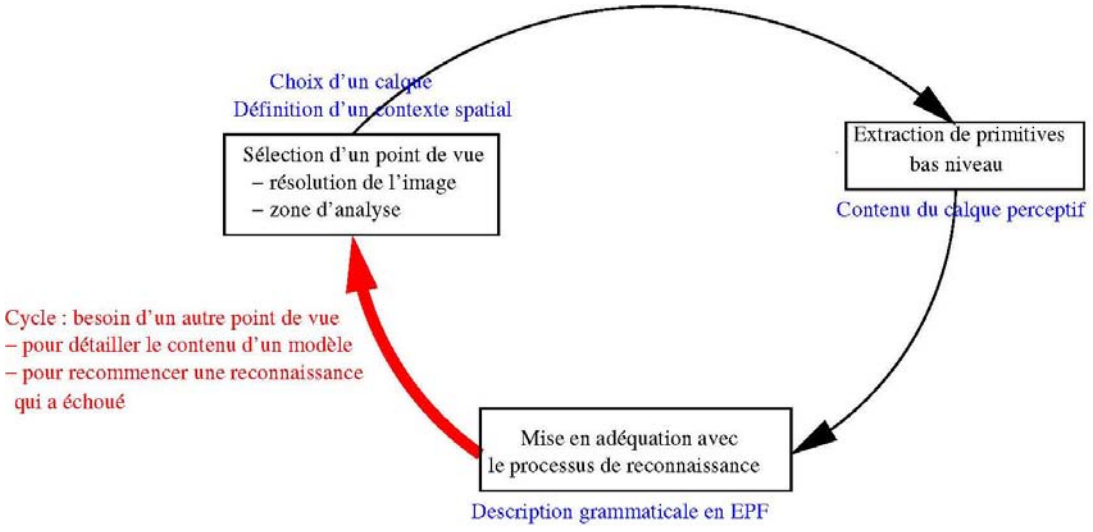


FIG. 7.19 – Formation d'un cycle

Détail d'informations

Le fait de demander un nouveau point de vue en cours d'analyse pour préciser les détails d'un objet à reconnaître est permis par l'utilisation de l'opérateur `USE_LAYER`. En effet, cet opérateur permet de changer le niveau de perception étudié en cours d'analyse.

Un exemple de cycle perceptif produit par l'opérateur `USE_LAYER` est illustré sur la figure 7.20. Il correspond à la règle, déjà présentée, de la ligne de texte vue de loin comme un segment et de près comme un ensemble de composantes connexes :

```
ligneDeTexte ::=
      TERM_SEG condSegHoriz noCondS etiq S &&
      AT(zoneSegment S) &&
      USE_LAYER("Normale") FOR(ensCompConnexes).
```

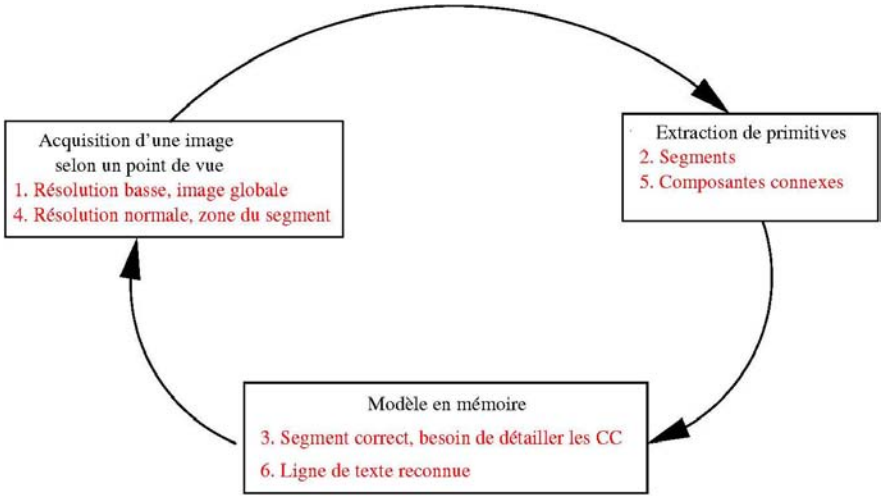


FIG. 7.20 – Exemple de cycle réalisé avec l'opérateur USE_LAYER

L'analyse démarre à résolution basse. La première étape consiste donc à se placer dans ce point de vue. Les primitives intéressantes sont les segments, parmi lesquels on est capable de trouver le segment `S`. Ensuite, le modèle en mémoire (la description en EPF) commande une focalisation d'attention. On s'intéresse donc à un autre niveau de perception, ici `"Normale"`, avec un contexte spatial réduit à `zoneSegment`. Dans ce nouveau point de vue, on extrait les composantes connexes. Si ces composantes connexes correspondent avec le modèle en mémoire, la reconnaissance est réussie.

Cet exemple met donc en évidence la manière dont l'opérateur `USE_LAYER` permet de parcourir le cycle perceptif, comme une succession de prise en compte de points de vue différents.

Remise en cause des informations

La deuxième utilisation d'un cycle perceptif intervient lorsque la mise en correspondance des primitives avec le modèle en mémoire échoue. Dans ce cas, le choix du point de vue initial est remis en cause et doit être réalisé de nouveau.

Ce mécanisme est bien présent dans notre méthode, de manière légèrement cachée. En effet, il est permis par le fait que la méthode DMOS est basée sur de la programma-

tion logique en λprolog. Ceci permet de gérer de manière transparente le retour arrière, en cas d'échec sur une règle.

Un échec sur une règle peut produire des remises en cause à plusieurs niveaux :

- si la primitive ne convient pas, la règle peut être essayée avec une autre primitive,
- si la zone de recherche ne convient pas, la règle peut être essayée avec une autre zone,
- si la résolution étudiée ne convient pas, la règle peut être essayée à une autre résolution,
- si la règle ne convient pas, elle est remise en cause.

Dans chacun de ces cas, la remise en cause provoque un cycle perceptif puisqu'elle nécessite un changement de point de vue.

Nous présentons un exemple de cycle produit suite à une erreur sur la reconnaissance d'un trait. Pour cet exemple, imaginons qu'on souhaite extraire un trait dans une zone de l'image `zoneIm`. La règle associée dans la grammaire est la suivante :

```
maRegle ::=
     AT(zoneIm) &&
     unTrait.
```

Nous avons vu dans la partie 7.1.2 qu'un trait peut être reconnu

- soit à partir de la résolution basse,
- soit à partir de la résolution moyenne.

Les deux clauses sont donc :

```
unTrait L ::=
     USE_LAYER(nomResol "Basse") FOR(unTraitLoin L).
unTrait L ::=
     USE_LAYER(nomResol "Moyenne") FOR(unTraitMoy L).
```

Supposons que dans notre cas, il n'y ait pas de segment perceptible à la résolution `Basse`. C'est donc la deuxième variante de la règle, basée sur `unTraitMoy` qui doit réussir. Nous présentons sur la figure 7.21 les différentes étapes avec la remise en cause : on commence par se positionner à résolution `Basse`, dans la zone `zoneIm`. On s'intéresse aux segments présents dans cette zone (étapes 1 et 2). Mais dans cet exemple, il n'existe pas de segment perceptible à cette résolution, ce qui fait échouer la règle `unTraitLoin` (étape 3). L'analyse redémarre donc à résolution `Moyenne` pour essayer de faire réussir la règle `unTraitMoy` et réussit finalement (étapes 4 à 9).

Cet exemple montre donc comment la remise en cause d'une règle constitue une nouvelle itération du cycle perceptif.

Notre méthode DMOS-P est donc réellement basée sur le cycle perceptif, qui est le socle de la vision perceptive.

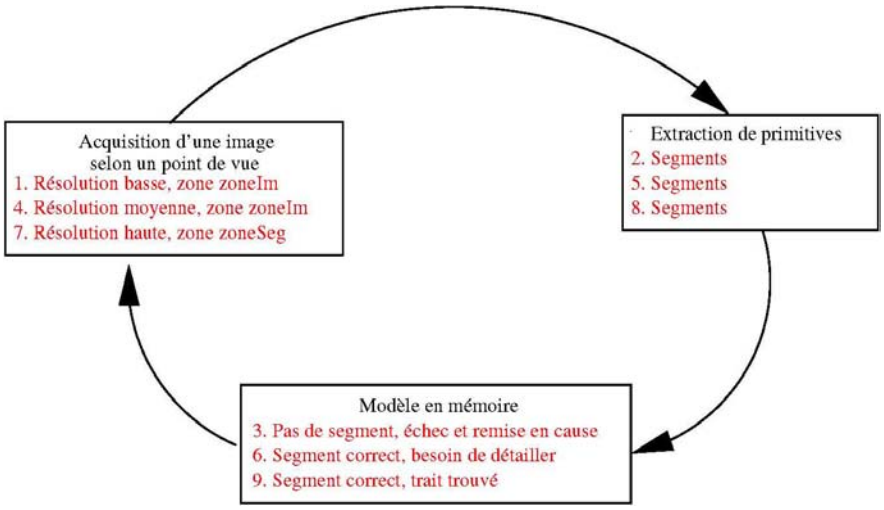


FIG. 7.21 – Exemple de cycle perceptif dû à une remise en cause

Attention visuelle

Comme nous l'avons montré dans le chapitre 1, le second constituant de la vision perceptive, après le cycle perceptif, est l'attention visuelle.

Nous rappelons que nous souhaitons utiliser les deux sortes d'attention visuelle : l'attention guidée par les éléments prégnants et l'attention guidée par un but.

Notre implémentation permet d'imiter l'attention guidée par des éléments prégnants. En effet, nous avons mis en place un formalisme, les calques perceptifs induits, dont le but est de faciliter l'utilisation d'objets construits indépendamment de connaissance liée à un type de document, selon des lois proches de l'organisation perceptive. Notre architecture permet la création d'autant de calques que de niveaux de perception à prendre en compte. Dans ce contexte, nous avons défini deux grammaires décrivant respectivement les lignes de texte et les traits comme des éléments structurels prégnants présents dans des documents.

La possibilité de décrire un document en étant guidé par un but est permise par l'utilisation du langage EPF pour décrire un objet à reconnaître. Nous l'avons déjà utilisée pour décrire les lignes de texte et les traits.

Nous allons montrer dans la partie suivante quatre grammaires dans le langage EPF, décrivant des structures de documents plus complexes. Ces grammaires sont des exemples d'utilisation de l'attention guidée par le but, basées en partie sur des éléments prégnants et décrits par une connaissance plus complexe.

Troisième partie

Apports pour la reconnaissance de la structure de documents

Introduction

Nous avons présenté le fonctionnement interne de la méthode DMOS-P, dont le but est de fournir un cadre pour la spécification de systèmes de reconnaissance de documents s'appuyant sur la vision perceptive. Pour l'illustrer, nous nous sommes basés sur des exemples ponctuels. Nous présentons maintenant des applications à grande échelle de notre méthode.

Ces travaux applicatifs ont un double objectif :

- valider la généricité et le pouvoir d'expression offerts par l'architecture de DMOS-P, par l'application de la méthode sur des problèmes variés à grande échelle et par son transfert industriel ;
- valider l'intérêt de la vision perceptive pour la reconnaissance de la structure de documents.

Pour chacune des applications présentées ci-dessous, nous avons créé une grammaire dans le langage EPF, représentant un mécanisme de coopération perceptive adapté à la nature de chaque type de documents. Ces grammaires s'appuyant sur la méthode DMOS-P ont ensuite été compilées de manière à produire automatiquement l'analyseur associé.

La première application, présentée dans le chapitre 8, se place dans le cadre d'une campagne d'évaluation nationale, le projet RIMES des ministères de la Recherche et de la Défense, dont une partie s'intéresse à la reconnaissance de la structure de courriers entrants manuscrits. Dans ce contexte, nous présentons les résultats obtenus avec notre approche perceptive, et présentons les intérêts de notre méthode vis-à-vis des approches proposées par d'autres équipes de recherche.

Nous présentons ensuite, de manière plus précise, les apports de la vision perceptive, en analysant les résultats obtenus avec DMOS-P sur un problème qui avait déjà été traités précédemment avec la méthode DMOS classique. Ainsi, le chapitre 9 présente les apports de la vision perceptive pour le traitement de documents d'archives : les registres de décrets de naturalisation.

Dans le chapitre 10, nous mettons en avant l'aspect générique de notre méthode en l'utilisant pour un problème différent : nous proposons en effet de positionner les lignes de base pour la reconnaissance de l'écriture manuscrite Bangla (un dialecte indien).

Le chapitre 11 présente enfin une application de segmentation de pages de journaux pour laquelle la vision perceptive des traits est particulièrement intéressante. Cette ap-

plication regroupe l'ensemble des concepts proposés dans DMOS-P. Cette application est le support d'un transfert industriel de la méthode DMOS-P qui permet de valider de manière externe l'architecture de DMOS-P.

Chapitre 8

Courriers manuscrits

La première application présentée consiste en une participation au projet RIMES : Recherche et Indexation de données Manuscrites et de fac-similES. Ce programme fait partie du projet Techno-vision du ministère de la recherche et du ministère de la défense. Un des aspects de ce projet consiste en la reconnaissance de la structure de courriers entrants manuscrits.

La participation à ce projet nous permet de proposer un premier mécanisme perceptif complet produit avec la méthode DMOS-P. Ainsi, dans ce chapitre, nous mettons en avant l'intérêt d'une approche grammaticale, et validons le principe du calque perceptif induit contenant les lignes de texte prégnantes. Enfin, nous comparons notre approche avec d'autres méthodes existant en reconnaissance de structure de documents.

Nous présentons donc le contexte du projet RIMES et la tâche de reconnaissance de courriers manuscrits. Puis nous introduisons notre processus de reconnaissance basé sur la vision perceptive, mis en œuvre par la méthode DMOS-P. Nous présentons ensuite les résultats obtenus avant de discuter des apports de la vision perceptive pour la reconnaissance de ce type de documents.

8.1 Contexte

8.1.1 Présentation du Projet RIMES

Le projet RIMES [GGC+06] a pour but de construire une grande base d'évaluation des techniques de reconnaissance de documents manuscrits. En effet, selon les organisateurs, « la communauté scientifique de l'analyse automatique de contenus manuscrits souffre d'un manque de bases de données à la fois cohérentes et de grande taille, ainsi que de métriques consensuelles ». Il était donc difficile avant, dans ce domaine, de comparer des méthodes sur des données communes.

Le projet RIMES a donc permis de construire une base composée de pages de courriers manuscrits fictifs et de pages de garde de fax. La seconde étape du projet consiste à mettre en place des métriques d'évaluation, puis à lancer des campagnes de tests permettant de comparer les résultats obtenus par différents laboratoires. Neuf laboratoires

de recherche français prennent part à ce projet en tant que participants à des tâches de reconnaissance.

Plusieurs tâches de reconnaissance sont proposées, couvrant ainsi l'essentiel des thèmes de l'analyse automatique de documents manuscrits :

- structuration sémantique,
- reconnaissance d'écriture manuscrite,
- reconnaissance de scripteur,
- reconnaissance de logo,
- extraction d'informations.

Dans la catégorie de la structuration sémantique, nous participons à la tâche de reconnaissance de la structure des courriers manuscrits.

8.1.2 Tâche de structuration de courriers manuscrits

La tâche de reconnaissance de la structure des courriers manuscrits consiste à localiser et à classer les 8 types de blocs suivants (figure 8.1) :

- coordonnées de l'expéditeur (en violet),
- coordonnées du destinataire (en turquoise),
- date et lieu (en rouge),
- objet (en vert clair),
- ouverture (en jaune),
- corps de texte (en orange),
- signature (en vert foncé),
- PS et pièce jointe (en bleu foncé).

Cette tâche doit faire face aux difficultés rencontrées dans les documents manuscrits non contraints. En effet, même si les usages français définissent la manière de structurer un courrier, cette structuration n'est pas toujours respectée dans les courriers manuscrits. Ainsi, nous devons traiter des documents à structure variable (figure 8.2).

La première difficulté est la segmentation de la page en blocs. Par exemple, considérons les quatre premières lignes en haut à droite des courriers 8.2(a) et 8.2(c). Avec une approche ascendante, ces lignes seraient probablement groupées en un seul bloc. Cependant, dans le cas du courrier 8.2(a), les trois premières lignes contiennent les coordonnées du destinataire, la quatrième la date et le lieu, alors que dans le courrier 8.2(c), les quatre lignes donnent les coordonnées du destinataire. Pour résoudre ce problème, il est indispensable de pouvoir prendre en compte la structure logique pour réaliser la segmentation physique.

De plus, les éléments ne sont pas tous présents dans chaque courrier. Ainsi, la date et le lieu sont présents uniquement sur les courriers 8.2(a) et 8.2(d) ; les coordonnées du destinataire sont absentes sur la figure 8.2(b). Nous devons donc pouvoir prendre en compte différentes configurations dans les pages de courrier.

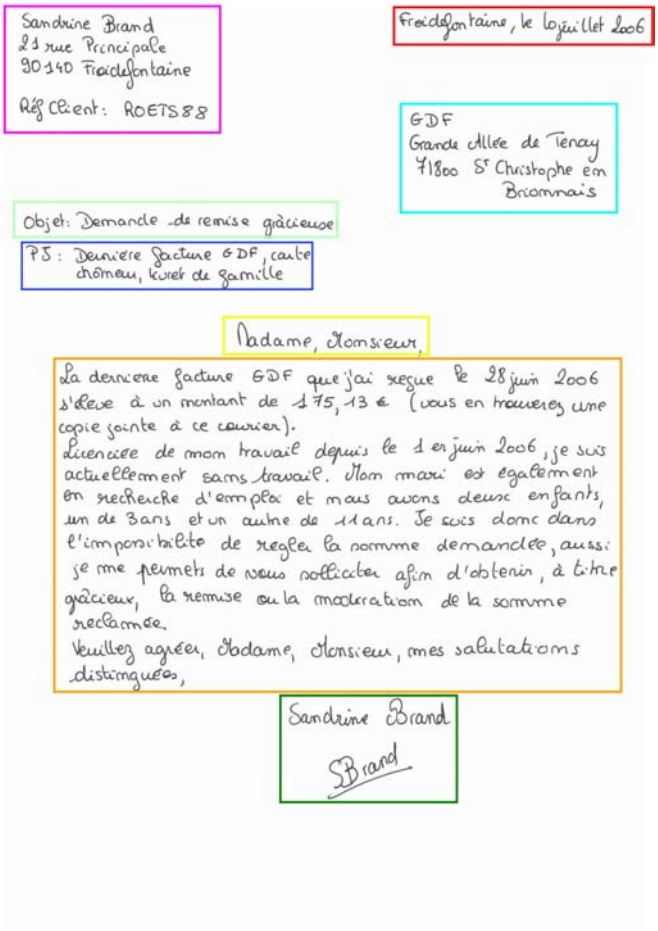
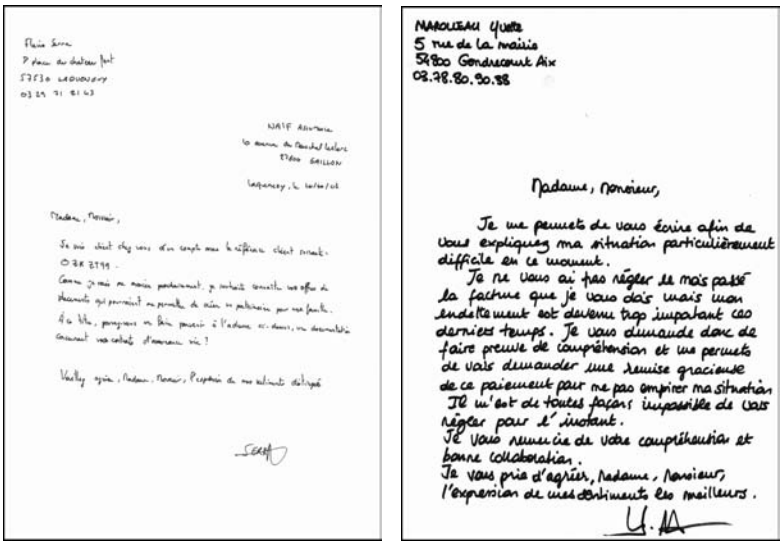
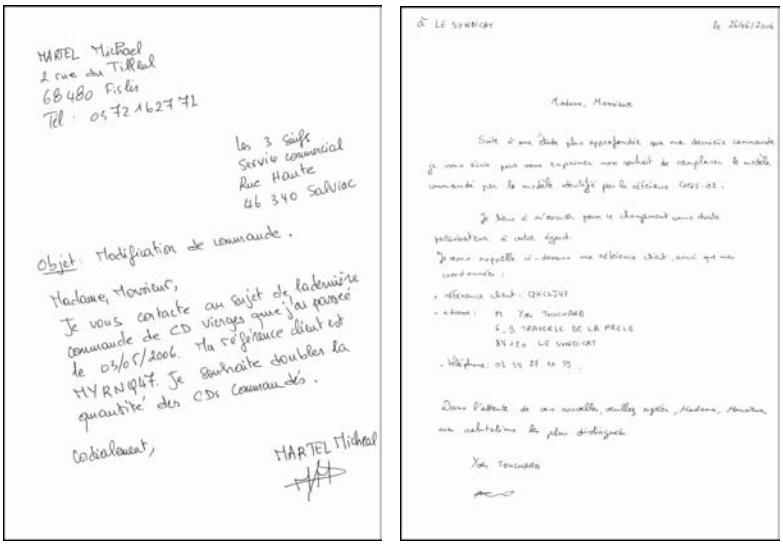


FIG. 8.1 – Blocs à localiser et à étiqueter dans les courriers manuscrits (voir les significations des couleurs dans la partie 8.1.2)



(a) (b)



(c) (d)

FIG. 8.2 – Exemples d'images de courriers manuscrits à structure variable

8.2 Processus de reconnaissance

Nous présentons le principe général de notre mécanisme de reconnaissance avant de détailler son implémentation grammaticale dans le langage EPF.

8.2.1 Principe général

Nous pouvons séparer le problème de reconnaissance en deux sous-problèmes :

1. la reconnaissance des lignes de texte,
2. l'organisation des lignes de texte en blocs.

Nous basons notre analyse sur l'entité de ligne de texte, décrite avec DMOS-P comme un élément prégnant (partie 7.1.1). Ces lignes de texte sont construites à partir des segments vus à basse résolution et des composantes connexes à haute résolution. Les calques perceptifs utilisés pour la reconnaissance des courriers manuscrits sont donc (figure 8.3) :

- un calque perceptif direct contenant les segments liés à une image à basse résolution,
- un calque perceptif direct contenant les composantes connexes liées à l'image initiale,
- un calque perceptif induit contenant les lignes de texte construites à partir des deux calques ci-dessus.

L'utilisation des lignes de texte contenues dans le calque induit simplifie l'expression de la structure. En utilisant ces informations, la description de l'organisation des lignes de texte en blocs est réalisable avec les concepts classiques proposés dans la méthode DMOS.

Les blocs de texte contenus dans le document sont décrits comme un agencement particulier des lignes de texte. Il s'agit d'une description guidée par un but. Notre analyse est basée sur les règles d'écriture couramment utilisées en français : les coordonnées de l'expéditeur sont en haut à gauche de la page, les coordonnées du destinataire, la date et le lieu sont placés en haut à droite, l'objet est placé avant l'ouverture, et le corps de texte est terminé par la signature. Cependant, comme nous l'avons montré précédemment, nous avons dû gérer une grande variété de configurations, et l'absence possible de chacun des éléments.

Afin de limiter les ambiguïtés dans l'analyse, nous avons choisi un ordre spécial pour la reconnaissance. En effet, la signature, le corps de texte et l'ouverture sont presque toujours présents dans un courrier. Nous avons donc choisi de démarrer l'analyse par le bas du document, en cherchant successivement, de bas en haut, une signature, le corps de texte puis l'ouverture. Nous pouvons alors nous concentrer sur la partie supérieure du document pour trouver les éléments restants.

Toute cette connaissance est décrite grâce au langage EPF.

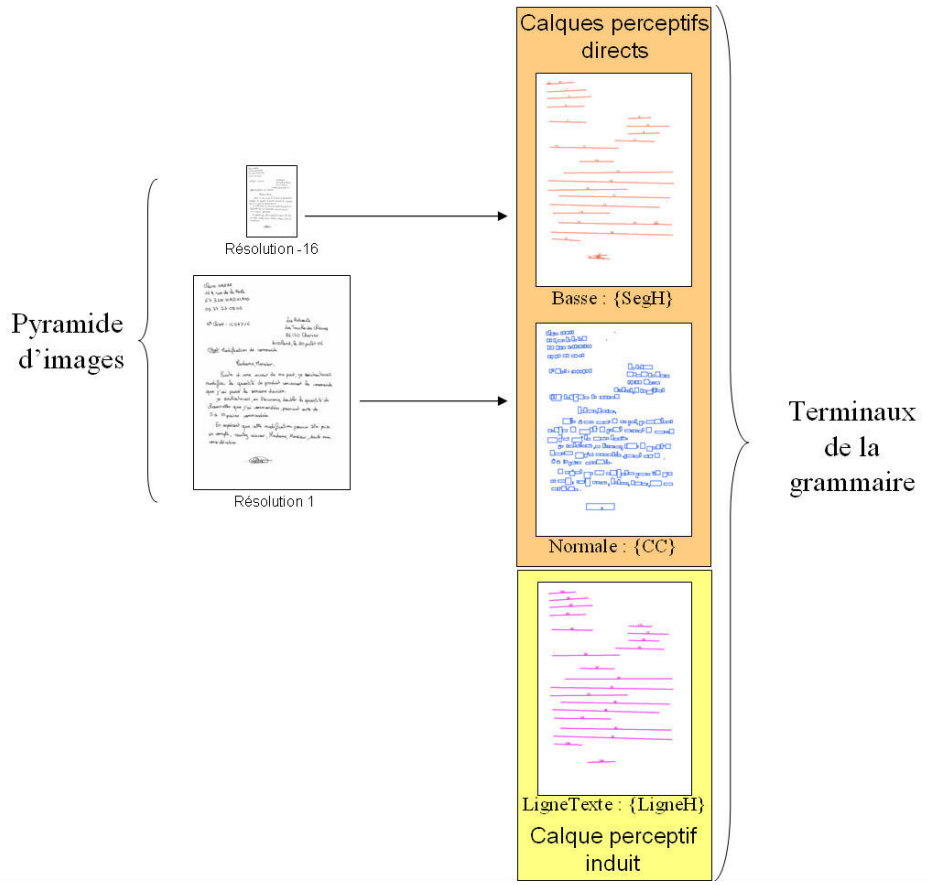


FIG. 8.3 – Calques perceptifs utilisés pour la reconnaissance des courriers manuscrits

8.2.2 Implémentation avec le langage EPF

L'implémentation réalisée avec le langage EPF met en évidence d'une part les avantages d'une description grammaticale selon les principes de DMOS, et d'autre part les apports des calques perceptifs contenus dans DMOS-P.

8.2.2.1 Description grammaticale

Le prédicat principal de reconnaissance du courrier, `courrier`, illustre l'ordre dans lequel nous avons choisi d'étudier les éléments afin de diminuer les ambiguïtés.

Nous rappelons que, dans le langage EPF, les attributs utilisés par les règles commencent par une majuscule et peuvent être, selon les cas, synthétisés ou hérités.

```
courrier ::=
     %Début de la structure
     AT_ABS(basPage)
     signat Sign &&
     ps Ps &&
     blocDeTexte Ouv Corps &&
     coorExp Exp &&
     dateLieu Date &&
     coorDes Dest &&
     objetRef Obj Exp2.
```

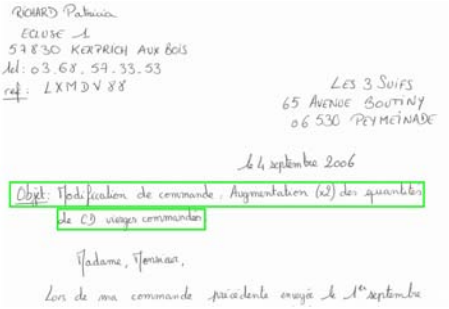
L'analyse consiste à reconnaître successivement la signature `Sign` avec la règle `signat`, le post-scriptum `Ps` avec la règle `ps`, le corps de texte `Corps` et l'ouverture `Ouv` avec la règle `blocDeTexte`, les coordonnées expéditeur `Exp` avec la règle `coorExp`, la date et le lieu `Date` avec la règle `dateLieu`, les coordonnées destinataire `Dest` avec la règle `coorDes` et l'objet `Obj` qui peut être lié par la règle `objetRef` à des informations supplémentaires sur les coordonnées de l'expéditeur `Exp2`.

Afin de montrer l'intérêt de notre approche dans le cas de structures variables, nous proposons de détailler la règle `objetRef`. Cette règle a pour but de reconnaître deux éléments : l'objet et la référence du client (partie des coordonnées de l'expéditeur). Ces deux éléments sont localisés dans la partie gauche du document, au dessus de l'ouverture, et sont parfois proche l'un de l'autre. Leur présence est facultative, et ils apparaissent dans un ordre variable. Il existe donc plusieurs combinaisons possibles, dont quelques unes sont illustrées sur la figure 8.4.

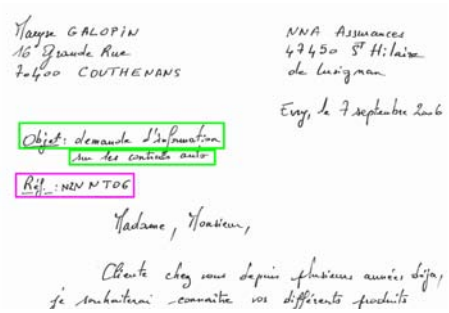
Dans les courriers 8.4(a) et 8.4(b), on trouve un objet mais pas de référence client. Par contre on peut trouver une référence client dans les courriers 8.4(c) et 8.4(d), respectivement sous et sur le champ de l'objet, lui-même constitué d'une ou deux lignes. Le courrier 8.4(e) ne contient ni référence client ni objet. Ces exemples montrent un cas précis pour lequel le choix de segmentation est fortement corrélé à l'étiquetage.

Nous émettons des hypothèses générales sur la structure permettant de différencier un objet d'une référence client :

(a) Une ligne d'objet (règle 6)



(b) Objet sur deux lignes (règle 3)



(c) Objet sur deux lignes suivi d'une référence client (règle 2)

(d) Réference client suivie d'une ligne d'objet (règle 5)

(e) Pas de référence client ni d'objet (règle 8)

FIG. 8.4 – Reconnaissance de l'objet (en vert) et de la référence client (en violet) : application d'une règle différente selon l'organisation spatiale des lignes

- une référence client est une ligne courte ;
- un objet est composé d'une ligne longue, voire de deux lignes ;
- lorsqu'un objet est composé de deux lignes, la deuxième ligne est décalée avec un alinéa.

Nous décrivons ces différentes combinaisons possibles au niveau de la règle `objetRef`.

Un bloc `objetRef` peut être composé de trois lignes, commençant soit par une référence, soit par un objet. Nous en déduisons deux règles possibles :

```
(1) objetRef [Objet1|Objet2] Ref::=
          ligneCourteReference Ref &&
          AT(sousLigne) &&
          ligneLongueObjet Objet1 &&
          AT(sousLigne) &&
          ligneComplementObjet Objet2.

(2) objetRef [Objet1|Objet2] Ref ::=
          ligneLongueObjet Objet1 &&
          AT(sousLigne) &&
          ligneComplementObjet Objet2 &&
          AT(sousLigne) &&
          ligneCourteReference Ref.
```

Un `objetRef` peut également être constitué de deux lignes de texte, soit deux lignes d'objet, soit une ligne d'objet et une référence client. Nous complétons donc la description avec trois nouvelles règles :

```
(3) objetRef [Objet1|Objet2] ::=
          ligneLongueObjet Objet1 &&
          AT(sousLigne) &&
          ligneComplementObjet Objet2.

(4) objetRef Objet Ref ::=
          ligneLongueObjet Objet &&
          AT(sousLigne) &&
          ligneCourteReference Ref.

(5) objetRef Objet Ref ::=
          ligneCourteReference Ref &&
          AT(sousLigne) &&
          ligneLongueObjet Objet.
```

Sinon, s'il y a une seule ligne de texte, nous décrivons les règles suivantes :

```
(6) objetRef Objet [] ::=
          ligneLongueObjet Objet.
```

```
(7) objetRef [] Ref ::=
        ligneCourteReference Ref.
```

La dernière règle décrit l'absence totale d'objet et de référence client :

```
(8) objetRef [] [].
```

Dans ces règles, `ligneCourteReference` produit une ligne `Ref` qui porte l'étiquette des *Coordonnées expéditeur*. Les règles `ligneLongueObjet` et `ligneComplementObjet` produisent des lignes ayant l'étiquette *Objet*. La différence entre une `ligneCourteReference` et une `ligneComplementObjet` vient de la présence d'un alinéa au début de la `ligneComplementObjet`.

Pour la reconnaissance, l'analyseur essaie d'appliquer successivement chacune des règles, dans l'ordre donné, jusqu'à ce qu'une réussisse. Ainsi, sur la figure 8.4, la légende précise le numéro de la première règle ayant réussi dans chacun des cas. Si rien n'a été reconnu, c'est la règle `(8)` qui réussit.

Nous pouvons noter que la segmentation finale en blocs, consistant à regrouper des lignes de texte dans une même classe, est décidée après la classification des lignes. Ceci n'est pas le cas dans les méthodes usuelles de la littérature où la classification est réalisée après la segmentation.

Cet exemple met également en avant l'adaptabilité de notre système. En effet, lorsqu'on rencontre une nouvelle configuration d'un document, il suffit d'ajouter une règle la décrivant. Cet ajout est réalisé sans remettre en cause le fonctionnement des règles précédentes.

8.2.2.2 Utilité des calques de DMOS-P

Les règles de grammaire présentées ci-dessus sont entièrement basées sur des non-terminaux décrivant les lignes de texte. Grâce au calque perceptif contenant les lignes de texte vues comme des éléments prégnants, la reconnaissance de ces non-terminaux s'exprime de manière très simple. Par exemple, la recherche d'une ligne assez longue est réalisée par un simple appel à l'extracteur de terminaux `TERM_SEG`, dans le calque associé aux lignes de texte.

```
ligneTexteAssezLongue X L ::=
     TERM_SEG (condAssezLongueLigne X) noCondS ligne L.
```

L'utilisation de pré et de post-conditions permet d'augmenter le pouvoir d'expression pour la description des lignes à reconnaître.

L'intérêt de cette approche est qu'elle permet de simplifier considérablement la description des blocs de texte. En effet, la reconnaissance des lignes étant réalisée de manière séparée, leur analyse est totalement transparente lors de la description de l'organisation des courriers. De plus, la reconnaissance des lignes de texte étant réalisée par combinaison de différents niveaux de perception, leur reconnaissance est considérée comme fiable, ce qui évite d'avoir besoin de prévoir des cas de mauvaise détection des lignes dans l'organisation des blocs. Par exemple, une ligne en biais va être correctement reconnue dans le calque perceptif, et pourra être considérée comme une ligne ordinaire au

niveau de la reconnaissance des blocs.

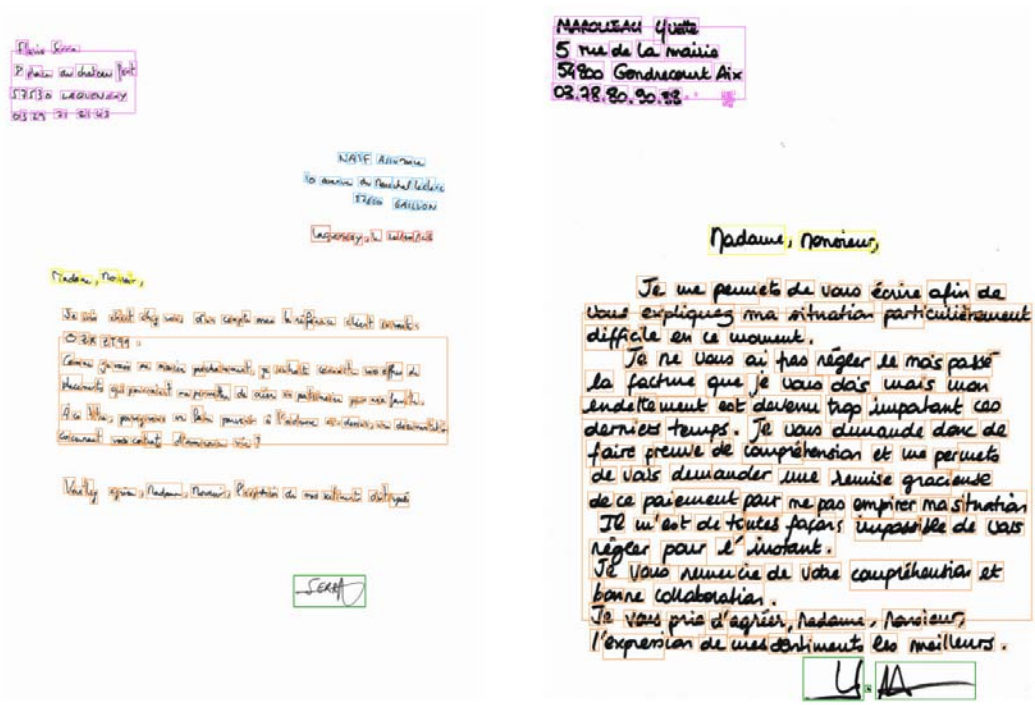
L'utilisation de notre méthode DMOS-P permet la simplification de l'écriture de la grammaire grâce à l'utilisation des calques perceptifs.

Une version plus complète de la grammaire décrivant les courriers manuscrits est présentée dans l'annexe B.

Nous avons donc produit un système perceptif complet dans le langage EPF décrivant les différentes configurations d'une page. La force de cette méthode réside dans la possibilité de remettre en cause la segmentation et l'identification des blocs au fur et à mesure de l'analyse.

La figure 8.5 présente des exemples de structures analysées grâce à notre méthode. Le code couleur utilisé est celui décrit dans la partie 8.1.2.

Nous présentons maintenant les résultats obtenus dans le cadre du concours RIMES.



(a) (b)

FIG. 8.5 – Exemples de résultats obtenus avec DMOS-P

8.3 Applications

8.3.1 Base de documents

Pour la campagne de reconnaissance des courriers manuscrits, le projet RIMES a fourni une base de 1050 images de courriers pour l'apprentissage des systèmes. La première compétition, en juin 2007 a eu lieu sur une base de 100 nouvelles images. Une seconde compétition a eu lieu en juin 2008 sur une autre base de 100 images.

Ces images sont stockées en PNG et ont une taille d'environ 2500*3500 pixels. Elles ont été manuellement annotées par RIMES. Ainsi, chaque zone à reconnaître est représentée par les coordonnées de son rectangle englobant et la classe associée. Les annotations sont stockées comme une liste de boîtes dans un fichier XML associé à chaque image. Ces fichiers de vérité terrain sont appelés *validation*.

L'objectif est de produire en résultat de notre système une liste de boîtes étiquetées, décrivant le document, dans un fichier XML appelé *hypothèse*. Les résultats sont ensuite calculés par comparaison des fichiers *hypothèse* et *validation*. Le format des fichiers XML utilisés est imposé par le projet RIMES.

8.3.2 Métriques utilisées

Nous distinguons la métrique utilisée par le concours RIMES (existant avec deux variantes) d'une métrique secondaire que nous avons proposée pour mieux évaluer les points forts et faibles de notre méthode.

8.3.2.1 Métrique primaire du concours RIMES

La métrique RIMES permet de produire un taux global d'erreur sur l'étiquetage complet d'un ensemble de documents.

Ce taux d'erreur est calculé en comparant les étiquettes attribuées à chacun des pixels de l'image, avec celles contenues dans les vérités terrain. Afin de donner plus d'importance aux pixels noirs qu'aux pixels blancs, le taux d'erreur est pondéré par le niveau de gris des pixels. Le taux d'erreur global représente donc le taux de niveaux de gris de l'image ayant été mal étiquetés. C'est cette métrique que nous nommons plus loin *métrique-Rimes-2007*.

Après lancement à grande échelle, il apparaît néanmoins que certains pixels du fond, étant légèrement grisés, causent anormalement une erreur dans l'étiquetage et la reconnaissance liée aux niveaux de gris. La métrique est donc revue en 2008, pour prendre en compte uniquement les pixels noirs, dans une image binarisée. Dans cette version, le taux d'erreur global représente le taux de pixels noirs de l'image ayant été mal étiquetés. C'est cette métrique que nous nommons plus loin *métrique-Rimes-2008*.

8.3.2.2 Métrique secondaire

Afin de nous relier aux métriques habituellement utilisées en classification, nous proposons d'évaluer des mesures de précision et de rappel. Dans notre cas, ces mesures

sont définies de la manière suivante :

$$Rappel = \frac{Nombre\ de\ pixels\ corrects\ classés}{Nombre\ de\ pixels\ attendus}$$

$$Précision = \frac{Nombre\ de\ pixels\ corrects\ classés}{Nombre\ de\ pixels\ classés}$$

Le rappel permet donc d'évaluer le taux de reconnaissance, tandis que la précision permet d'évaluer la pertinence des résultats ramenés.

D'autre part, afin d'identifier les points forts et les points faibles de notre méthode, nous proposons d'appliquer ces mesures en séparant les résultats classe par classe.

Munis de ces deux métriques, nous présentons les résultats obtenus par notre méthode.

8.3.3 Résultats

Nous présentons dans un premier temps les résultats officiels obtenus lors de la participation au concours RIMES avant de présenter des résultats à plus grande échelle.

8.3.3.1 Participation aux concours RIMES

Nous avons participé à deux évaluations organisées par RIMES, l'une en juin 2007, l'autre en juin 2008. Dans chacun des cas, il s'agissait d'appliquer notre méthode sur une base nouvelle de 100 images.

Les résultats obtenus sont présentés dans les tableaux 8.1 et 8.2. Les noms des autres participants ont volontairement été masqués puisque les résultats n'ont pas encore été rendu publics de manière nominative.

Système	Taux d'erreur global
DMOS-P	8.58%
Laboratoire X	10.05%

TAB. 8.1 – Résultats obtenus par les différents participants lors de l'évaluation RIMES 2007, *métrique-Rimes-2007*

Système	Taux d'erreur global
Laboratoire X	8.53%
DMOS-P	8.97%
Laboratoire Y	12.62%
Laboratoire Z	12.88%

TAB. 8.2 – Résultats obtenus par les différents participants lors de l'évaluation RIMES 2008, *métrique-Rimes-2008*

Ces résultats montrent que d'un point de vue global, notre méthode se place au niveau des approches proposées par les autres laboratoires. Notons ici qu'on ne peut pas comparer directement les résultats fournis par RIMES en 2007 et en 2008, puisque la *métrique-Rimes-2007* était différente de la *métrique-Rimes-2008*.

Afin d'effectuer cette comparaison, nous présentons dans le tableau 8.3 nos résultats calculés pour 2007 et 2008 avec la *métrique-Rimes-2008*. Dans ce tableau, le résultat de 7.40% obtenu en 2007 correspond à la valeur de 8.58% (tableau 8.1) qui était calculée avec la *métrique-Rimes-2007*.

Concours	Juin 2007		Juin 2008
Métrique	2007	2008	2008
Nombre d'images	100	100	100
Taux d'erreur global	8.58%	**7.40%**	**8.98%**

TAB. 8.3 – Comparaison des résultats obtenus par la méthode DMOS-P lors des deux évaluations RIMES, métrique 2008

La différence entre les résultats des deux évaluations vient du fait que 100 images ne sont pas suffisantes pour obtenir des résultats significatifs. En effet, à cette échelle, il suffit d'une ou deux images particulièrement difficiles pour modifier le résultat global. D'autre part, il nous semble important de connaître le taux d'erreur classe par classe.

Nous présentons donc des résultats plus complets dans la partie suivante.

8.3.3.2 Résultats à plus grande échelle

Afin de présenter des résultats à plus grande échelle, nous avons appliqué notre méthode sur l'ensemble des images fournies par le projet RIMES. Parmi ces images, 300 ont été regardées une à une, manuellement, pour pouvoir définir les règles de grammaire à écrire dans le langage EPF. Ces images peuvent donc être considérées comme une base d'apprentissage de la grammaire. Les 950 images restantes forment la base de test. Nous présentons dans le tableau 8.4 le détail des résultats obtenus avec notre méthode, classe par classe, sur les bases d'apprentissage, de test, puis sur la base globale.

Les résultats finaux principaux sont situés en bas à droite du tableau. Toutes classes confondues, on obtient un rappel de 92.0%, soit un taux d'erreur de 8,0%. Ce taux sur 1250 images peut être comparé aux résultats des évaluations 2007 et 2008 portant chacune sur 100 images (tableau 8.3). Il confirme l'idée que la base de 2007 était particulièrement facile ; celle de 2008 plus compliquée.

La colonne *Pixels concernés* donne la répartition des pixels dans la base globale ; la colonne *Pages concernées* donne le nombre de pages présentant chacune des classes. Ainsi, le corps de texte représente la majorité des pixels, 61.5% ; il est présent dans les 1250 images. Le tableau de résultats montre qu'on obtient de bons taux de reconnaissance pour le corps de texte, qui est très présent, et pour les coordonnées de l'expéditeur. En effet, celles-ci sont situées très souvent en haut de la gauche de la page et présentent assez peu d'ambiguïtés.

Classe	Pages concernées	Pixels concernés	Base d'apprentissage 300 images		Base de test 950 images		Base complète 1250 images	
			Rappel	Précision	Rappel	Précision	Rappel	Précision
Corps de texte	1250	61.5%	96.3%	98.4%	96.3%	97.6%	96.3%	97.8%
Expéditeur	1242	15.1%	94.0%	93.1%	92.1%	92.0%	92.5%	92.2%
Destinataire	1173	9.1%	85.1%	92.4%	86.1%	91.3%	85.9%	91.5%
Signature	1239	4.1%	86.1%	89.8%	88.0%	90.8%	87.5%	90.5%
Objet	698	4.0%	76.8%	70.3%	66.5%	71.0%	68.8%	70.8%
Date, Lieu	953	3.2%	77.0%	82.4%	75.9%	84.8%	76.2%	84.2%
Ouverture	1222	2.9%	81.9%	77.2%	76.4%	74.1%	77.7%	74.8%
PS/PJ	35	0.2%	55.0%	40.9%	9.6%	16.5%	17.9%	24.8%
Total	-	100%	92.6%	94.2%	91.8%	93.7%	**92.0%**	**93.8%**

TAB. 8.4 – Résultats sur 1250 courriers manuscrits

Le taux de reconnaissance plus faible pour les autres classes est globalement lié au taux de représentativité de chacune des classes. Concernant la classe PS/PJ, notre méthode obtient un faible taux de reconnaissance (17.9%). En effet, compte tenu de la faible représentativité de cette classe (0.2% des pixels, 35 exemples sur 1250 courriers), nous avons assez peu étudié la structure de cette classe. Les résultats seraient cependant améliorables en précisant quelques règles de description pour ces éléments.

Nous pouvons également noter que les résultats de la base de test sont proches de ceux obtenus sur la base d'apprentissage.

8.3.3.3 Expérience complémentaire

Les résultats ci-dessus permettent de comparer DMOS-P avec d'autres approches, mais pas d'évaluer l'apport de la vision perceptive. Afin d'évaluer plus précisément cet apport pour la reconnaissance des lignes de texte, nous proposons une expérience complémentaire.

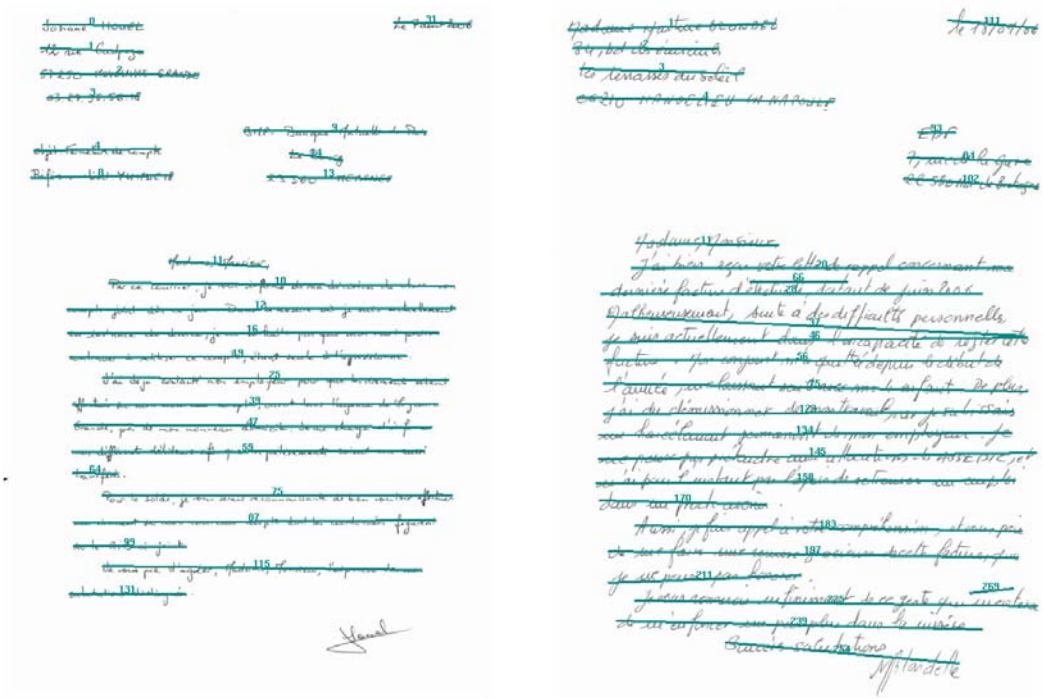
Ainsi, nous réutilisons une grammaire décrivant les lignes de texte de manière monorésolution, comme une succession horizontale de composantes connexes vues dans l'image de résolution initiale. Cette description a été développée précédemment dans l'équipe Imadoc, pour la description de documents d'archives.

Nous créons donc un nouveau système de reconnaissance de courriers, qui se base non plus sur les lignes de texte reconnues de manière perceptive, mais sur les lignes de texte décrites de manière monorésolution. Ceci est réalisé de manière très simple, grâce au formalisme des calques perceptifs. En effet, pour que la grammaire décrivant les courriers se base sur des lignes de texte construites en monorésolution, il suffit de modifier la création du contenu du calque perceptif `LigneTexte`. Cet échange est transparent d'un point de vue de la grammaire décrivant les courriers (présentée dans la partie 8.2.2.1).

La figure 8.6 présente des exemples de lignes de texte détectées avec l'approche monorésolution. Avec cette méthode, les lignes de texte sont correctement détectées dans le cas où les lignes sont horizontales, sans courbure ni inclinaison (figure 8.6(a)). En revanche, la méthode de recherche de composantes connexes alignées pose des problèmes dès que le texte est plus dense (figure 8.6(b)) ou que les lignes sont inclinées (figure 8.8(a)). Notons que sur ces mêmes images, les lignes de texte sont correctement reconnues avec notre approche perceptive (figure 8.7).

Lorsque les lignes de texte sont mal reconnues, cela entraîne des erreurs de reconnaissance pour la structure des courriers. La figure 8.8 présente un exemple de courrier pour lequel les lignes de texte sont mal perçues en monorésolution (figure 8.8(a)). Le résultat produit par la grammaire décrivant les courriers (figure 8.8(b)) est erroné car cette grammaire n'est pas prévue pour gérer des lignes de texte morcelées. Pour cet exemple en revanche, la reconnaissance des lignes de texte avec notre approche perceptive est correcte (figure 8.8(c)) ; le courrier est donc globalement bien reconnu (figure 8.8(d)).

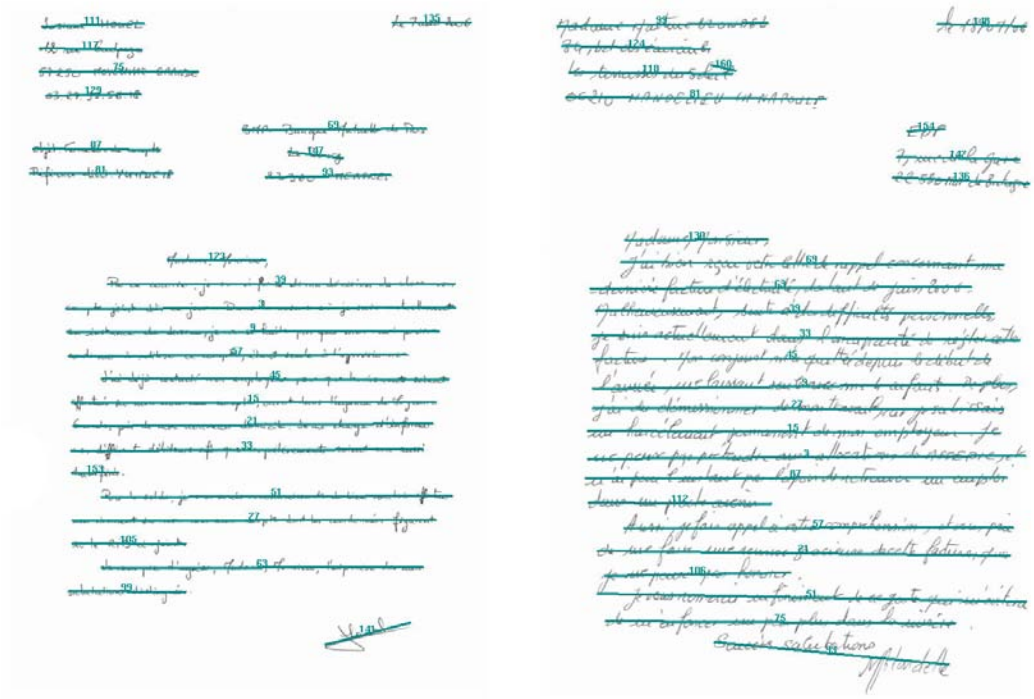
Nous avons évalué de manière plus globale l'impact des lignes de texte sur la reconnaissance des courriers. Le tableau 8.5 présente les résultats comparatifs obtenus pour



(a) Lignes correctement détectées

(b) Difficultés dues à la grande densité du texte

FIG. 8.6 – Lignes extraites par l'approche monorésolution créée précédemment dans DMOS



(a) Lignes correctement détectées

(b) Lignes correctement détectées malgré la densité

FIG. 8.7 – Lignes extraites par notre approche perceptive

(a) Lignes de texte monorésolution : mauvaise gestion de la pente

(b) Résultat associé en monorésolution : erreurs dans les blocs

(c) Lignes de texte perceptives : bonne perception de la pente

(d) Résultat associé à la vision perceptive : aucune erreur

FIG. 8.8 – Comparaison des résultats selon la méthode de reconnaissance des lignes de texte

la reconnaissance de 1250 courriers en se basant sur :

- les lignes de texte reconnues de manière perceptive,
- les lignes de texte reconnues en monorésolution comme des alignements de composantes connexes.

Classe	Approche perceptive	Monorésolution
Corps de texte	96.3%	81.4%
Expéditeur	92.5%	90.1%
Destinataire	85.9%	85.7%
Signature	87.5%	87.5%
Objet	68.8%	61.7%
Date, Lieu	76.2%	71.8%
Ouverture	77.7%	52.5%
PS/PJ	17.9%	16.2%
Total	**92.0%**	**81.3%**

TAB. 8.5 – Comparaison des taux de rappel (reconnaissance) selon la méthode de reconnaissance utilisée pour les lignes de texte, sur 1250 images

Ce tableau permet de mettre en avant l'intérêt de l'approche perceptive pour la reconnaissance des lignes de texte. En effet, avec cette méthode, on observe 92.0% de reconnaissance des courriers, contre seulement 81.3% avec l'approche monorésolution. Cette baisse de taux provient principalement des difficultés de reconnaissance du corps de texte qui contient davantage de lignes longues, pentues ou courbes, lesquelles sont plus difficiles à reconnaître en monorésolution.

Il serait vraisemblablement possible d'améliorer les résultats obtenus avec l'approche monorésolution : il faudrait pour cela adapter la grammaire des courriers pour pouvoir prendre en compte le fait que les lignes de texte puissent être mal reconnues. Cela compliquerait donc l'expression de cette grammaire, et nécessiterait un temps de mise au point important. L'autre possibilité consisterait à essayer de mieux décrire les alignements possibles entre composantes connexes dans la version monorésolution. Cependant, il serait difficile d'envisager tous les cas d'alignements possibles, et cette description perdrait en généricité.

Même si la grammaire décrivant les lignes de texte en monorésolution n'a pas été adaptée spécifiquement au traitement des courriers, cette expérience mène à plusieurs conclusions. L'approche perceptive permet de reconnaître les lignes de texte avec beaucoup plus de précision et de confiance que l'approche monorésolution. La description des courriers s'en trouve donc largement simplifiée, puisqu'elle n'a pas besoin de prendre en compte des éventuelles erreurs dans la reconnaissance des lignes de texte.

De plus, les temps de calcul présentés dans le tableau 8.6 montrent que l'approche perceptive pour la reconnaissance des lignes de texte est 10 fois plus rapide que la version monorésolution. Ceci est lié à la diminution de la combinatoire provoquée par la vision perceptive : l'utilisation de la vision globale permet de définir un contexte qui guide la

recherche des alignements de composantes connexes à haute résolution.

	Approche perceptive	Monorésolution
Temps d'exécution moyen par image	4.85 sec.	51.40 sec.

TAB. 8.6 – Comparaison des temps d'exécution selon la méthode de reconnaissance des lignes de texte, sur 1250 images

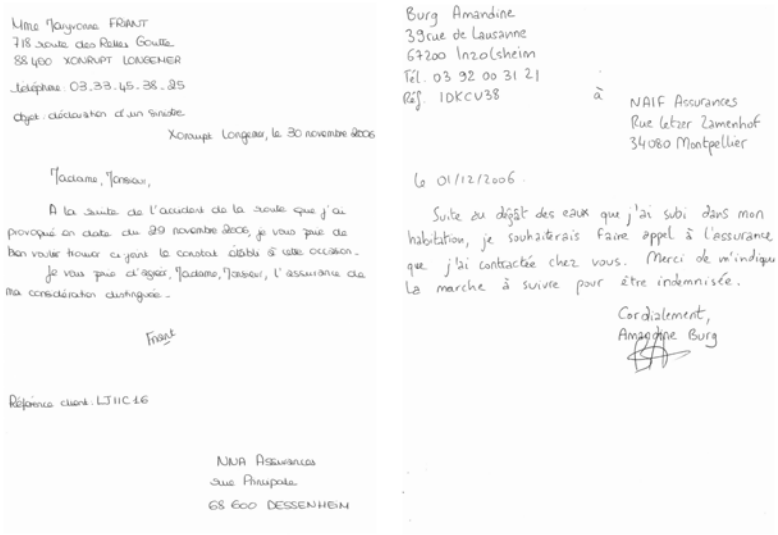
8.4 Discussion

Nous discutons maintenant des limites de notre approche, comparées aux approches proposées dans la littérature. Nous mettons ensuite en avant les avantages de la vision perceptive.

8.4.1 Limites actuelles de notre approche structurelle

Nous pouvons noter deux sources d'erreurs liées à l'approche grammaticale. La structure peut être mal reconnue lorsque l'on rencontre un cas qui n'a pas été prévu par les règles de grammaire. La figure 8.9(a) présente par exemple un courrier dans lequel les coordonnées du destinataire sont localisées en bas à droite. Cependant, si ce cas apparaissait plus fréquemment, il serait facile d'ajouter une règle permettant de décrire ce cas dans la grammaire.

Le second cas d'erreur est rencontré lorsque la structure seule ne suffit plus à classer une ligne de texte. Par exemple, sur la figure 8.9(b), la date est positionnée à l'emplacement habituel de l'ouverture. Sans reconnaissance du contenu, il n'est donc pas possible de classer cette date correctement. Pour répondre à ce problème, nous avons amorcé quelques travaux pour introduire de la connaissance en utilisant un classifieur qui étiquette les composantes connexes comme *chiffre* ou *non chiffre*. En effet, la méthode DMOS-P permet une communication aisée avec un classifieur : en fonction du type d'informations souhaitées, la description en EPF permet de faire appel à des segmenteurs ou des classifieurs adaptés au problème étudié, tout en ajoutant la connaissance du contexte local de l'analyse. Dans le cas des courriers, la présence d'un chiffre permet de distinguer par exemple un champ ouverture d'un champ de date. Cependant, les premiers travaux montrent qu'il est difficile pour le classifieur de prendre une décision entièrement hors contexte. De plus, nous sommes confrontés à des problèmes de segmentation : une composante connexe correspond en effet parfois à plusieurs chiffres liés. Ces travaux n'ont donc pas encore produit de résultats satisfaisants mais ils sont à poursuivre en améliorant les performances des classifieurs et des segmenteurs.



(a) Position inhabituelle pour les coordonnées destinataire

(b) Date positionnée à la place de l'ouverture

FIG. 8.9 – Images de courriers difficiles pour notre méthode

8.4.2 Autres méthodes utilisées

Les autres méthodes proposées par les participants au projet RIMES sont basées principalement sur un apprentissage statistique, utilisant par exemple les champs de Markov [LGGP08]. Ces méthodes se basent notamment sur les probabilités de transition entre des sites étiquetés de manière homogène.

Nous pouvons noter plusieurs points faibles de ces méthodes.

D'une part, ces méthodes ne peuvent pas facilement gérer des cas qui apparaissent peu fréquemment dans la base d'apprentissage (exemple de la figure 8.9(a)). Dans notre cas en revanche, nous avons montré qu'il suffit d'ajouter une règle de grammaire en dernière position pour pouvoir traiter une nouvelle configuration. Avec les méthodes statistiques, cette nouvelle configuration ne pourra pas être correctement traitée si la base d'apprentissage ne contient pas suffisamment d'exemples.

Le cas présenté sur la figure 8.9(b) est également une difficulté pour les méthodes basées sur la localisation statistique, puisque la base d'apprentissage aura permis d'apprendre le site contenant ici la date comme celui de l'ouverture.

D'autre part, les méthodes statistiques proposent un étiquetage pixel par pixel, ce qui peut entraîner l'étiquetage de deux pixels d'un même caractère avec des classes différentes, ce qui provoque des résultats très incohérents. Un exemple de ce type d'erreur est présenté sur la figure 8.10.

Enfin, ces méthodes souffrent de difficultés pour introduire de la connaissance. Par

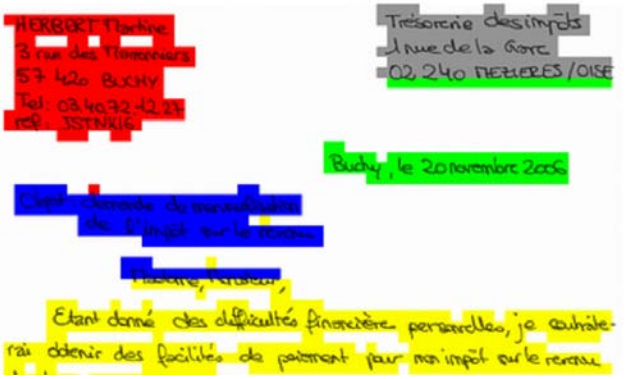


FIG. 8.10 – Exemple d'étiquetage incohérent avec une méthode statistique : des pixels d'une même composante connexe ont une étiquette différente

exemple, il serait difficile de préciser que l'ouverture est constituée d'une seule ligne de texte courte. Il serait également plus complexe que dans une méthode grammaticale d'introduire un lien avec un reconnaisseur de caractères ou de mots.

8.4.3 Intérêts de notre méthode

Face aux problèmes rencontrés par les autres méthodes, notre approche présente plusieurs avantages, liés à l'aspect grammatical et perceptif.

Une approche grammaticale permet la description intuitive du document, et la gestion d'un grand nombre de configurations possibles. Ainsi, même si une configuration est assez peu représentée, elle peut être correctement reconnue grâce à une règle appropriée. Ceci est permis par l'utilisation de la méthode DMOS.

Notre approche perceptive permet d'appréhender le document comme un agencement particulier de lignes de texte. Grâce à la méthode DMOS-P, les lignes de textes sont considérées ici comme des éléments prégnants dont la construction n'est pas remise en cause. Ceci est permis par la grande qualité de la perception des lignes de texte avec notre approche multirésolution.

Grâce au mécanisme de calque perceptif, la description du document est largement simplifiée puisqu'elle n'a pas besoin de prendre en compte les variations internes aux lignes de texte. De plus, l'utilisation de la vision perceptive pour la description des lignes de texte permet une meilleure reconnaissance et facilite donc la description des courriers manuscrits.

Enfin, notre méthode utilise la ligne de texte comme entité de base, ce qui permet de classer systématiquement toute la ligne de texte avec la même étiquette, et évite ainsi les erreurs d'incohérence produites par les méthodes statistiques utilisées par les autres participants.

8.4.4 Validation de la méthode DMOS-P

La description des courriers manuscrits constitue une première validation de la méthode DMOS-P. En effet, pour ces documents, nous avons mis en œuvre un mécanisme de coopération perceptive adapté au but applicatif. Ceci a pu être réalisé de manière simple, grâce au formalisme des calques perceptifs directs et induits, et à l'existence des éléments prégnants que sont les lignes de texte.

Ce mécanisme perceptif illustre l'intuition évoquée dans la partie 3.2.2.1 : lorsque l'information structurelle est physiquement diffuse, la vision perceptive permet de *reconstituer* la structure du document. Ceci est notamment permis par la synthèse d'éléments prégnants, analysés dans les résolutions les plus adaptées, qui facilitent la description de l'organisation globale du document.

Enfin, nous pouvons noter que notre description doit théoriquement fonctionner de manière transparente pour la reconnaissance de courriers imprimés, dans la mesure où les lignes de texte imprimées sont détectées correctement comme des éléments prégnants. L'application que nous allons présenter dans le chapitre suivant fonctionne d'ailleurs sur ce principe en reconnaissant de la même manière des documents manuscrits et imprimés.

Chapitre 9

Décrets de naturalisation

La seconde application présentée porte sur des documents d'archives de types décret de naturalisation. Nous avons réalisé une description perceptive guidée par un but pour ce type de document, qui nous permet de valider la possibilité de décrire un mécanisme de coopération simple dans DMOS-P. Nous comparons cette approche perceptive avec une version monorésolution développée précédemment dans l'équipe Imadoc, avec DMOS. Cette comparaison nous permet de valider les apports de la vision perceptive pour la reconnaissance de documents, c'est à dire une simplification de la description de la structure et une meilleure reconnaissance.

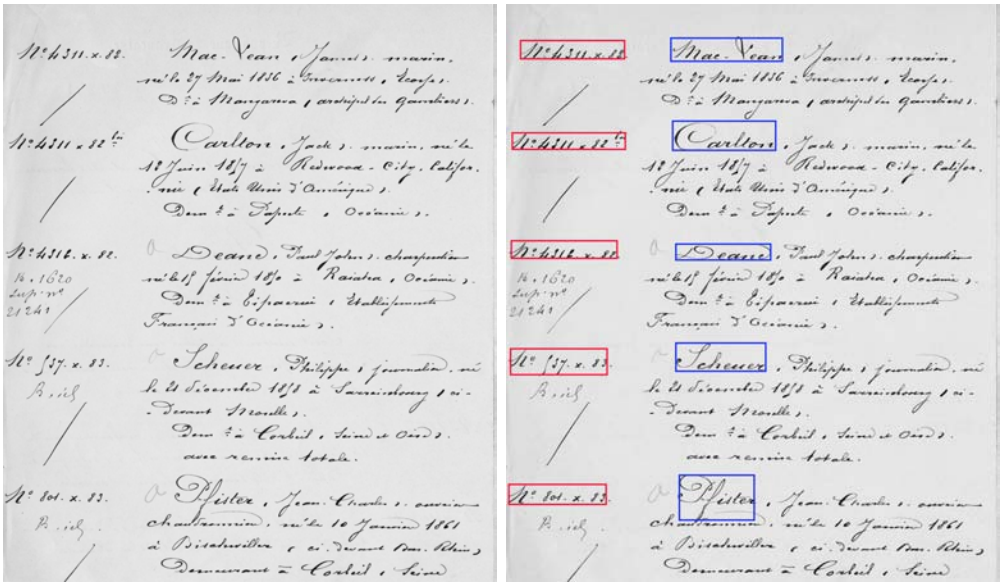
Dans ce chapitre, nous présentons donc les spécificités des décrets de naturalisation, puis le processus de reconnaissance exprimé dans DMOS-P. L'interprétation des résultats obtenus nous permet de conclure sur les apports de la vision perceptive pour ce type de documents.

9.1 Présentation des documents

Les décrets de naturalisation sont des documents d'archives justifiant l'acquisition de la nationalité française par des personnes d'origine étrangère. Pour ces personnes, ces documents sont parfois la seule preuve de leur nationalité.

Nous nous intéressons à des décrets datés de 1883 à 1930. Un décret est un document composé d'une dizaine de pages, manuscrites ou imprimées. Chaque page est composée d'une succession d'actes, disposés sur deux colonnes : la marge et le corps de texte (figure 9.1(a)). Pour chacun des actes, un numéro de dossier est placé dans la marge. Le corps de texte contient un paragraphe commençant par le nom et le prénom de la personne concernée. La figure 9.1(b) présente un exemple de page de décret, dans laquelle les numéros et les patronymes sont mis en évidence.

Ces décrets sont triés selon leur date de publication. Cependant, à l'intérieur d'un décret, les actes ne sont triés ni par ordre alphabétique du patronyme, ni par numéro de dossier. Il est donc particulièrement fastidieux de retrouver l'acte de naturalisation d'une personne précise, puisque le lecteur doit feuilleter une à une toutes les pages de tous les décrets s'il ne connait pas la date de publication.



(a) Exemple de page (b) Numéros de dossier et patronymes

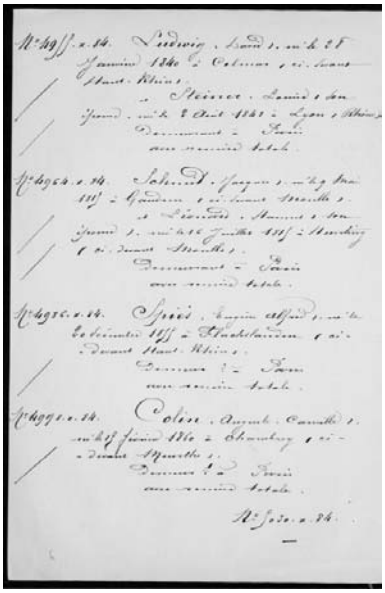
FIG. 9.1 – Exemple de page de décret de naturalisation

Le but de notre analyse est donc d'extraire les éléments clés du document, patronymes et numéros de dossier, afin de faciliter le feuilletage en proposant un résumé visuel des documents (figure 9.11).

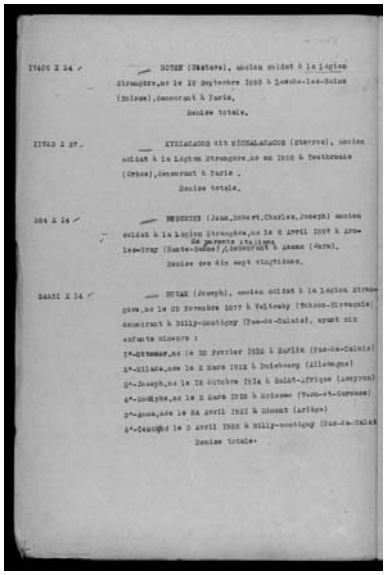
Même si les documents sont tous basés sur la même structure logique, les pages peuvent être très variées d'un décret à l'autre. Citons par exemple des documents manuscrits avec une écriture large (figure 9.2(a)), des documents imprimés avec une petite police de caractère (figure 9.2(b)), des premières et dernières pages ayant une structure légèrement différente (figures 9.2(c) et 9.2(d)). Notre description d'une page de décret doit être suffisamment générique pour pouvoir traiter tous ces cas. Il est donc nécessaire de se baser uniquement sur une description de la structure logique, et non sur des seuils ou des dimensions. Par conséquent, une méthode grammaticale est bien adaptée pour reconnaître ce genre de documents.

9.2 Processus de reconnaissance

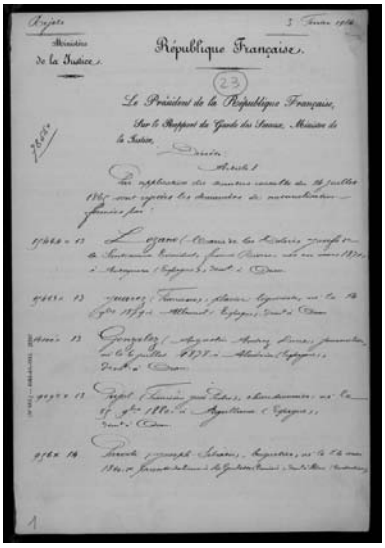
Des travaux précédents de l'équipe ont conduit à la description des décrets de naturalisation avec la méthode DMOS dans sa version initiale. Nous présentons brièvement cette mise en œuvre, afin de mettre en avant ses limites. Nous présentons ensuite notre description perceptive réalisée avec DMOS-P.



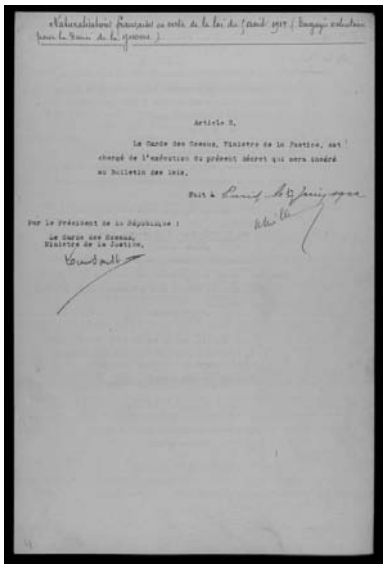
(a) Page manuscrite de 1884



(b) Page imprimée de 1928



(c) Première page d'un décret



(d) Dernière page d'un décret

FIG. 9.2 – Variabilité des pages de décrets de naturalisation

9.2.1 Approche monorésolution existante

Nous exposons le principe de fonctionnement de la grammaire décrivant les décrets de naturalisation dans la version initiale de la méthode DMOS [CCL04].

Le processus de recherche se base sur l'extraction de deux alignements verticaux, permettant de délimiter une zone de marge et une zone de corps de texte. On recherche ensuite les numéros et les noms associés, situés dans chacune de ces zones. L'analyse est basée uniquement sur les composantes connexes situées dans l'image de résolution initiale.

La figure 9.3 illustre la recherche des deux alignements verticaux. Le processus de reconnaissance d'un alignement vertical est le suivant :

1. Sélectionner une composante de base (en rouge sur la figure), vérifiant les propriétés « avoir quelque chose à droite » et « rien juste à gauche », c'est à dire une composante située à l'extrémité gauche d'une ligne.
2. Rechercher des composantes connexes alignées verticalement à la composante de base (en bleu sur la figure) situées également en extrémité de ligne.
3. L'alignement est validé si on peut trouver suffisamment de composantes connexes alignées à la première.

Le principal problème de cette approche est qu'elle est très sensible au bruit présent dans la marge, et que le résultat peut varier selon le choix de la composante connexe de base. D'autre part, la recherche d'un nombre suffisant de composantes alignées est réalisée par essais successifs, ce qui peut entraîner une combinatoire importante.

Une fois que la marge est détectée, les numéros et noms sont recherchés de chaque côté comme des composantes connexes :

- alignées horizontalement,
- suffisamment grandes (pour éviter les petites tâches),
- pas trop grandes (pour éviter les caractères qui se chevauchent entre lignes).

Cette méthode a été conçue pour des documents manuscrits de 1883 et 1884. Elle est assez spécifique à un style particulier d'écriture, en incluant notamment des seuils de distance et de taille. Cette méthode n'est donc pas adaptée pour la reconnaissance de manuscrits de style trop différent ou de documents imprimés, comme le montreront les résultats des expérimentations.

9.2.2 Approche perceptive

Au vu des limites de la première méthode, nous proposons une nouvelle description des décrets de naturalisation, en utilisant la vision perceptive.

Notre description se base sur deux résolutions :

- la résolution *normale* correspond à l'image initiale (240 dpi) ;
- la résolution *basse* correspond à une image sous-échantillonnée (15 dpi), de dimensions 16 fois plus petites.

L'analyse est réalisée à partir de deux ensembles d'éléments, les segments extraits à basse résolution (figure 9.4(a)) et les composantes connexes extraites à haute résolution (figure 9.4(b)). Les calques perceptifs utilisés sont présentés sur la figure 9.5.

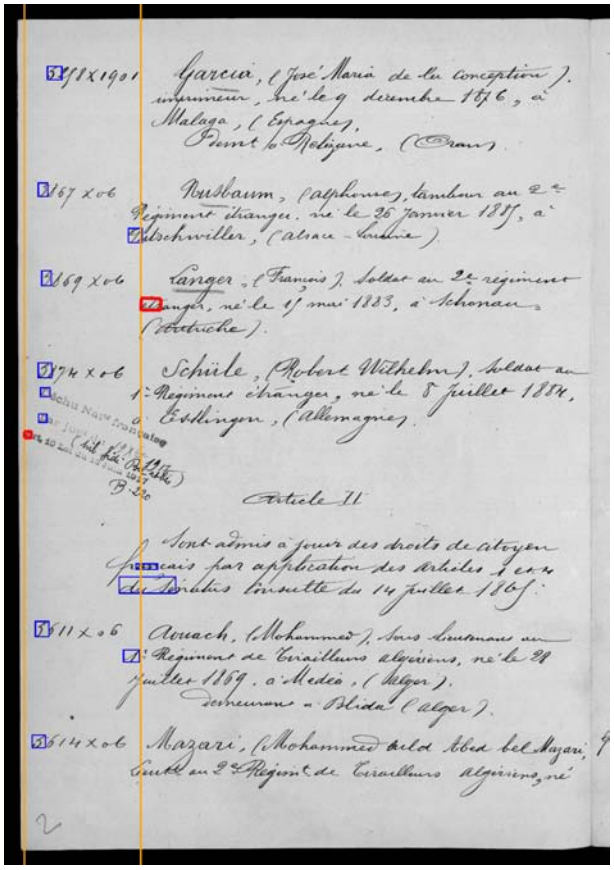
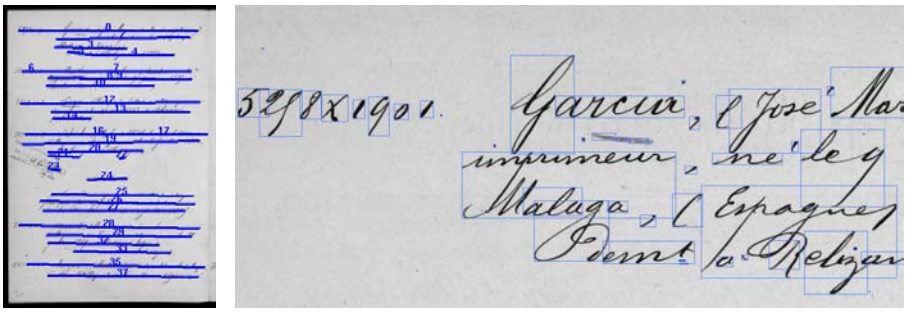


FIG. 9.3 – Recherche des alignements verticaux (en orange) à partir de composantes de base (en rouge) et de composantes alignées (en bleu), méthode monorésolution avec DMOS



(a) Segments extraits dans la résolution Basse

(b) Composantes connexes extraites dans la résolution Normale

FIG. 9.4 – Primitives utilisées pour la reconnaissance des décrets de naturalisation avec la méthode perceptive DMOS-P

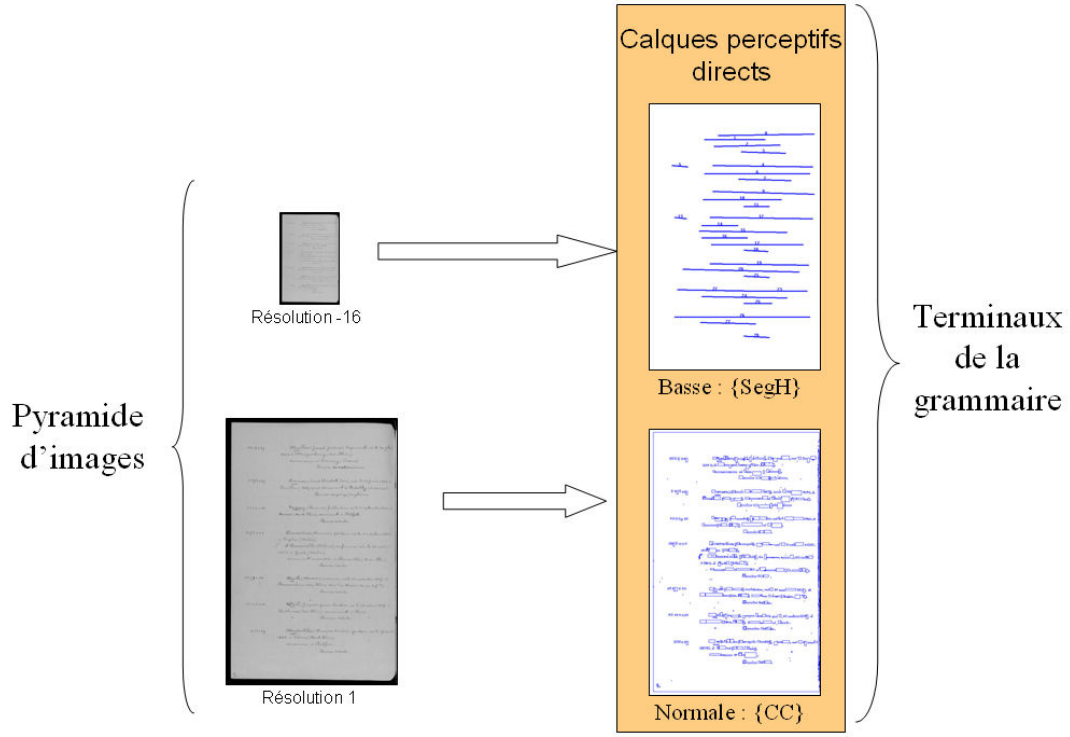


FIG. 9.5 – Calques perceptifs utilisés pour la description des décrets de naturalisation dans la méthode DMOS-P

Dans cette application, il n'était pas pertinent d'utiliser le calque perceptif contenant les lignes de texte construites de manière prégnante. En effet, dans ces documents, la position de la marge influe sur notre manière de segmenter les lignes de texte. Il est donc nécessaire d'introduire une connaissance sur la présence d'une marge. D'autre part, dans cette description, on se contente dans la plupart des cas d'observer les lignes de texte de loin. On ne détaille le contenu que de la première ligne de l'acte qui contient le nom. Il n'est donc pas nécessaire de rentrer dans le détail des lignes de texte fourni par la description comme des éléments prégnants.

La flexibilité de notre architecture nous permet d'utiliser les segments pour positionner la marge, puis de détailler uniquement les lignes de texte souhaitées en étant guidé par le but, sans s'encombrer du calque perceptif des lignes de texte qui apporte des informations mal adaptées au problème.

Nous présentons le principe général de notre description avant de l'exprimer dans le langage EPF.

9.2.2.1 Principe général

L'analyse se déroule selon le principe suivant :

1. détection de la marge, à basse résolution :
   (a) repérer les segments correspondants aux lignes de texte,
   (b) en déduire la position de la marge (figure 9.6(a)) ;
2. reconnaissance des actes ; pour chaque acte :
   (a) à la résolution normale, dans la marge, extraire un numéro de dossier comme une succession de composantes connexes alignées horizontalement (figure 9.6(b)),
   (b) à la résolution basse, trouver un segment correspondant à une ligne de texte, dans le corps de texte, en face du numéro trouvé précédemment,
   (c) à la résolution normale, à l'endroit de la ligne de texte repérée à l'étape précédente, et extraire le patronyme (figure 9.6(c)).

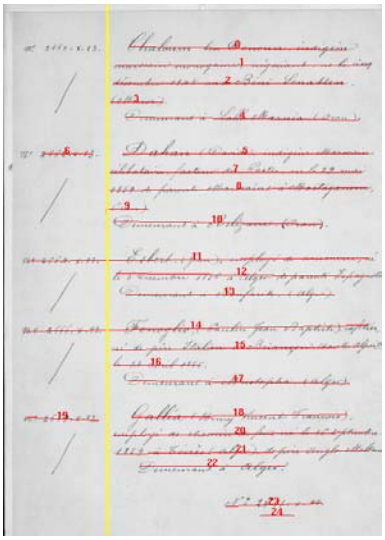
La particularité de cette analyse est qu'elle combine des allers-retours successifs entre les résolutions. Elle permet de créer une coopération, guidée par la connaissance, entre les données contenues dans les calques perceptifs directs.

9.2.2.2 Description dans le langage EPF

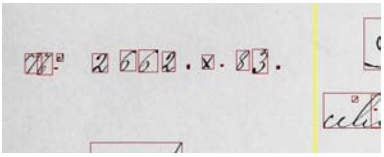
La description dans le langage EPF est basée sur des terminaux contenus dans deux calques perceptifs directs, contenant les données issues de deux résolutions : la résolution `Normale` qui correspond à l'image initiale, et la résolution `Basse` qui est la résolution `-16`.

Cette description met en avant la simplicité d'expression de la connaissance dans le langage EPF avec les outils de multirésolution. Seuls quelques attributs de la grammaire ont été omis pour faciliter la lecture.

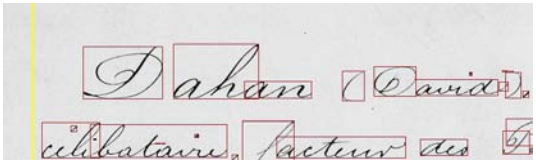
L'analyse d'une page consiste à trouver une marge puis à extraire des actes :



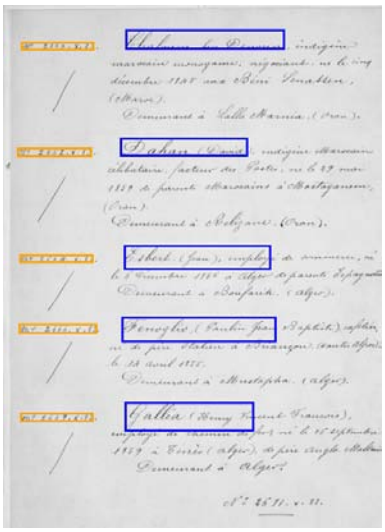
(a) Segments perçus à basse résolution, et la marge déduite (en jaune)



(b) Focalisation dans la marge et localisation d'un numéro



(c) Focalisation dans le corps de texte et localisation d'un nom



(d) Résultat final : numéros et noms

FIG. 9.6 – Mécanisme d'analyse des décrets de naturalisation avec la méthode perceptive DMOS-P

```
pageDeDecrets ::=
      USE_LAYER("Basse") FOR(marge M) &&
      AT_ABS(hautPage) &&
      USE_LAYER("Basse") FOR(ensembleActes M).
```

La recherche de la marge se base sur la reconnaissance des lignes de textes comme segments. L'analyse démarre à basse résolution :

```
marge M ::=
      AT_ABS(hautPage) &&
      ensembleLignesTexte &&
      calculerPositionMoyenneMarge M.
```

L'ensemble des lignes de texte est extrait récursivement par :

```
ensembleLignesTexte ::=
      TERM_SEG noCond noCond SegmentTrouvé &&
      AT(sousSeg SegmentTrouvé) &&
      ensembleLignesTexte.
```

La reconnaissance de l'ensemble des actes `ensembleActes` est effectuée récursivement et chaque acte est décrit de la manière suivante :

```
acte M ::=
      AT(zoneMarge M) &&
      USE_LAYER("Normale") FOR(detailsNumero Nb) &&
      AT(enFaceNumero Nb) &&
      TERM_SEG noCond noCond LigneDeNom &&
      AT(zoneLigneNom LigneDeNom) &&
      USE_LAYER("Normale") FOR(detailsNom).
```

Il faut noter que l'analyse est réalisée à résolution `Basse`, sauf pour les prédicats appelés à l'intérieur de l'opérateur `USE_LAYER`.

Selon le principe de la méthode DMOS-P, cette grammaire décrite en EPF permet de produire, par une étape de compilation, un analyseur dédié aux décrets de naturalisation.

9.3 Application

9.3.1 Base de documents

Nous disposons d'une large base de 15 699 registres, datés de 1883 à 1930, ce qui représente 85 088 pages de documents. Les images initiales ont une résolution de 240 dpi (taille d'environ 2000*3000 pixels) et sont stockées en JPEG.

Les images initiales sont très variées, tel que présenté sur la figure 9.2, et présentent les difficultés habituellement liées aux documents d'archives : page abîmées, tachées, encre traversant de l'autre côté de la page.

9.3.2 Evaluation des résultats

Nous présentons maintenant l'application de nos travaux sur les décrets de naturalisation et l'évaluation des résultats.

9.3.2.1 Bases d'évaluation

Afin d'estimer les taux de reconnaissance de notre méthode, nous avons construits manuellement trois bases de vérités terrains.

La base manuscrite est composée de 1999 images manuscrites, datées de 1883 et 1884.

La base variée est composée de 320 pages manuscrites ou imprimées, sélectionnées aléatoirement parmi les 47 années, avec approximativement le même nombre de pages pour chaque année.

La base représentative est construite en prenant, selon l'ordre chronologique, une image sur 250. Cette base de 347 images est donc représentative du ratio imprimé *vs* manuscrit et des différents problèmes rencontrés dans la base.

Les résultats les plus pertinents sont donc ceux de la base représentative.

9.3.2.2 Métriques d'évaluation

Les fichiers de vérité terrain sont composés des coordonnées des rectangles englobants des noms et des numéros. Les résultats de la méthode sont également rendus sous cette forme. L'évaluation du taux de reconnaissance revient donc à évaluer la mise en correspondance des rectangles attendus avec ceux qui sont reconnus (figure 9.7).

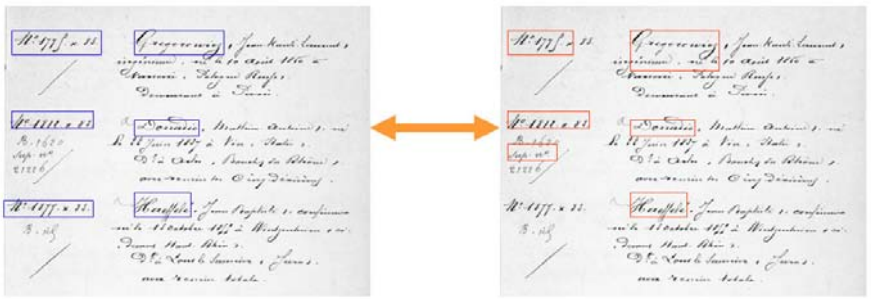


FIG. 9.7 – Evaluation : confronter la liste de rectangles attendus, à gauche, avec la liste de rectangles reconnus, à droite

Nous avons écarté la métrique proposée par le projet RIMES (voir chapitre 8), à cause de la difficulté pour mettre en place une vérité terrain pour chaque pixel, dans le cas des mots avec beaucoup de chevauchement.

On trouve dans la littérature plusieurs méthodes pour calculer la concordance entre des boîtes englobantes attendues et reconnues. Certaines se basent sur la surface en

commun [Gar95], sur les pixels noirs correctement étiquetés [GGC+06]. D'autres méthodes prennent en compte le taux de bruit reconnu [HJBJ+96] [YV98]. Dans tous les cas, il est nécessaire de définir un seuil minimal de recouvrement pour pouvoir dire que deux boîtes se recouvrent. Ce seuil est toujours propre à l'application.

Dans le cas de notre application aux décrets de naturalisation, nous rappelons que le but est de définir des zones pour un feuilletage rapide. Il est donc nécessaire que le champ recherché soit lisible entièrement. Par contre, le rectangle reconnu peut être un peu plus grand.

La figure 9.8 présente un exemple de comparaison entre un rectangle reconnu et le rectangle attendu. Nous calculons l'intersection entre ces deux rectangles, ce qui correspond à la partie utile que pourra lire l'utilisateur final de l'application. Pour être lisible correctement, cette intersection doit avoir une largeur très proche de la largeur du rectangle attendu. La hauteur de l'intersection doit également être suffisamment grande.

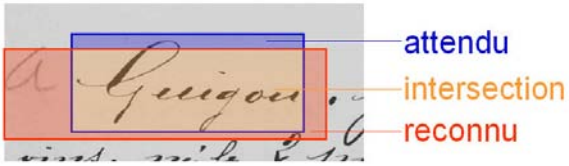
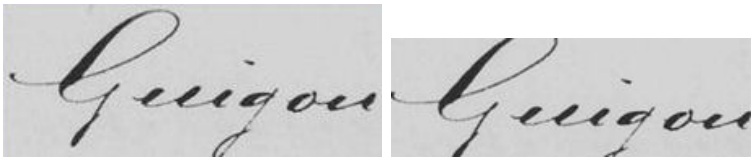


FIG. 9.8 – Calcul de l'intersection entre les rectangles attendus et reconnus

Comme dans les approches de la littérature, nous mettons en place un seuil propre à l'application. Ici, nous avons évalué de manière empirique que, pour être lisible, le rectangle contenant un mot doit être reconnu à au moins 95% de sa largeur et 75% de sa hauteur (figure 9.9).



(a) Rectangle idéal attendu

(b) Rectangle minimal : 95% de la largeur, 75% de la hauteur

FIG. 9.9 – Rectangle minimal, déterminé expérimentalement, pour que le mot soit lisible

En résumé, pour déterminer si un rectangle attendu est correctement reconnu, nous utilisons le principe suivant :

1. Calculer l'intersection entre les rectangles attendus et reconnus.

2. Le recouvrement est correct si :
   - la largeur de l'intersection est supérieure à 95% de l'attendu
     et
   - la hauteur de l'intersection est supérieure à 75% de l'attendu.

C'est cette métrique que nous utilisons pour calculer les taux de reconnaissance des deux méthodes décrivant les décrets de naturalisation.

9.3.2.3 Résultats

Les résultats obtenus sont regroupés dans le tableau 9.1.

Base	Nombre de pages	Nombre d'actes	Version monorésolution	Vision perceptive
Manuscrite	1 999	13 896	99.06%	**98.63%**
Variée	320	2 706	86.66%	**98.93%**
Représentative	347	3 186	92.69%	**98.31%**

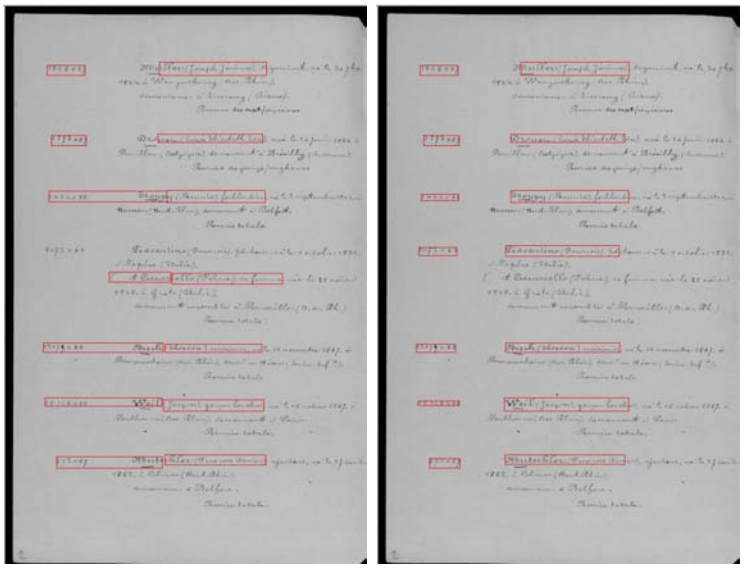
TAB. 9.1 – Comparaison des taux de reconnaissance entre la version monorésolution et la version avec vision perceptive

Comme nous l'avons expliqué dans la partie 9.2.1, la méthode initiale monorésolution a été plus spécifiquement adaptée à des documents manuscrits. Elle obtient donc de bons résultats pour la base *manuscrite* : 99.06% de reconnaissance. Cependant, cette méthode obtient seulement 92.69% de reconnaissance pour la base *représentative*. En effet, cette méthode n'est pas assez générique pour traiter indifféremment les documents manuscrits et imprimés.

Les résultats sur la base *représentative* montrent que la méthode basée sur la vision perceptive a permis d'améliorer le taux de reconnaissance, principalement grâce à l'aspect générique de notre description qui n'est pas spécifique aux documents manuscrits. Cette méthode obtient ainsi 98.31% de reconnaissance sur la base *représentative*. Précisons que pour cette base, nous obtenons un taux de fausse reconnaissance (faux-positifs) de 4.33%, ce qui n'est pas problématique dans le cas de notre application dont le but est un feuilletage rapide (cela rajoute juste quelques mots en plus).

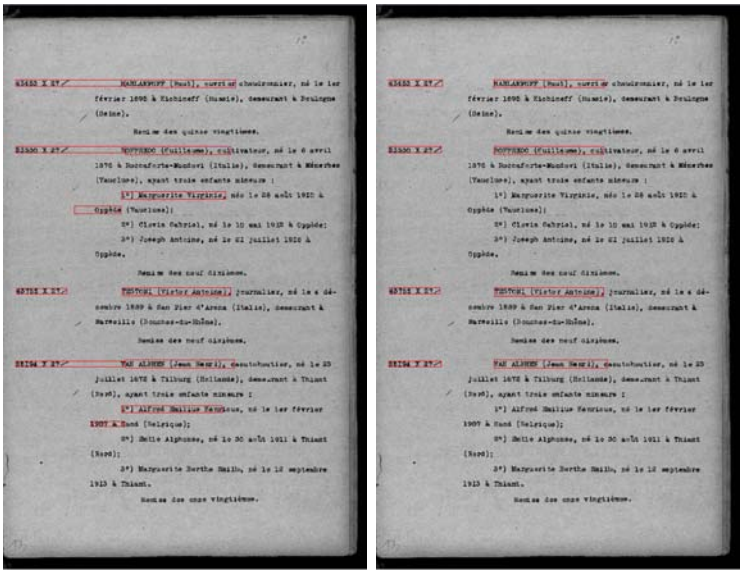
La figure 9.10 met en avant les différences entre les résultats produits par les deux versions. Dans la version monorésolution, la recherche de la marge produit plus souvent un résultat erroné. Dans les exemples présentés, la marge est trouvée trop à droite, ce qui entraîne de mauvaises hypothèses pour la présence des numéros et décale la position des patronymes trop à droite.

Enfin, même si ce n'est pas l'objectif premier, nous pouvons noter que la vision perceptive améliore le temps de calcul. En effet, le temps de traitement en monorésolution est en moyenne de 6.4 secondes par acte, alors qu'il n'est plus que de 1.2 secondes par acte avec la vision perceptive. En effet, l'utilisation de la vision perceptive permet de diminuer le nombre d'hypothèses étudiées par l'analyseur.



(a) Page manuscrite, version monorésolution

(b) Page manuscrite, avec vision perceptive



(c) Page imprimée, version monorésolution

(d) Page imprimée, avec vision perceptive

FIG. 9.10 – Dans ces images, avec la monorésolution, la marge est localisée trop à droite, ce qui provoque des erreurs de localisation des noms et numéros. Cette localisation est correcte avec la méthode perceptive.

9.3.3 Application réelle

Dans le cadre d'un contrat de recherche avec la Direction des Archives de France pour le Centre Historique des Archives Nationales (CHAN), nous avons appliqué notre système à la base complète des 85 088 pages de 1883 à 1930. Dans cette base, nous avons reconnu 433 230 actes (couples numéro, nom), soit environ 5 par page en moyenne. 106 pages ont été reconnues, à tort, comme vides, ce qui représente un taux d'omission de 0.1%.

Les résultats produits sont maintenant accessibles en salle de lecture pour les usagers du CHAN, grâce à une plateforme de consultation. Cette plateforme (figure 9.11), permet un feuilletage rapide des décrets de naturalisation par le patronyme.

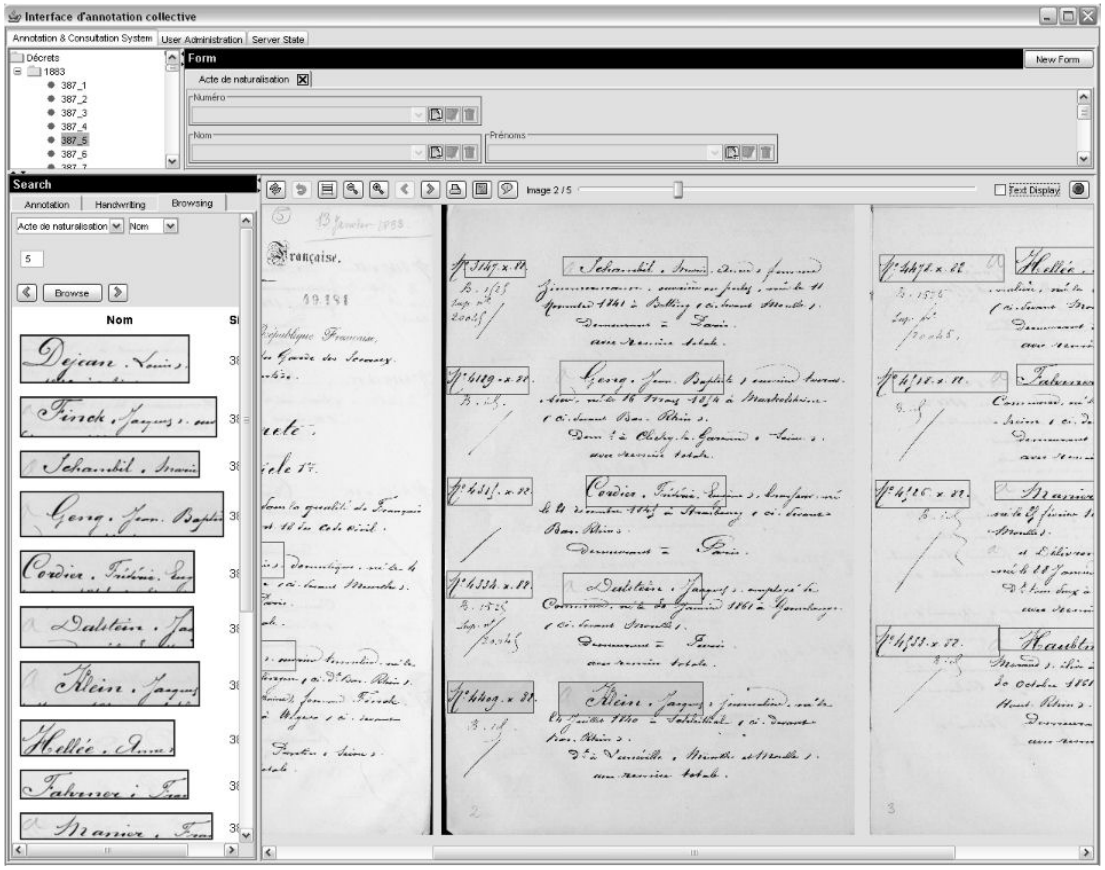


FIG. 9.11 – Plateforme de consultation : feuilletage rapide par patronymes à gauche, visualisation de la page entière à droite.

9.4 Apports de notre approche

Grâce à la méthode DMOS-P, nous avons créé facilement un mécanisme perceptif dédié à la reconnaissance des décrets de naturalisation. Cette application démontre la souplesse de la méthode DMOS-P, qui permet d'utiliser uniquement les concepts perceptifs nécessaires. En particulier, dans cette application, le calque perceptif induit contenant les lignes de texte n'est pas utilisé. En effet, il aurait apporté une information superflue (détail de toutes les composantes connexes) et légèrement erronée (gestion non prévue du découpage en deux colonnes). Grâce à la flexibilité des calques, il a donc été possible de décrire un mécanisme perceptif dédié aux décrets de naturalisation, qui soit à la fois précis et efficace.

La comparaison avec une grammaire existante liée à la méthode DMOS permet de montrer la simplification de l'expression de la connaissance grâce à la coopération multirésolution. Ainsi, on peut noter une simplification au niveau :

- de la recherche de la position de la marge,
- de la recherche de la position des lignes de texte,
- du traitement d'écriture variée, manuscrite ou imprimée.

Le mécanisme perceptif ainsi produit est suffisamment générique pour obtenir des résultats homogènes sur toute la base, que les documents soient imprimés ou manuscrits.

Le mécanisme perceptif décrivant les décrets de naturalisation illustre l'intuition présentée dans la partie 3.2.1.1 : l'utilisation de la vision perceptive permet de diminuer l'impact du bruit présent dans ces documents, et de *sélectionner* uniquement les informations structurelles utiles. Ceci est illustré par exemple par la recherche de la marge. Dans la version existante, la recherche de composantes connexes alignées était très sensible à la présence de composantes connexes parasites ; avec notre mécanisme perceptif, la marge est détectée globalement et n'est pas perturbée par de petits bruits.

Enfin, les décrets de naturalisation sont des documents faiblement structurés, dans le sens où leur structure n'est matérialisée par aucun élément physique direct, tel que des traits. Les résultats obtenus illustrent donc également l'apport proposé dans la partie 3.2.2.1 : la combinaison des visions globales et locales permet de *reconstituer* la structure propre aux décrets de naturalisation à partir de la fusion d'indices obtenus aux différentes résolutions.

Chapitre 10

Documents indiens

Afin de montrer la généricité de notre approche, nous proposons d'utiliser notre méthode perceptive DMOS-P pour un problème de structuration un peu en marge de ceux habituellement rencontrés : la recherche de ligne de base dans des documents indiens, et plus généralement dans l'écriture manuscrite. Ces travaux ont également pour but de valider la détection des lignes de texte prégnantes ainsi que les outils de transfert d'informations entre résolutions, gérant l'adaptation de la localisation et le changement de primitive visuelle (présentés dans la partie 6.3).

Ces travaux se placent dans le cadre d'une coopération internationale avec le professeur Bidyut Baran Chaudhuri de l'*Indian Statisical Institute* de Calcutta. Dans ce contexte, nous abordons le problème de la reconnaissance de l'écriture manuscrite d'une langue spécifique : le Bangla. Nous proposons de faciliter la reconnaissance de l'écriture en recherchant les lignes de base par une approche structurelle.

Nous présentons dans un premier temps les spécificités du Bangla. Nous détaillons ensuite notre méthode de reconnaissance des lignes de base dans l'écriture manuscrite, avant de présenter l'application pratique de notre méthode, tant sur les documents indiens que sur de l'écriture manuscrite en français. Nous concluons sur les apports de la vision perceptive pour ce type de problème.

10.1 Contexte de l'écrite manuscrite Bangla

10.1.1 Ecriture Bangla

Le Bangla ou Bengali est une langue utilisée par plus de 230 millions de personnes de l'est de l'Inde et du Bangladesh. Il utilise un alphabet particulier, dérivé de l'alphabet Brahmi, qui est particulièrement complexe vis à vis de l'écriture latine. Un exemple de texte imprimé est présenté sur la figure 10.1.

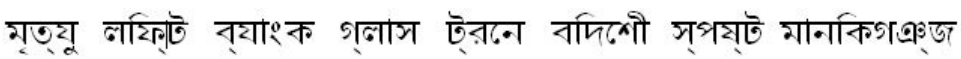
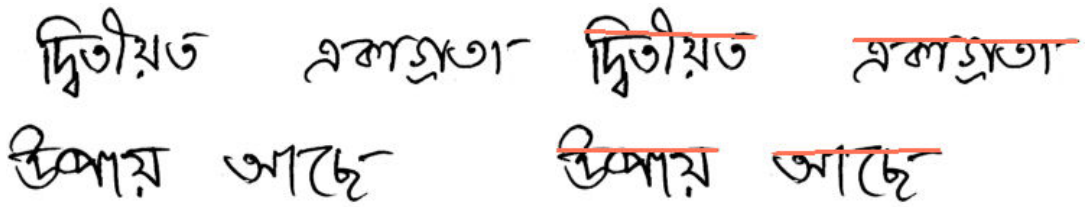


FIG. 10.1 – Exemple de texte bangla imprimé

L'alphabet présente environ 50 voyelles et consonnes basiques, auxquelles on ajoute 10 modifieurs, qui peuvent être connectés aux caractères basiques à différentes positions, dans le but de créer de nouveaux caractères. De plus, il existe environ 250 caractères composés, obtenus par la fusion des formes de deux ou trois consonnes, et environ trois fois plus de nouvelles formes qui sont créées par l'attachement des modifieurs de voyelles aux caractères composés.

Un des aspects intéressants des caractères bangla est l'existence d'une ligne horizontale dans la partie supérieure de la plupart des caractères. Cette ligne de base est appelée *headline*, littéralement « ligne de tête »[1]. Un exemple de *headline* est donné sur la figure 10.2.



(a) Exemple de manuscrit Bangla (b) *Headlines* associées aux mots

FIG. 10.2 – Exemple de *headline*

La *headline* est très marquée dans le texte imprimé (figure 10.1), mais moins présente dans le texte manuscrit. En effet, la plupart des scripteurs ne la marquent pas dans 60 à 70% des caractères. Cependant, une personne connaissant l'écriture Bangla est capable de trouver la position approximative de la *headline*.

10.1.2 Reconnaissance d'écriture manuscrite

La reconnaissance automatique de manuscrit Bangla est un problème de recherche naissant. Pal et Chaudhuri présentent dans [PC04] un état de l'art des différentes techniques de reconnaissance automatique de caractères indiens imprimés. Des travaux sur la reconnaissance de caractères isolés en Bangla ont également été présentés [BPS07], ainsi que des essais de lecture de noms de familles sur des courriers Bangla [RVB+05].

Une des manières usuelles de reconnaître un texte manuscrit est d'identifier les lignes de textes individuelles, puis les mots de chaque ligne. Ceci est encore un sujet de recherche ouvert comme le montrent les travaux récents de Roy *et al.* [RPL08].

Les mots peuvent être reconnus comme des entités globales. Dans ce cas, la détection de la *headline* est utile puisque les classifieurs de mots, tels que les HMM, peuvent faire une bonne approximation du point terminal d'un caractère et du début du caractère suivant. Une autre approche consiste à segmenter les mots en caractères individuels qui

[1] Dans la suite du manuscrit, nous emploierons le terme *headline*, n'ayant pas trouvé de traduction satisfaisante en français.

sont envoyés à un extracteur de caractéristiques puis à un classifieur. Dans ce cas, La *headline* agit comme un guide pour segmenter les caractères.

Nous proposons donc, grâce à notre approche par vision perceptive, d'identifier la position de la *headline* dans des mots, dans le but de faciliter l'étape suivante de reconnaissance de l'écriture. Les travaux basés sur l'extraction de la *headline* sont peu nombreux dans la littérature à notre connaissance, ce qui ne nous permet pas de comparer directement notre approche avec l'état de l'art.

10.2 Processus de reconnaissance

Nous proposons de suivre un processus d'analyse en trois étapes :

1. la localisation des lignes de texte,
2. le découpage des lignes en mots,
3. le positionnement des headlines.

10.2.1 Localisation des lignes de texte

Les lignes de texte sont perçues comme des éléments prégnants. Leur localisation est donc réalisée grâce à l'approche perceptive présentée dans la partie 7.1.1, et ce sans nécessiter aucune modification ni ajustement de paramètre par rapport à la version présentée dans le chapitre 7. Cette localisation est basée sur la vision conjointe des segments à basse résolution et des composantes connexes à haute résolution.

La première étape du processus d'analyse est donc réalisée de manière générique, grâce à notre approche perceptive. Le résultat de cette étape est contenu dans le calque perceptif des lignes de texte. Notre approche est donc basée sur trois calques perceptifs, présentés sur la figure 10.3.

10.2.2 Découpage des lignes en mots

La seconde étape de l'analyse consiste à un regroupement en mots des composantes connexes présentes dans la ligne.

Pour effectuer ce regroupement, nous calculons toutes les distances entre deux composantes connexes successives de la ligne. Nous appliquons ensuite une classification selon la méthode des k-moyennes (avec k=2) pour séparer les distances intra-mots des distances inter-mots. Cette classification permet d'obtenir un seuil de distance pour le regroupement des composantes connexes.

Concernant le calcul des distances, nous pouvons noter ici que les composantes connexes sont représentées par leur rectangle englobant, ce qui entraîne que la distance entre deux composantes connexes est une distance entre deux rectangles englobants. Cette distance n'est pas toujours représentative dans le cas de composantes connexes dont les rectangles se recouvrent. Pour obtenir des distances plus proches de la réalité des composantes connexes, nous utilisons un pavage de Voronoï et les distances entre

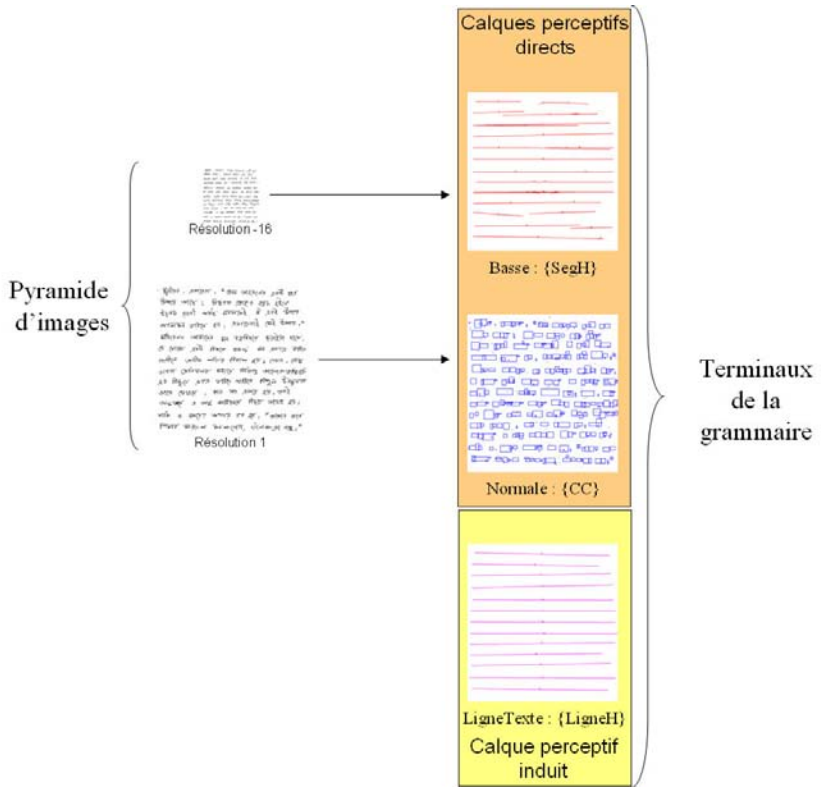


FIG. 10.3 – Calques perceptifs utilisés pour la localisation des *headlines*

composantes connexes sont calculées dans le graphe de Delaunay associé[2]. Plus de détails sur cette méthode sont présentés dans deux articles que nous avons écrits [7] [8].

10.2.3 Localisation des *headlines*

Pour chacun des mots trouvés, la dernière étape consiste à trouver le positionnement exact de la headline. Pour cela, on utilise la ligne de texte globale qui donne une allure générale de la pente et de la position de la ligne. A partir de cette ligne, on génère le sous-segment limité à la longueur du mot étudié.

Dans un deuxième temps, on utilise l'outil de recalage des lignes abstraites, présenté dans la partie 6.3.1.3, pour positionner la *headline* sur les pixels du haut du mot. Cet outil de recalage permet d'absorber localement des courbures et des variations fortes des caractères. En effet, la vision globale fournit un contexte et une zone de recherche pour le recalage, ce qui permet de gagner en précision pour les pixels à prendre en compte localement dans le recalage. Par exemple, le recalage est capable d'ignorer les hampes et les jambages qui dépassent des mots.

10.2.4 Bilan

Les différentes étapes de reconnaissance sont regroupées sur la figure 10.4.

Cette grammaire est décrite en EPF, mais elle est en réalité très simple puisqu'elle est basée principalement sur trois outils que nous avons évoqués précédemment :

- le calque perceptif induit des lignes de texte, construites selon une approche perceptive, considérées comme des éléments prégnants
- le graphe de Voronoï/Delaunay pour le calcul des seuils de distance
- la fonctionnalité de recalage des lignes abstraites.

10.3 Application

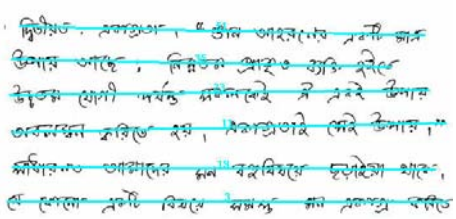
Nous présentons ici les résultats obtenus avec notre méthode pour la recherche de la *headline* dans des documents bangla. Puis, nous proposons d'étendre l'application au cas de la recherche de lignes de base dans des documents manuscrits français, ce qui est possible par la simple modification d'un paramètre.

10.3.1 Positionnement des *headlines*

10.3.1.1 Base de test

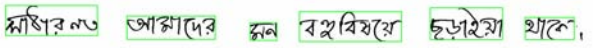
Nous appliquons notre système sur 61 pages de documents manuscrits Bangla, écrites par 27 scripteurs différents. Les images initiales ont une résolution de 300 dpi, et une taille d'environ 2000*3200 pixels pour la moitié des documents, et de 2000*1900 pixels pour l'autre moitié des documents.

[2]En ce qui concerne ces documents indiens, un test sur la reconnaissance de 3293 mots montre que l'approche basée sur les rectangles englobants fait deux fois plus d'erreur de segmentation que la méthode basée sur le pavage de Voronoï/Delaunay.

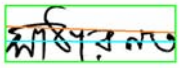


(a) Utilisation des lignes de texte construites comme éléments prégnants dans le calque perceptif induit

(b) Composantes connexes d'un même ligne utilisées pour le calcul de la distance entre mots



(c) Découpage d'une ligne en mots



(d) Pour chaque mot, utilisation de la ligne initiale (en bleu) et recalage (en orange) sur le haut des pixels noirs

(e) Résultat final : l'ensemble des *headlines*

FIG. 10.4 – Etapes de reconnaissance des *headlines*

Nous avons manuellement mis en place une base d'évaluation composée de 3293 *headlines* de mots, réparties sur 39 pages.

10.3.1.2 Résultats

Les résultats obtenus sont présentés dans le tableau 10.1.

Pages	39
Mots attendus	3293
Mots correctement segmentés	2948 89.52%
Headlines correctes lorsque le mot est bien segmenté	2886 97.89%

TAB. 10.1 – Reconnaissance des mots et des *headlines* avec notre approche perceptive, sachant que l'objectif est le positionnement des *headlines* et non la segmentation en mots

Il faut noter que notre priorité n'était pas ici la segmentation en mot mais le positionnement de la *headline*, qui est correct dans 97.89% des cas.

Il est difficile d'évaluer ces résultats puisque nous n'avons pas trouvé dans la littérature d'autres travaux dédiés à la recherche de la headline. Néanmoins, grâce à l'utilisation de l'approche perceptive, notre méthode est assez générique. Ainsi, notre méthode peut facilement s'adapter pour la recherche de lignes de base dans le cas de manuscrits français.

10.3.2 Extension au cas de texte français

10.3.2.1 Méthode générale

La différence entre le français et le bangla est la position de la ligne de base : sur le bas des mots et non sur le haut.

Positionner les lignes de base dans des documents français revient donc à appliquer le même principe que pour les documents indiens : recherche des lignes, recherche des mots, puis recalage de la ligne de base, en changeant uniquement un seul paramètre du recalage pour recaler en tenant compte des pixels extrêmes du bas des mots.

10.3.2.2 Base de test

Nous choisissons d'appliquer notre méthode de recherche des lignes de base sur des documents manuscrits, non contraints, avec des scripteurs variables, en langue française. Les courriers fournis par le projet RIMES (présentés dans le chapitre 8) réunissent tous ces critères.

Nous avons donc étiqueté manuellement les lignes de base de 25 courriers, soit 2691 mots.

10.3.2.3 Résultats

Un exemple de résultat obtenu est présenté sur la figure 10.5.

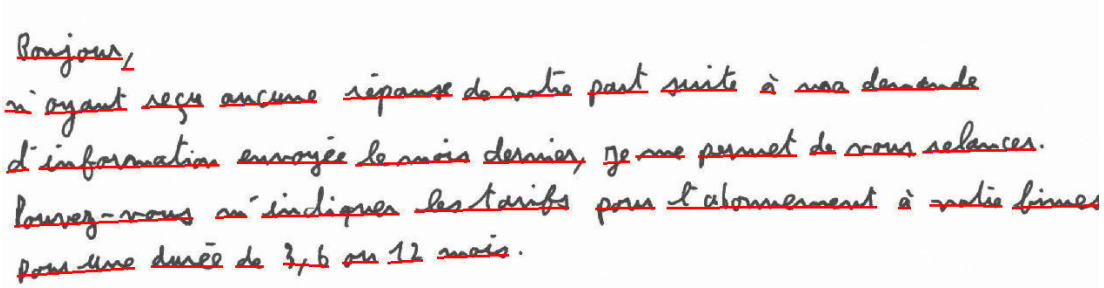


FIG. 10.5 – Exemples de lignes de bases localisées dans un document manuscrit français

Les résultats obtenus sur la base de test sont présentés dans le tableau 10.2.

Pages	25
Mots attendus	2691
Lignes de base correctes lorsque le mot est bien segmenté	98.22%

TAB. 10.2 – Reconnaissance des lignes de base avec notre approche perceptive dans des documents français

Le taux de reconnaissance des lignes de base dans les documents manuscrits français est proche du taux de reconnaissance des *headlines* dans les documents manuscrits banglas (98.2% contre 97.9%).

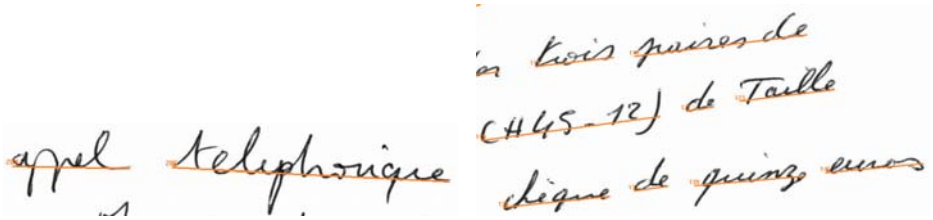
De nombreuses méthodes ont été proposées dans la littérature pour la recherche des lignes de base [MSB+08]. Par rapport à ces approches, notre méthode a la particularité d'utiliser un contexte global pour faciliter le positionnement local de la ligne de base, ce qui permet entre autres de diminuer les perturbations provoquées habituellement par les jambages (figure 10.6(a)), et de gérer les irrégularités (biais, courbure) de l'écriture (figure 10.6(b).

10.4 Intérêts de notre approche

Grâce à la méthode DMOS-P nous avons pu créer un mécanisme de coopération perceptive dédié au problème de la recherche de lignes de base dans l'écriture manuscrite.

Ces travaux permettent, d'une part de valider le caractère prégnant des lignes de texte, et la généricité de la description associée au calque perceptif. En effet, notre système de détection des lignes de texte est applicable en dehors de l'alphabet latin.

D'autre part, ces travaux montrent l'intérêt des outils de transfert d'informations entre résolutions, présentés dans la partie 6.3. En effet, les lignes abstraites servent de support à la synthèse des lignes de base. Elles permettent de gérer une représentation par



(a) Pas de perturbation par les jambages

(b) Adaptation locale au biais

FIG. 10.6 – Intérêts de l'utilisation d'un contexte global pour le positionnement local des lignes de base

des primitives différentes selon la résolution : segments ou composantes connexes. Elles facilitent également la recherche du positionnement précis des lignes de base en fonction des pixels présents dans l'image. Ainsi, elles permettent la définition d'un contexte pour sélectionner les pixels pertinents qui seront utiles pour ajuster la localisation des lignes de base.

Cette application correspond à l'intuition présentée dans la partie 3.2.2.1 : la vision perceptive permet de *reconstituer* et de positionner des éléments structurels ayant une existence physique faible, c'est à dire les lignes de base. Ceci est permis par le mécanisme de recalage des lignes abstraites, permettant une coopération entre les résolutions guidée par la description symbolique du document.

Chapitre 11

Pages de presse ancienne

Nous présentons dans ce chapitre une étude réalisée sur un quatrième type de documents : des pages de presse ancienne. Dans ce contexte, la méthode DMOS-P a été utilisée sous tous ses aspects, pour produire une chaîne de traitement complète, réalisée en lien avec un transfert industriel.

Avec cette application, nous montrons donc comment les formalismes et les outils de la méthode DMOS-P ont permis de créer simplement un mécanisme perceptif puissant, dédié à la reconnaissance des pages de presse.

Dans un premier temps, nos travaux sur les pages de presse consistent en une application dont le but est de reconnaître les filets présents dans ces pages. En effet, les filets représentent des composants élémentaires de la structure.

Ensuite, nous proposons de découper les pages de journaux en *cases*, selon les filets visibles. Le but de cette application n'est pas de produire une structuration parfaite des cases de journaux, mais de montrer les apports de notre approche perceptive vis à vis d'une méthode monorésolution déjà existante dans DMOS pour traiter ce genre de problèmes.

Enfin, nous présenterons une application plus globale, développée par la société Evodia[1], illustrant le transfert industriel de la méthode DMOS-P.

11.1 Présentation des documents

Les documents que nous étudions sont des pages de presse datées de 1859 à 1944, issues de 4 périodiques différents : le « Journal de Mantes », « La Concorde de Seine et Oise », « Le Courrier de Versailles et de Seine et Oise » et « Le Progrès ». Ces images nous sont fournies par les Archives Départementales des Yvelines.

La figure 11.1 présente plusieurs exemples de ces pages de périodiques. Ces documents anciens regroupent différentes formes de bruits (selon la définition présentée dans l'introduction) : pages déchirées, pliures, encre pâle.

[1] www.evodia.fr

(a) Première page de 1875

(b) Première page de 1888

(c) Dernière page de 1905

(d) Dernière page de 1940

FIG. 11.1 – Exemples de pages de presse ancienne étudiées

Nous pouvons noter que ces documents sont très fortement structurés. En effet, les colonnes sont systématiquement délimitées par des filets verticaux ; des séparateurs horizontaux permettent de structurer les articles.

Dans la suite de l'analyse, nous distinguerons les résultats obtenus sur les premières pages (figures 11.1(a) et 11.1(b)) et les dernières pages du périodique (figures 11.1(c) et 11.1(d)). En effet, les premières pages présentent un contenu assez homogène : une zone de titre du périodique, suivie d'articles avec des polices de caractères utilisées de manière régulière. En revanche, les dernières pages des périodiques contiennent fréquemment des encarts publicitaires et des petites annonces de nature très hétérogènes : polices de caractères variables, dessins, gravures, contours décoratifs. Ces dernières pages, bien que très structurées, présentent donc plus de difficultés pour la reconnaissance.

La structure de ces documents est marquée par des filets dont la reconnaissance est délicate. En effet, ces filets sont de nature variée et regroupent les différentes difficultés présentées dans la partie 3.1.2 : filets épais mouchetés, filets doubles, filets fins et discontinus.

Dans ces documents, nous proposons donc dans un premier temps une méthode de reconnaissance des filets, avant d'utiliser les filets trouvés pour découper ces pages en cases. Pour ces deux applications, nous mettrons en avant les avantages de notre méthode perceptive en la comparant avec la méthode monorésolution existant dans DMOS.

11.2 Reconnaissance des filets

11.2.1 Approche monorésolution

Nous créons une grammaire spécifique, appelée TRAIT_MONO dont le but est de construire des filets en se basant uniquement sur les segments extraits dans l'image de résolution initiale. La figure 11.2 présente le calque utilisé.

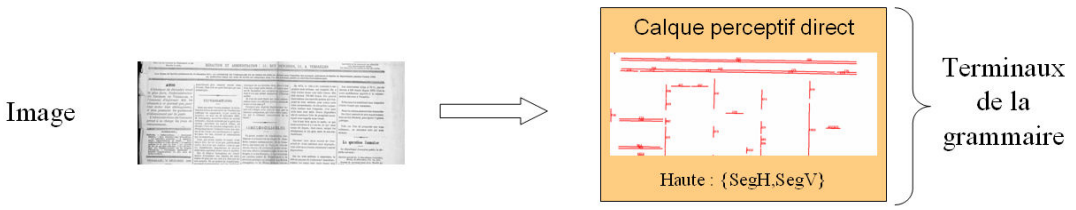


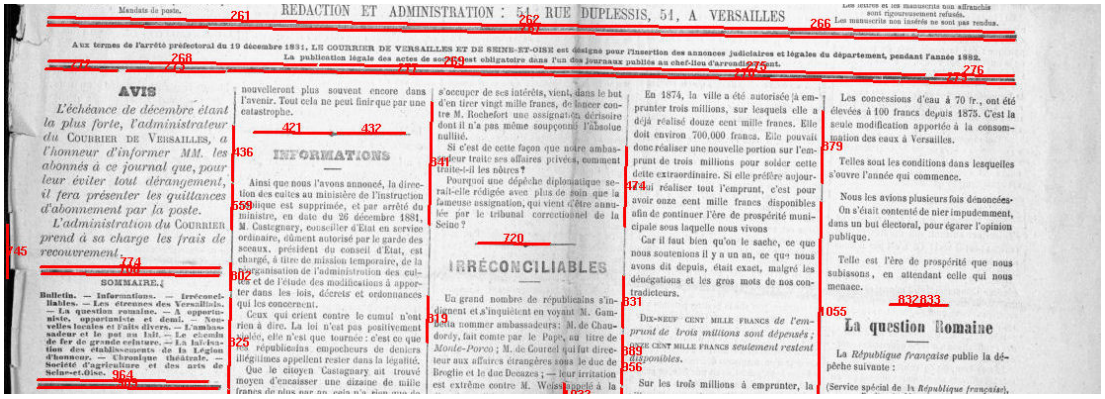
FIG. 11.2 – Calque utilisé pour l'analyse monorésolution des traits

Dans l'image de résolution initiale, les segments extraits ne correspondent souvent qu'à une partie de filet. En effet, la vision locale des segments limite leur reconnaissance dans le cadre de ces documents bruités.

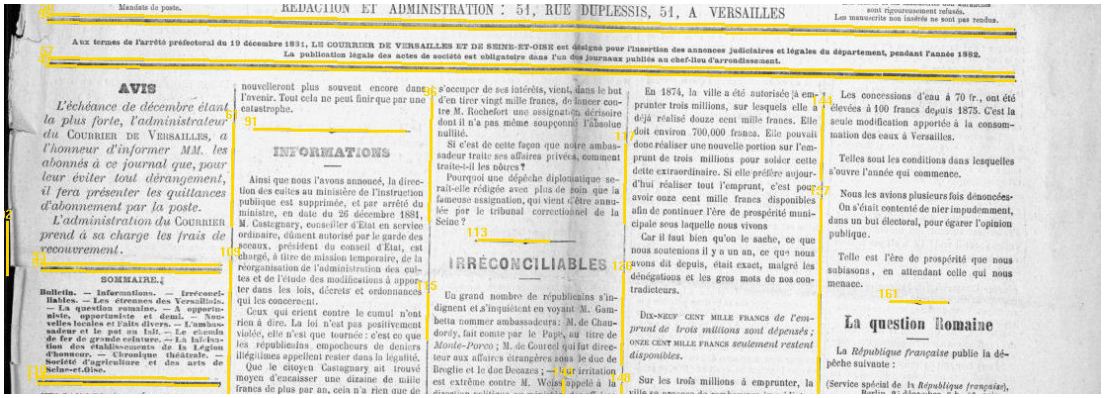
La grammaire de reconnaissance des traits en monorésolution TRAIT_MONO a donc pour rôle de regrouper les segments alignés pour construire des filets typographiques. Cependant, ce regroupement nécessite des seuils de distance, souvent spéci-

fiques au type de documents étudiés. Afin de préserver la généricité de la grammaire TRAIT_MONO, nous avons limité au minimum les regroupements des segments, afin de laisser la possibilité à l'utilisateur de cette grammaire d'affiner l'utilisation des traits en fonction du type de documents étudiés.

La figure 11.3 présente l'effet produit par la grammaire, c'est à dire la combinaison des segments perçus dans la résolution initiale pour produire des traits plus élaborés.



(a) Segments reconnus dans l'image de résolution initiale



(b) Traits construits par la grammaire TRAIT_MONO, avec DMOS : les filets sont mal reconnus (morcèlement, dédoublement)

FIG. 11.3 – Exemple de résultat produit par la grammaire de reconnaissance des traits en monorésolution TRAIT_MONO (les numéros permettent d'identifier les segments)

L'exemple de la figure 11.3 met en avant plusieurs limites de l'approche monorésolution :

– certains filets fins sont morcelés,

– certains filets épais sont dédoublés.

Cela signifie que pour utiliser les traits produits avec cette méthode, il faudrait prévoir dans la description le fait que les filets comportent des erreurs de morcèlement ou de dédoublement.

11.2.2 Approche perceptive

L'approche perceptive pour la reconnaissance des traits a été largement détaillée dans la partie 7.1.2. Grâce à sa généricité, nous la réutilisons sans avoir besoin d'y apporter de modifications ni d'ajustement de code. Nous rappelons les calques perceptifs utilisés pour cette approche sur la figure 11.4.

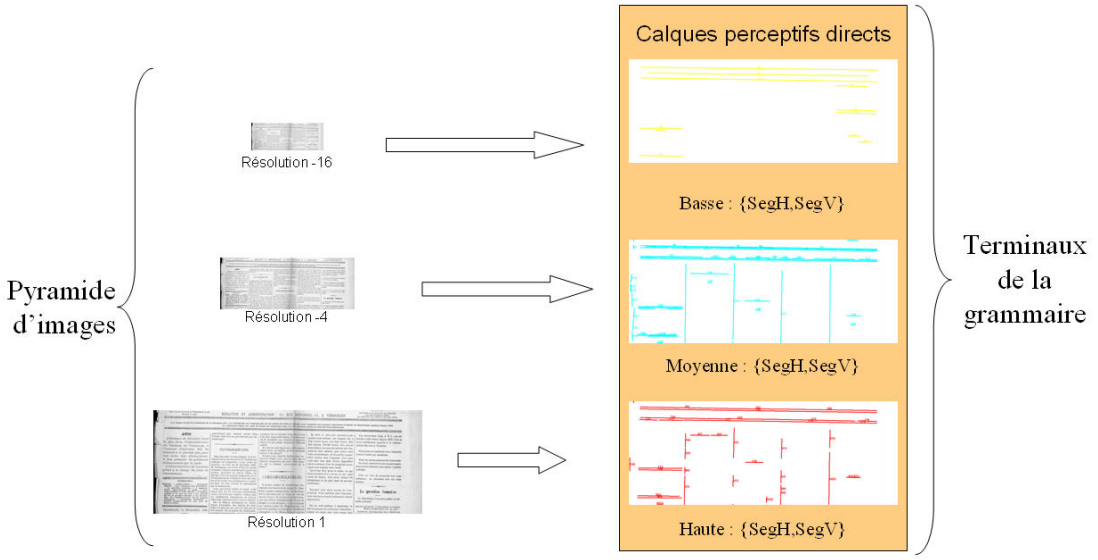


FIG. 11.4 – Calques utilisés pour la reconnaissance des traits avec une approche perceptive

Nous rappelons que l'avantage de cette méthode est de combiner les différents niveaux de vision pour reconnaître les traits par un système de prédiction/vérification. Un exemple de résultat produit est présenté sur la figure 11.5.

11.2.3 Evaluation des résultats

Nous comparons maintenant les résultats obtenus par la méthode monorésolution, face à l'approche perceptive.

11.2.3.1 Base de test

Afin de constituer une base de test, nous avons étiqueté manuellement 4967 filets à reconnaître parmi 157 pages de journaux, issus de la base présentée dans le paragraphe 11.1.

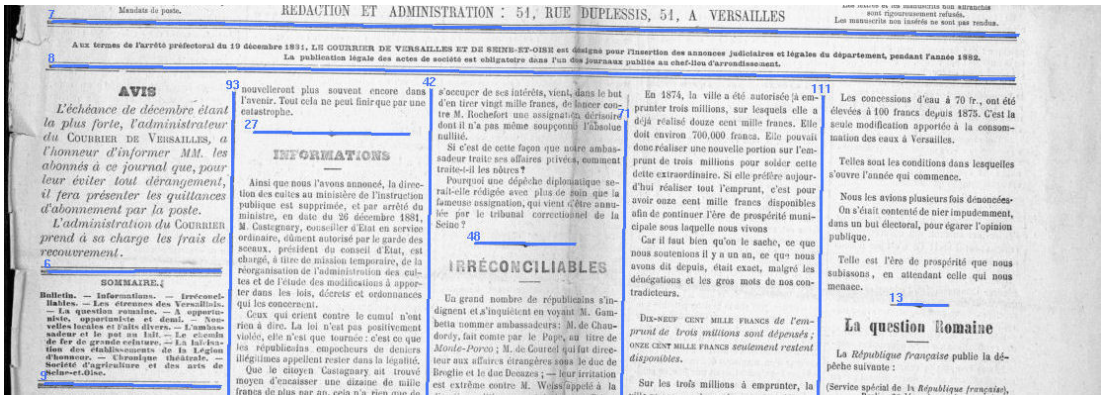


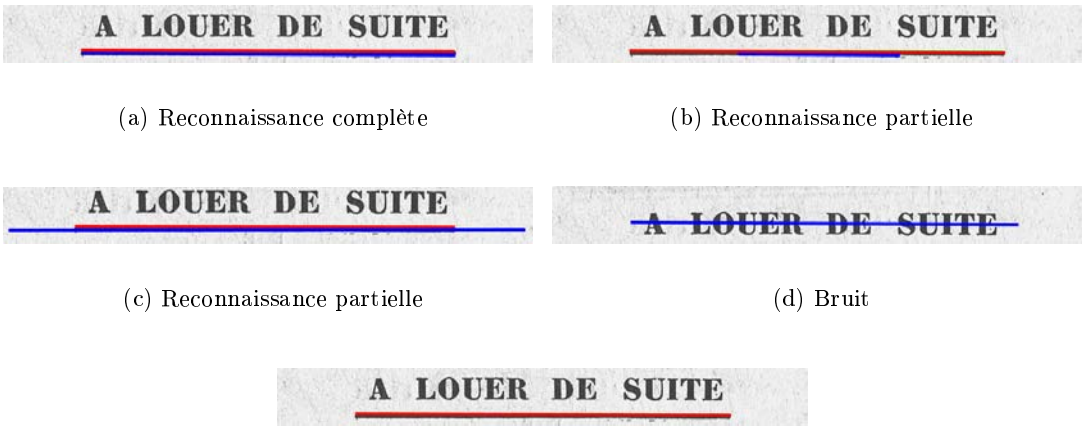
FIG. 11.5 – Exemple de traits reconnus avec l'approche perceptive (à comparer avec la version monorésolution sur la figure 11.3(b))

Nous avons ensuite appliqué les deux méthodes de reconnaissance des traits présentées dans le paragraphe 11.2 pour extraire les filets contenus dans ces images.

11.2.3.2 Métrique utilisée

Pour chaque image, nous avons dû comparer une liste de traits attendus avec une liste de traits reconnus.

Nous mettons en avant, sur la figure 11.6, les différentes configurations possibles de mise en correspondance des éléments attendus et reconnus.



(a) Reconnaissance complète

(b) Reconnaissance partielle

(c) Reconnaissance partielle

(d) Bruit

(e) Oubli

FIG. 11.6 – Les différentes configurations entre les éléments attendus (en rouge) et reconnus (en bleu)

La comparaison entre les éléments attendus et reconnus est basée sur la position de leurs extrémités, avec une zone de tolérance proportionnelle à la longueur du trait attendu. Nous proposons d'utiliser la métrique suivante :

- la *reconnaissance complète* correspond au nombre de traits attendus qui ont été correctement reconnus (figure 11.6(a)),
- la *reconnaissance partielle* correspond au nombre de traits qui ont été reconnus partiellement, c'est à dire que le trait a été reconnu soit plus court (figure 11.6(b)), soit plus long (figure 11.6(c)) que la vérité terrain,
- le *bruit* correspond au nombre de traits reconnus qui n'ont pas de correspondance avec un élément attendu (figure 11.6(d)),
- l'*oubli* correspond au nombre de traits attendus qui n'ont pas été reconnus (figure 11.6(e)).

11.2.3.3 Résultats

Le tableau 11.1 regroupe les résultats obtenus pour la reconnaissance des traits, d'une part avec la grammaire monorésolution, d'autre part avec la méthode perceptive. Les taux proposés sont donnés par rapport au nombre de traits attendus.

Méthode	Monorésolution	Vision perceptive
Reconnaissance complète	69.09%	94.42%
Reconnaissance partielle	21.22%	3.02%
Oublis	9.67%	2.55%
Bruit	93.85%	31.14%
Temps par image	10.9 sec	15.4 sec

TAB. 11.1 – Evaluation de la reconnaissance des traits avec une ancienne méthode monorésolution et notre approche perceptive, sur 4967 filets dans 157 pages de journaux

Ces résultats mettent en évidence l'intérêt de la vision perceptive pour la reconnaissance des traits. En effet, avec la méthode perceptive de DMOS-P, 94.42% des traits sont entièrement reconnus, contre seulement 69.09% avec l'approche monorésolution. Le temps de calcul est plus important pour la méthode perceptive. Ceci est du au besoin de combiner des segments issus de résolutions différentes. Cependant, la grande amélioration de taux de reconnaissance compense, dans le cas de nos travaux, le temps d'exécution plus long.

Le faible taux d'oublis pour l'approche perceptive (2.55% contre 9.67%) met en évidence la meilleure reconnaissance des filets épais et bruités qui sont mieux perçus à résolution faible qu'à résolution haute.

Le fort taux de reconnaissance partielle pour l'approche monorésolution (21.22% contre 3.02%) met en évidence l'intérêt d'être guidé par des hypothèses émises à basse résolution, pour pouvoir combiner des segments fins à haute résolution, afin de former un seul trait.

Le taux de bruit plus faible avec l'approche perceptive (31.14% contre 93.85%) est permis par la stratégie de prédiction/vérification qui permet de ne pas valider la présence

d'un trait si celui-ci n'a pas été perçu à au moins deux résolutions. Le bruit important restant est principalement dû aux segments verticaux détectés à plusieurs résolutions dans les lettres majuscules du titre du périodique. L'introduction de connaissance plus approfondie sur les pages de journaux permettrait d'en tenir compte.

Devant les faibles résultats obtenus par l'analyse monorésolution dans la résolution initiale, nous nous sommes demandé si une analyse monorésolution dans une résolution plus faible ne fonctionnerait pas mieux. Nous avons donc expérimenté notre système monorésolution en changeant la résolution étudiée. Les résultats obtenus sont présentés dans le tableau 11.2

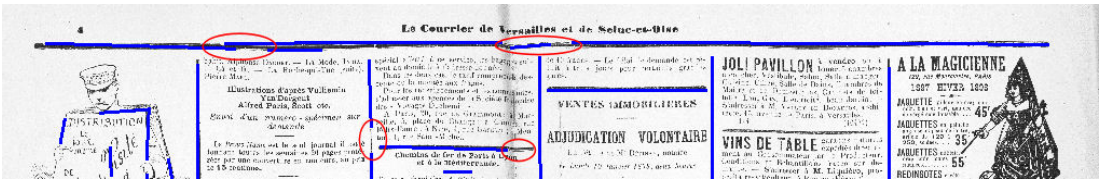
Méthode Résolution	Monorésolution 1	Monorésolution -4	Monorésolution -16
Reconnaissance complète Reconnaissance partielle	69.09% 21.22%	52.62% 8.53%	5.57% 2.41%
Oublis	9.67%	38.83%	92.01 %
Bruit	93.85%	16.54%	2.49 %
Temps par image	10.9 sec	3.3 sec	2.6 sec

TAB. 11.2 – Comparaison des approches monorésolutions, basées sur des résolutions différentes, sur 4967 filets dans 157 pages de journaux

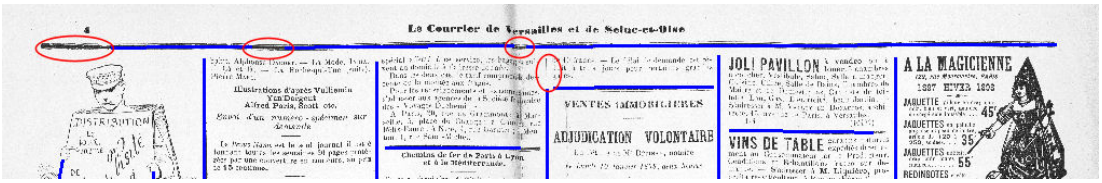
Cette expérience est illustrée par un exemple sur la figure 11.7, qui compare les traits extraits aux différentes résolutions avec ceux construits selon l'approche perceptive. La résolution 1 (figure 11.7(a)) permet de reconnaître de nombreux traits mais rencontre des difficultés face aux discontinuités : de nombreux traits sont reconnus partiellement ; quelques traits trop épais sont oubliés. La résolution -16 (figure 11.7(c)), en revanche, permet de reconnaître complètement et correctement les traits épais, mais occulte tous les traits fins. La résolution -4 (figure 11.7(b)) présente des résultats intermédiaires qui ne sont pas parfaits.

Les résultats du tableau 11.2 montrent que les segments perçus à faible résolution sont utiles pour la reconnaissance. En effet, les faibles résolutions sont caractérisées par des faibles taux de bruit. Ceci est dû à la présence de traits épais, qui sont perçus uniquement à faible résolution, car trop mouchetés pour être détectés dans une vision détaillée. En revanche, si on ne peut s'intéresser qu'à une résolution, la résolution initiale est bien celle qui offre le meilleur taux de reconnaissance complète : 69.09%.

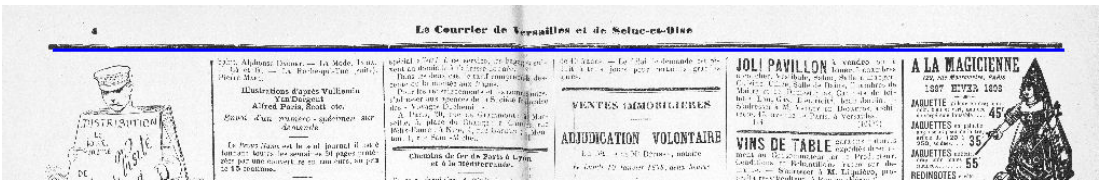
Ces résultats complémentaires ne font que renforcer l'intérêt de la comparaison présentée dans le tableau 11.1, qui met en avant les meilleurs résultats obtenus grâce à l'approche perceptive, illustrés sur la figure 11.7(d).



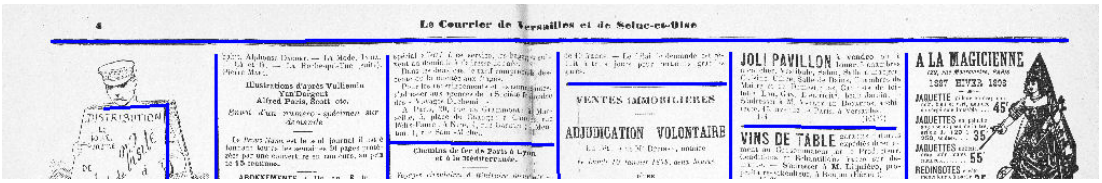
(a) Traits construits en monorésolution, à partir de la résolution 1 ; sur-segmentations entourées en rouge



(b) Traits construits en monorésolution, à partir de la résolution -4 ; sur-segmentations entourées en rouge



(c) Traits construits en monorésolution, à partir de la résolution -16 ; oublis de nombreux traits fins



(d) Traits construits selon l'approche perceptive

FIG. 11.7 – Comparaison des traits construits selon différentes méthodes : intérêt de l'approche perceptive

11.3 Découpage en cases

Munis de traits reconnus lors de l'étape présentée ci-dessus, nous proposons d'effectuer un découpage récursif des pages, selon les lignes et les colonnes.

11.3.1 Principe de reconnaissance

Dans l'optique du découpage en cases, nous considérons les traits comme des éléments prégnants qui ont pu être reconnus de manière indépendante au type de documents. Nous créons ensuite une grammaire dédiée au découpage en cases, JOURN, basée sur ces traits.

11.3.1.1 Base : les traits

La grammaire de découpage JOURN se base donc sur le calque perceptif induit des traits. Cependant, grâce à la généricité de notre approche, la grammaire JOURN n'a pas besoin d'avoir de connaissance sur la manière dont sont construits les traits contenus dans le calque perceptif induit.

Ainsi, le contenu du calque perceptif des traits peut avoir été construit soit de manière monorésolution (figure 11.8(a)), soit selon la vision perceptive (figure 11.8(b)). Dans les deux cas, la grammaire JOURN utilise de manière « aveugle » le contenu du calque de traits.

11.3.1.2 Découpage récursif

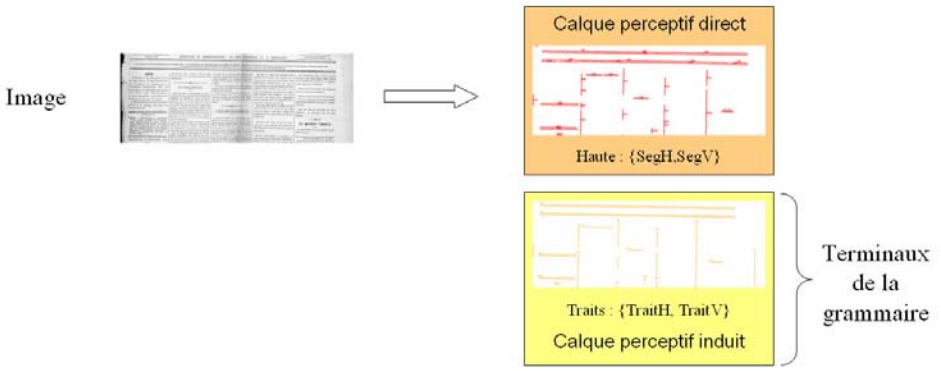
Nous présentons le principe de la grammaire JOURN, chargée d'effectuer le découpage récursif de la page en cases selon les traits trouvés.

Le principe général est illustré sur la figure 11.9. L'analyse se base sur une case initiale qui correspond à la page complète (figure 11.9(a)). Dans cette case, on cherche des séparateurs horizontaux ou verticaux correspondant à toute la largeur de la case (figure 11.9(b)). Ces séparateurs vont être utilisés pour produire autant de sous-cases (figure 11.9(c)). Chacune des sous-cases produites est alors récursivement découpée selon le même principe (figures 11.9(d) et 11.9(e)). L'analyse se termine lorsque chacune des cases est découpée au maximum (figure 11.9(f)).

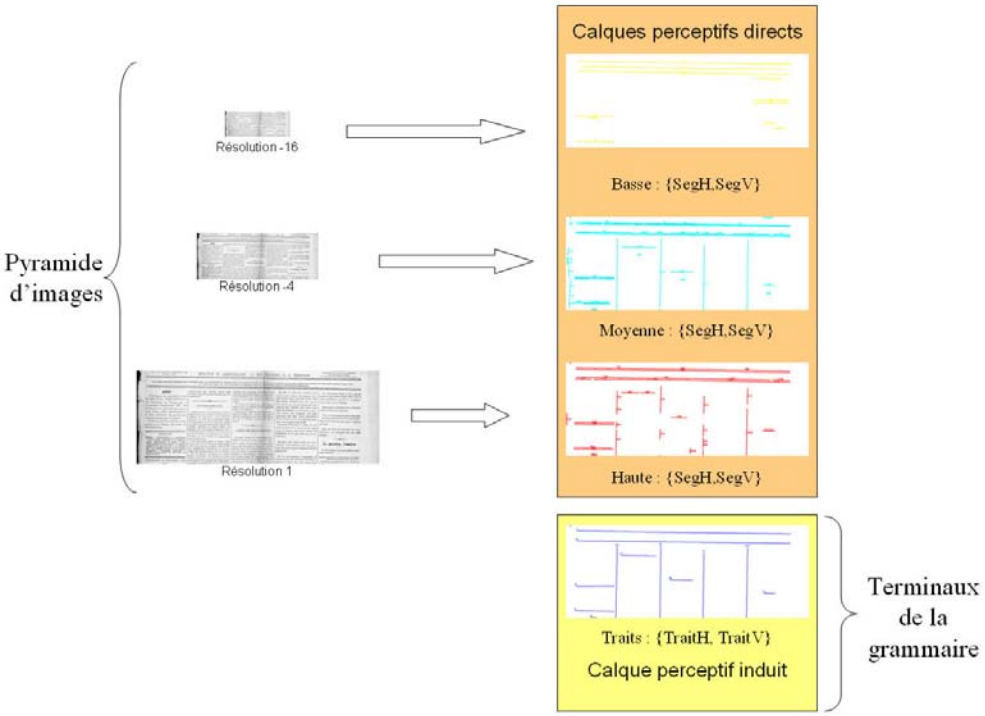
La description associée dans le langage EPF est très simple.

La case de départ est constituée de la page complète de journal, `CaseInit`. Dans cette case, on cherche les séparateurs `Traits` horizontaux ou verticaux , que l'on utilise pour découper la case en un ensemble de cases filles `CasesFille`. Chacune des sous-cases ainsi produite est ensuite découpée de manière récursive.

```
separationCase CaseInit ::=
     IN(dansZoneCase CaseInit) DO(listeSeparateurs CaseInit Traits) &&
    consListeCase CaseInit Traits CasesFilles &&
     separationCase CasesFilles.
```



(a) Calque perceptif des traits construit à partir de la monorésolution



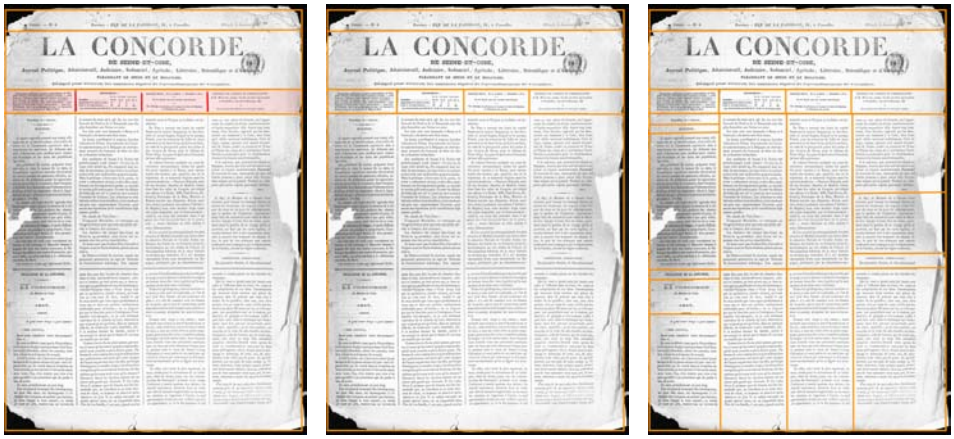
(b) Calque perceptif des traits construit par vision perceptive

FIG. 11.8 – Calques perceptifs pouvant être utilisés pour le découpage du journal en cases

(a) Case initiale

(b) Recherche de séparateurs horizontaux

(c) Construction des cases associées



(d) Recherche des séparateurs verticaux dans la sous-case sélectionnée

(e) Construction des cases associées

(f) Résultat final

FIG. 11.9 – Principe récursif de découpage d'une page en cases

L'analyse de la liste de séparateurs `listeSeparateurs` se fait simplement par la recherche de terminaux dans le calque `Traits`.

```
separateurHoriz CoorDeb CoorFin Case  Sep ::=
     AT(zoneCase CoorDeb CoorFin) &&
     USE_LAYER(Traits) FOR (
       TERM_SEG (condLigneHoriz CoorDeb CoorFin Case) noCondS filet Sep).
separateurVerti CoorDeb CoorFin Case  Sep ::=
     AT(zoneCase CoorDeb CoorFin) &&
     USE_LAYER(Traits) FOR (
       TERM_SEG (condLigneVerti CoorDeb CoorFin Case) noCondS filet Sep).
```

La figure 11.10 présente des exemples de cases extraites grâce à cette méthode.

11.3.2 Application

Nous proposons de comparer les résultats obtenus par la grammaire de découpage en cases, appliquée dans deux conditions :

- avec le calque perceptif des traits construits en monorésolution,
- avec le calque perceptif des traits construits selon l'approche perceptive.

11.3.2.1 Base de test

Nous appliquons la grammaire de découpage de pages de journaux en cases sur un échantillon représentatif des quatre périodiques présentés dans la partie 11.1. Ainsi, nous avons sélectionné une page par an pour chacun des périodiques et chacun des numéros disponibles.

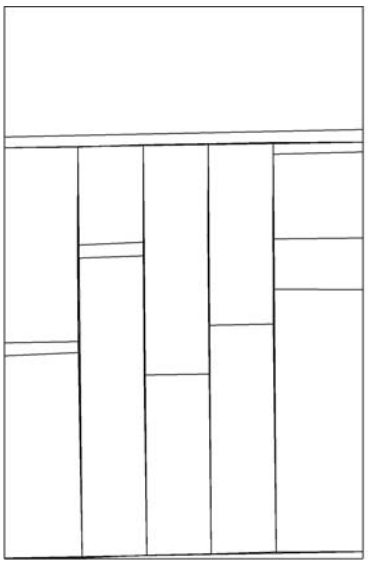
Nous avons constaté que certaines pages présentaient plus de traits difficiles à reconnaître que d'autres. Ainsi, dans les premières pages de journaux, l'épaisseur des filets est relativement constante, alors que les dernières pages contiennent davantage d'encarts publicitaires avec des lignes d'épaisseurs variable. Nous avons donc créé deux bases : une base constituée de 179 premières pages de journaux, dans laquelle nous avons établi manuellement une vérité terrain constituée de 4148 cases, et une base contenant 79 dernières pages de journaux et 3480 cases.

11.3.2.2 Métrique utilisée

La comparaison entre les résultats et la vérité terrain est en fait un problème de sur et de sous-segmentation. Nous utilisons donc la métrique proposée par Silva [CeS07], basée sur les notions de complétude et de pureté, qui semble bien adaptée à ce problème de segmentation. Nous en rappelons la définition.

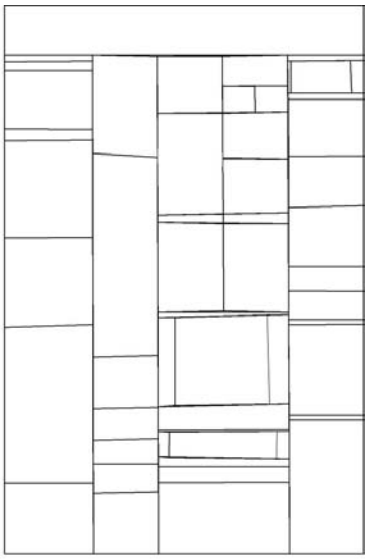
Lorsqu'une case présente dans le vérité terrain est trop segmentée par la méthode (sur-segmentation), elle est *incomplète*. L'*incomplétude* est la proportion de cases attendues trouvées de manière incomplète par rapport au nombre total de cases attendues.

(a) Exemple de première page de journal



(b) Cases extraites dans la première page

(c) Exemple de dernière page de journal



(d) Cases extraites dans la dernière page

FIG. 11.10 – Exemple de segmentation de pages en cases

Lorsque deux cases de la vérité terrain sont regroupées en une seule dans le résultat produit (sous-segmentation), cette case reconnue est dite *impure*. L'*impureté* correspond au taux de cases impures reconnues, par rapport au nombre total de cases reconnues.

Les taux d'*incomplétude* et d'*impureté* doivent être les plus petits possible.

11.3.2.3 Résultats

Les résultats obtenus sur la base des premières pages sont présentés dans le tableau 11.3, et ceux sur la base des dernières pages dans le tableau 11.4.

Les apports de la vision perceptive sont plus marquants sur la base de dernières pages. En effet, c'est sur ces pages que les filets sont particulièrement difficiles à reconnaître. Sur cette base, la version perceptive, diminue l'impureté (sous-segmentation) de 45%, tout en diminuant l'incomplétude (sur-segmentation) de 20%.

Version	Cases	Incomplétude	Impureté
Monorésolution	4148	10.46%	10.73%
Vision perceptive	4148	10.17%	7.87%
Gain		- 3%	- 33%

TAB. 11.3 – Application de l'extraction de filets pour le découpage de 179 premières pages de journaux en cases (filets de bonne qualité)

Version	Cases	Incomplétude	Impureté
Monoresolution	3480	17.06%	11.34%
Vision perceptive	3480	13.70%	6.23 %
Gain		- 20%	- 45%

TAB. 11.4 – Application de l'extraction de filets pour le découpage de 79 dernières pages de journaux en cases (filets variés et dégradés)

Ces résultats sont à utiliser uniquement dans un but de comparaison de deux méthodes de détection de lignes, et non pour évaluer les performances d'une segmentation de pages de journaux. En effet, on pourrait améliorer la grammaire de description des pages de journaux, en tenant compte de certaines spécificités liées à la presse, si on voulait obtenir de meilleurs résultats quant à la segmentation en cases.

De plus, nous pouvons noter que ces résultats ne reflètent pas entièrement l'écart du taux de reconnaissance entre les traits, présentés dans le tableau 11.1. Cela provient principalement d'une des limites de la métrique qui ne permet pas d'évaluer correctement le degré d'impureté et d'incomplétude au sein de cases impures et incomplètes. Ainsi, une case segmentée à tort en deux partie comptera pour une incomplétude, mais une case segmentée à tort en dix parties comptera au même titre pour une incomplétude. Cependant, il nous semble plus grave de découper une case en dix qu'en deux. La métrique proposée par Silva [CeS07] n'est donc en réalité pas entièrement adaptée à notre problème.

Néanmoins, les résultats permettent de montrer l'intérêt de la vision perceptive.

11.4 Transfert industriel

La méthode DMOS-P a fait l'objet d'un transfert industriel avec la société Evodia. Ce transfert s'est concrétisé par le développement, par Evodia, d'un système permettant la reconnaissance de la structure et du contenu des pages de presse ancienne.

Afin de créer leur système perceptif adapté à leurs besoins applicatifs, les ingénieurs d'Evodia ont dû prendre en main la méthode DMOS-P. Ceci a été relativement simple puisque, pour des utilisateurs connaissant le principe de fonctionnement de DMOS, les fonctionnalités nouvelles offertes par DMOS-P sont cohérentes avec la philosophie globale.

La description perceptive proposée par Evodia exploite pleinement les possibilités de coopération entre les différents calques. En effet, cette description est basée sur l'étude de trois calques directs issus d'images à trois résolutions, ainsi que du calque induit contenant les traits.

De plus, la grande souplesse du formalisme des calques perceptifs a permis aux ingénieurs d'Evodia de définir un nouveau niveau de perception. En effet, ils souhaitaient utiliser le résultat de l'application locale d'un OCR commercial (Fine Reader) comme terminaux de la grammaire. Ceci a pu être permis par simple ajout d'un nouveau calque perceptif direct contenant l'étiquetage des composantes connexes réalisé par l'OCR. Ce nouveau calque forme ainsi une extension cohérente au système perceptif.

Les calques utilisés par le système perceptif sont donc au nombre de 5, et sont regroupés sur la figure 11.11. L'introduction du nouveau calque OCR permet d'utiliser les informations fournies par FineReader, telles que la séparation texte/images, la structuration grossière des tableaux, la reconnaissance de mots clés. Le calque OCR constitue ainsi une nouvelle perception particulière de l'image, qui peut interagir avec les informations contenues dans les autres calques.

Cette méthode a été appliquée à l'heure actuelle sur environ 45 000 pages de journaux anciens et sera utilisée d'ici la fin 2008 pour 50 000 autres documents. Les résultats du découpage des pages sont mis à disposition des lecteurs grâce à une plateforme de consultation présentée sur la figure 11.12. Cette plateforme permet d'effectuer des recherches plein texte et donne accès à un système d'annotations collectives qui vient compléter les annotations produites automatiquement.

11.5 Intérêts de notre approche

Nos travaux sur la presse ont permis de valider l'utilité du calque perceptif induit contenant les traits. En effet, les résultats obtenus montrent nettement l'intérêt de combiner la vision à plusieurs résolutions pour reconnaître des traits de types variés.

De plus, nous pouvons noter la souplesse de notre approche basée sur les calques perceptifs. Dans le cas de la presse, la grammaire d'analyse se base sur les traits construits dans le calque induit sans avoir connaissance de la manière dont ces traits ont été construits. Cette application illustre donc la coopération entre l'attention guidée par

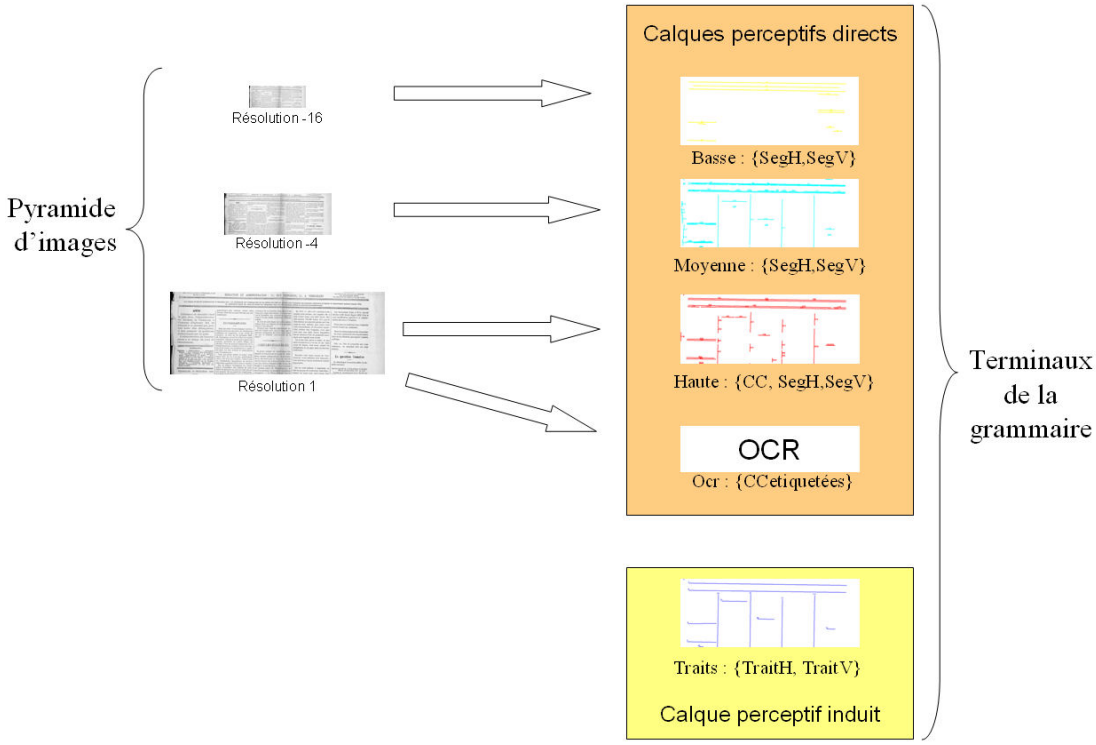


FIG. 11.11 – Calques perceptifs utilisés dans la description réalisée par Evodia

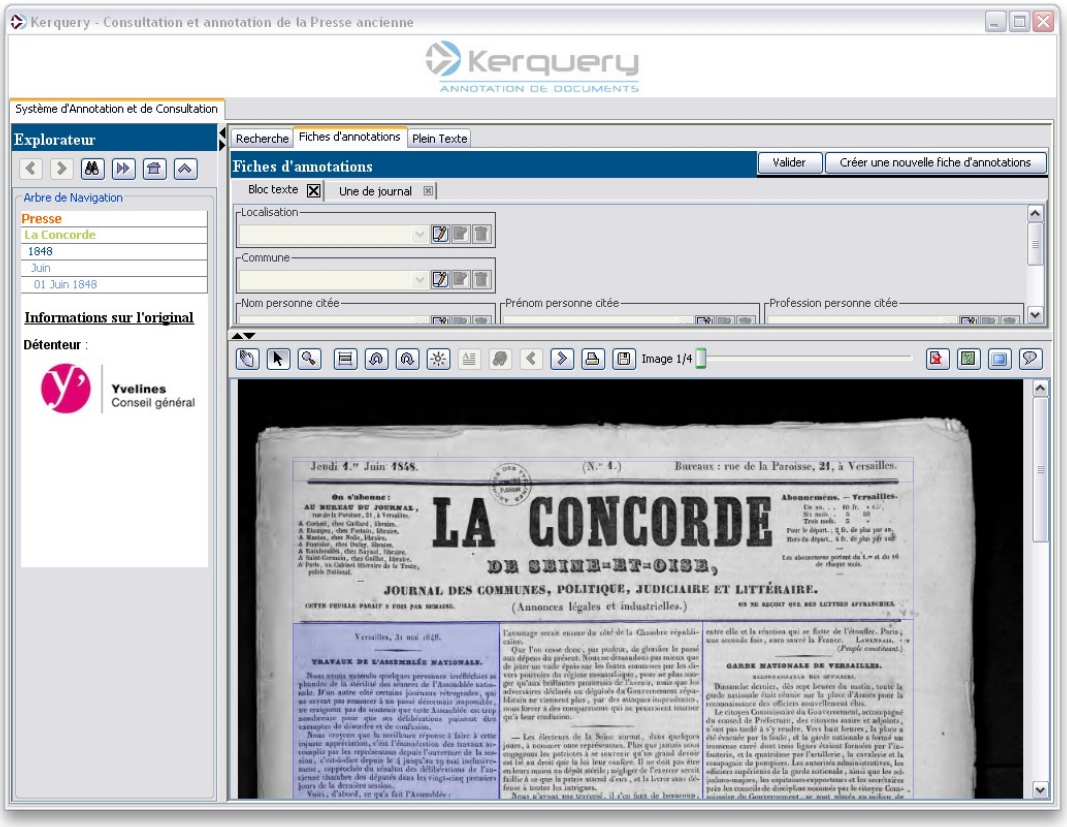


FIG. 11.12 – Plateforme de consultation proposée par Evodia : recherche textuelle, visualisation des différents types de cases et accès à des fiches d'annotations

des éléments prégnants, et l'attention guidée par un but.

Les travaux sur la presse ont été le support du transfert industriel de la méthode DMOS-P. La simplicité de sa prise en main pour des ingénieurs connaissant la méthode DMOS, et le grand pouvoir d'expression permis par les nouveaux formalismes liés à la vision perceptive ont permis la création d'un nouveau système complexe, dédié à la presse. Ce système décrit une coopération perceptive adaptée au problème précis de la structuration et de l'indexation de la presse, en incluant un nouveau niveau de perception provenant d'une analyse OCR.

Enfin, l'étude des pages de presse permet de mettre en avant les apports de la vision perceptive pour l'analyse de documents fortement structurés et bruités, confortant ainsi l'hypothèse émise dans la partie 3.2.1. En effet, l'utilisation du calque perceptif des traits vus comme des éléments prégnants permet de retrouver plus facilement les éléments structurels. Grâce à l'utilisation de ces éléments prégnants, la *sélection* d'une structure précise, ici le découpage en cases, est simplifié.

Conclusion de la troisième partie

Dans cette troisième partie, nous avons présenté des applications de la vision perceptive pour des documents de nature variée :

- des documents manuscrits récents : les courriers entrants,
- des documents anciens manuscrits et imprimés : les décrets de naturalisation,
- des documents manuscrits récents avec un alphabet différent : le bangla,
- des documents imprimés anciens : la presse.

Validation de la méthode DMOS-P

Grâce à la méthode DMOS-P, nous avons proposé, pour chaque problème traité, un mécanisme perceptif adapté à l'objectif applicatif. En effet, le formalisme de calque perceptif permet de gérer de manière modulaire les éléments nécessaires à la description de la structure. Au travers des différentes applications, nous avons validé de manière unitaire les différents composants de DMOS-P, mais aussi leurs imbrications qui permettent de créer des systèmes complexes.

Les travaux sur les courriers ont permis d'une part de rappeler l'intérêt d'une approche grammaticale, et d'autre part de montrer la simplification apportée par l'utilisation du calque induit contenant les lignes de texte reconnues comme des éléments prégnants.

L'analyse des décrets de naturalisation a montré un exemple de coopération forte entre les différents niveaux de résolution, en étant totalement guidé par la connaissance. Là encore, cet application démontre la simplification permise par l'approche perceptive, à la fois pour la description réalisée par l'utilisateur, et pour la combinatoire lors de l'exécution.

Les travaux sur la recherche des lignes de base ont montré l'intérêt des outils de mise en correspondance entre résolutions, tant au niveau de la localisation que de la fusion de primitives de natures différentes.

Les applications développées autour de la presse ont permis d'une part de valider l'approche des traits perçus comme des éléments prégnants. D'autre part, le transfert technologique et la prise en main par de nouveaux utilisateurs a permis de valider la cohérence de nos formalismes par rapport à la méthode initiale DMOS. Enfin, l'utilisation d'un nouveau niveau de perception par Evodia démontre la souplesse du formalisme des calques perceptifs.

La variété des documents et des problématiques traitées démontre la généricité de

notre approche. De plus, notre méthode a été appliquée à grande échelle, puisqu'elle a été testée sur plus de 86 000 pages de documents, puis utilisée de manière industrielle sur 45 000 pages supplémentaires.

Intérêt de la vision perceptive

Les différents types de documents étudiés sont représentatifs des difficultés couramment rencontrées dans la littérature, recensées dans le chapitre 3 et regroupées dans le tableau 11.5. En effet, les courriers manuscrits sont des documents faiblement structurés, dans lesquels nous avons également proposé le positionnement de lignes de base. Les décrets de naturalisation sont des documents faiblement structurés et bruités. Les pages de presse sont des documents bruités et très structurés. Enfin, les documents banglas ont été le support au positionnement d'éléments structurels : les *headlines*.

	Courriers	Décrets	Presse	Bangla
Documents bruités		x	x	
Documents très structurés			x	
Documents faiblement structurés	x	x		x
Recherche de lignes de base	x			x

TAB. 11.5 – Synthèse des types de problèmes rencontrés

Dans le cas des documents bruités et très structurés, l'utilisation d'informations prégnantes en coopération avec l'attention guidée par un but a permis de simplifier la description en *sélectionnant* uniquement les informations utiles à la reconnaissance de la structure.

Dans le cas des documents pour lesquels l'information structurelle physique est diffuse, c'est à dire les documents faiblement structurés ou pour le positionnement de lignes de base, l'utilisation du principe de prédiction/vérification de la vision perceptive a permis de *reconstituer* des éléments structurels à partir d'indices globaux et de positionnements locaux.

Ces applications permettent donc de démontrer les intuitions que nous avions indiquées dans le chapitre 3.

Les résultats obtenus ont donné lieu à plusieurs publications. La description multirésolution des lignes de texte et des traits est réalisée respectivement dans [1] et [4]. Notre participation au concours RIMES est décrite dans [5]. [3] présente l'enrichissement de DMOS appliqué aux cas des décrets de naturalisation. Les travaux en collaboration avec B.B. Chaudhuri, de l'Indian Statisical Institute de Calcutta sont présentés dans [6].

Tous ces travaux applicatifs ont permis de synthétiser les apports de la vision perceptive pour la reconnaissance de la structure de documents, dans un article de la revue IJDAR [2].

Conclusion générale

Ces travaux ont permis de produire et de valider une méthode générique de reconnaissance de la structure de documents, s'appuyant sur des mécanismes s'inspirant de la vision perceptive. Grâce à cette méthode, nous avons montré les apports de la vision perceptive pour plusieurs problèmes rencontrés classiquement en analyse de documents.

Concepts de la vision perceptive

Dans la première partie, nous avons décrit les principes généraux de la vision perceptive. Nous avons identifié deux composantes utiles : le cycle perceptif et l'attention visuelle qui elle-même peut prendre deux formes selon qu'elle est guidée par des éléments prégnants ou par la recherche d'un modèle précis.

Nous avons proposé de créer un système perceptif complet qui s'inspire du cycle perceptif tout en combinant les deux formes d'attention visuelle. En effet, l'attention guidée par la prégnance permet notamment de structurer des documents sans connaissance *a priori* sur leur contenu. L'attention guidée par un but précis permet, en décrivant des modèles spécifiques, de limiter l'impact du bruit.

Plusieurs approches ont déjà été proposées dans la littérature, qui cherchent à s'inspirer de l'attention visuelle, parfois même en proposant une interprétation de manière cyclique. Cependant, à notre connaissance, ces méthodes imitent uniquement l'une des deux formes d'attention. L'utilisation combinée des deux formes d'attention est donc une nouveauté par rapport aux approches de la littérature. Elle permet une description plus naturelle des documents et améliore également les performances de reconnaissance.

Une architecture complète de vision perceptive : DMOS-P

Création d'une nouvelle version de la méthode DMOS

Dans la deuxième partie, nous avons recensé les propriétés requises par un système générique de vision perceptive incluant à la fois le cycle perceptif et l'attention visuelle. A l'issue du chapitre 4, nous avons proposé une solution de mise en œuvre qui est une nouvelle version d'une méthode existante, la méthode DMOS. En analysant DMOS

sous l'angle de la vision perceptive, nous avons montré que cette méthode offrait déjà implicitement plusieurs bonnes propriétés pour la création d'une architecture perceptive.

Utilisation de données multirésolution

Pour produire la méthode DMOS-P, nous avons introduit la gestion de la multirésolution. Ainsi, nous avons proposé de baser l'analyse sur une pyramide d'images, qui constituent différents niveaux de perception.

Afin d'en organiser le contenu, nous avons proposé un formalisme spécifique : le *calque perceptif*. L'introduction de ce formalisme est un des atouts majeurs pour la simplicité et la modularité des descriptions que peut réaliser un utilisateur de DMOS-P. Un nouvel opérateur du langage EPF, `USE_LAYER` permet de choisir le calque utilisé à chaque étape de l'analyse.

Nous avons également introduit des abstractions, la *ligne abstraite* et le *rectangle abstrait* qui permettent de manipuler des objets de manière indépendante de la résolution de l'image, et donc de faciliter la mise en correspondance d'éléments issus de résolutions différentes.

Intérêt des éléments prégnants

Dans le chapitre 7, nous avons présenté la gestion des éléments prégnants en nous appuyant sur les exemples des lignes de texte et des traits. Leur synthèse, basée sur la vision perceptive, fournit de nouvelles primitives pour le mécanisme de reconnaissance. Ces primitives peuvent être considérées comme des terminaux de la grammaire utilisables de manière transparente grâce au principe des calques perceptifs.

La détection de ces éléments prégnants utilise une fusion d'informations entre différents niveaux de perception, ce qui la rend plus fiable. Une description utilisant ces éléments est donc plus simple car il y a beaucoup moins besoin de tenir compte de l'existence de bruit.

Une architecture cohérente

Pour créer la méthode DMOS-P, nous avons donc introduit des éléments clés pertinents, qui par leur assemblage permettent de démultiplier les possibilités offertes initialement par DMOS. Cet enrichissement a été réalisé en respectant autant que possible la philosophie de la méthode initiale, afin de garder une architecture cohérente. Ceci a nécessité une bonne prise en main du fonctionnement interne de la méthode, ainsi qu'une étape de conception pour envisager la meilleure intégration possible.

DMOS-P : un générateur de systèmes perceptifs

Avec l'utilisation conjointe des calques perceptifs et des outils associés (opérateur EPF et abstractions), la coopération entre les niveaux de perception est totalement guidée par la connaissance symbolique. Pour chaque nouveau problème, l'utilisateur est donc libre de mettre en place un système perceptif adapté à la difficulté du document.

Grâce à la coopération entre différents niveaux de perception, la méthode DMOS-P offre un gain en puissance d'expression, mais aussi en facilité d'utilisation. En effet, les

applications ont montré qu'il est plus simple d'exprimer une connaissance directement au bon niveau de résolution, plutôt que de rentrer dans des détails non adaptés à l'objectif de reconnaissance. La qualité des résultats se trouve également améliorée.

De plus, l'architecture proposée est extensible, en permettant notamment d'intégrer des niveaux de perception supplémentaires. Ceci a été validé avec l'introduction d'un nouveau calque perceptif (OCR) pour la presse.

Validation

Les applications présentées nous ont permis de valider les différents axes de nos travaux : la généricité, l'aspect perceptif, l'application à grande échelle et industrielle.

Validation de la généricité

Nous avons présenté des applications de la vision perceptive pour des documents variés : manuscrits, imprimés, anciens, récents, avec un alphabet latin ou bangla. Nous avons également abordé des applications diverses : localisation de lignes de texte, reconstitution de filets, découpage et classification de blocs de texte, localisation de champs précis dans un document, positionnement de ligne de base dans l'écriture.

La grande variété des documents et des applications étudiées nous permet de valider l'aspect générique de notre approche perceptive. Cette généricité est issue d'une part de la séparation de la connaissance du reste du système et d'autre part de la modularité des concepts perceptifs qui peuvent se combiner pour produire des systèmes complexes adaptés à chaque problème.

Cette grande généricité est un apport important par rapport aux approches de la littérature qui se concentrent généralement sur un seul type de problème, en utilisant des mécanismes perceptifs dédiés, difficilement réutilisables dans un autre contexte applicatif.

Validation de l'aspect perceptif

Dans les applications, nous avons mis en évidence le mécanisme de prédiction/vérification qui est induit par le cycle perceptif. En effet, en imitant le cycle perceptif, nous avons mis en place un système de coopération entre niveaux de perception, guidé par la connaissance symbolique. Ainsi, une vision globale du document permet d'émettre des hypothèses sur la position et sur la nature des objets présents dans l'image, en tenant compte du contexte général du document. Ces hypothèses permettent ensuite une focalisation d'attention sur l'élément étudié, dans un contexte local issu de l'analyse globale. L'analyse locale de cet élément permet alors de vérifier les hypothèses émises. La méthode DMOS-P permet donc de fusionner des informations de nature hétérogène pour produire un tout cohérent.

Ce système permet à la fois :

- de *sélectionner* à basse résolution les éléments pertinents pour l'analyse, qui sont ensuite détaillés à haute résolution,

– mais aussi de *reconstituer* des éléments structurels à partir d'une vision globale, qui sont positionnés de manière plus fine à haute résolution.

Nous avons pu démontrer l'apport de ces deux aspects aux travers de nos expérimentations.

Les résultats chiffrés montrent que la vision perceptive améliore les taux de reconnaissance pour les différents problèmes étudiés. Les difficultés étudiées représentent généralement les derniers pourcentages d'erreur restant lors de l'évaluation des systèmes de reconnaissance dans la littérature ; la force de notre approche est donc, en plus de produire des résultats corrects sur des cas dits faciles, de pouvoir apporter une solution pour de nombreux cas complexes, qui sont les plus difficiles à résoudre dans l'état actuel des travaux en reconnaissance de structure de documents.

De plus, même si ce n'était pas notre but premier, nous pouvons noter généralement une nette amélioration des temps d'exécution. En effet, le principe de prédiction/vérification permet de diminuer la complexité des traitements, en diminuant les choix possibles à résolution haute grâce à ceux prédits à basse résolution.

Validation à grande échelle

La validation de la méthode DMOS-P a été réalisée à grande échelle puisqu'elle a été testée sur 86 637 pages de documents, puis appliquée de manière industrielle sur 45 000 autres pages.

Validation industrielle

La méthode DMOS-P a été validée de manière industrielle : elle a fait l'objet d'un transfert technologique vers la société Evodia. Ce transfert nous a permis de nous assurer que l'utilisation de DMOS-P reste facilement accessible aux concepteurs : la création par les ingénieurs d'Evodia (déjà utilisateurs de DMOS) d'un nouveau système perceptif dédié à l'analyse des pages de journaux est une preuve que cet objectif est atteint.

Bilan

En conclusion, notre architecture DMOS-P est une méthode générique et complète, qui permet de décrire simplement des mécanismes perceptifs pour la reconnaissance de documents structurés, adaptés à chaque type de problèmes.

Les éléments de base de cette méthode ont été validés séparément, mais aussi assemblés pour produire des descriptions complexes. La méthode a ensuite été appliquée à des documents variés, puis transférée vers l'industrie, pour une validation totale sur plus de 130 000 pages de documents.

Grâce à DMOS-P, nous avons créé des mécanismes de prédiction/vérification dont l'application montre l'intérêt de l'approche perceptive pour la reconnaissance de documents, notamment pour faciliter la sélection ou la reconstitution d'éléments structurels.

Perspectives

Nous présentons maintenant quelques axes pour les travaux futurs.

La structure des calques perceptifs offre une grande souplesse pour l'utilisation de niveaux de perception supplémentaires dans la méthode DMOS-P. Nous prévoyons donc, par exemple, de construire un calque induit contenant les images localisées dans une page de documents. Il nous semble en effet que la séparation texte/image peut être facilitée par une coopération entre plusieurs résolutions. D'autre part, les images construites pourraient être considérées comme prégnantes pour la suite de l'analyse. La reconnaissance des zones d'images pourraient servir de support pour la spécification des outils liés aux rectangles abstraits, qui n'ont pas été encore utilisés à l'heure actuelle.

Dans plusieurs de nos applications, les erreurs restantes laissent entrevoir les limites d'une analyse basée uniquement sur la structure. Dans ces cas, il serait nécessaire de reconnaître des mots clés de l'écriture pour pouvoir améliorer la reconnaissance. La méthode DMOS est prévue pour être interfacée avec un classifieur. Nous avons mené quelques expérimentations dans ce sens en interfaçant un classifieur de chiffres avec la grammaire décrivant les courriers. A l'heure actuelle, les résultats ne sont pas suffisants mais il nous semble que l'interfaçage avec un reconnaisseur d'écriture manuscrite est une piste importante pour améliorer encore les taux de reconnaissance.

Nous prévoyons également une nouvelle extension de la méthode DMOS-P, vers une quatrième dimension. En effet, si la méthode DMOS permet une analyse bidimensionnelle des documents, la méthode DMOS-P a permis de rajouter une troisième dimension qui consiste a prendre en compte plusieurs manières de percevoir une même image.

L'idée de la quatrième dimension consiste à prendre en compte l'ensemble d'une collection cohérente d'images. Ainsi, intégrer dans la méthode DMOS-P une quatrième dimension consisterait à construire une mémoire visuelle des images précédemment traitées. Cette mémoire serait un support pour l'apprentissage de caractéristiques telles que des zones de positionnements, des dimensions d'objets.

Grâce à l'apprentissage réalisé, la méthode DMOS-P pourrait être appliquée sur des grandes collections de documents, sans avoir besoin d'ajuster les paramètres manuellement. Cet apprentissage permettrait de lever les ambiguïtés dans les images difficiles.

Enfin, au vu de la généricité des mécanismes perceptifs offerts par DMOS-P, il serait envisageable d'appliquer notre méthode dans un autre contexte que celui de l'analyse de documents, par exemple pour la détection d'objets dans des images naturelles.

Annexes

Annexe A

Filtrage de Kalman appliqué à l'extraction de segments

Cet annexe présente le fonctionnement interne de l'extracteur de segments développé au sein de l'équipe Imadoc. Nous rappelons la théorie sur le filtre de Kalman et présentons son application pour l'extraction de segments.

Dans les parties suivantes, nous limitons notre présentation au cas des segments à tendance horizontale. Cependant, tous les mécanismes sont facilement transposables pour l'extraction de segments verticaux ou diagonaux.

A.1 Le filtrage de Kalman

Nous présentons le principe d'une approche par prédiction/vérification et les équations du filtre de Kalman qui sont basées sur ce principe.

A.1.1 Principe général de l'approche par prédiction/vérification

L'hypothèse de base est que tout objet observable peut être caractérisé par un état e_t, variant au cours du temps. L'approche permet de suivre l'évolution de l'état, de l'estimer et de le prédire. Un estimateur récursif linéaire est dont utilisé pour prédire l'état e_t d'après les états passés :

$$e_t = estimation\_lineaire(e_{t-1}, e_{t-2}, e_{t-3}, ..., e_0)$$

Une fois l'état e_t prédit, l'étape suivante consiste à le vérifier par une observation de l'état de l'objet à l'instant t, et de corriger l'estimateur en tenant compte de l'erreur entre la prédiction et l'observation mesurée.

A.1.2 Équations du filtre de Kalman

Le filtre de Kalman est une formalisation de l'approche de prédiction/vérification. C'est un estimateur linéaire récursif capable de prendre en compte les erreurs calculées entre les estimations et les observations, afin d'améliorer les estimations suivantes.

Cette méthode est basée sur deux modèles habituellement utilisés dans les applications du filtrage de Kalman [Sor85]. Nous rappelons brièvement les modèles mis en jeux :

- un modèle d'état S variant dans le temps,
- un modèle de mesure X, relié à S.

Ces deux modèles S et X sont représentés par un vecteur.

L'état prédit $\hat{S}$ est estimé grâce à l'équation d'évolution :

$$\hat{S}(t) = A(t-1).S(t-1) + W(t-1)$$

où A est la matrice d'évolution du vecteur S. W représente le modèle d'erreur, caractérisé par une moyenne nulle et une variance connue *a priori*.

La mesure prédite $\hat{X}$ est donnée par :

$$\hat{X}(t) = C(t).S(t) + N(t)$$

où C est une matrice utilisée pour déduire le vecteur de mesure X à partir du vecteur d'état S, et où N est une mesure de bruit. Il faut noter que W et N sont supposées non corrélées. On calcule également la matrice de covariance H de prédiction de l'erreur.

Lorsqu'une observation réelle $X(t)$ est réalisée, l'état S doit être mis à jour en considérant l'état prédit $\hat{S}$ et la mesure prédite $\hat{X}$.

$$S(t) = \hat{S}(t) + G(t).(X(t) - \hat{X}(t))$$

Le gain G pondère l'importance de la mesure relativement à l'état précédent. G est mis à jour après chaque mesure, tout comme la matrice de co-variance H de prédiction de l'erreur.

A.2 Application à l'extraction de segments

Afin d'appliquer la théorie présentée ci-dessus, nous devons déterminer le type d'objet à étudier, ses états possibles et son modèle d'évolution au cours du temps. Dans le cas d'un segment à tendance horizontale, une observation est un empan vertical, c'est-à-dire approximativement orthogonal à la direction du segment. L'index d'évolution t correspond à la colonne k de l'image.

La méthode contient deux modèles pour le traitement des images observées : un outil de prédiction/vérification basé sur le filtre de Kalman et une couche de contrôle permettant d'interpréter des cas attendus (discontinuités, croisements).

Le principe de prédiction/vérification est présenté globalement sur la figure A.1.

A.2.1 Extraction des observations

Le but est d'extraire des segments horizontaux dans des images en niveaux de gris. Ces lignes sont considérées comme des objets foncés sur un arrière-plan plus clair.

L'observation suivante est choisie parmi les empans présents dans la colonne suivante. L'empan suivant doit être cohérent avec l'état estimé, en tenant compte d'une possible erreur calculée par la matrice de co-variance H.

Lorsque l'observation est choisie, on en extrait l'épaisseur et le point milieu.

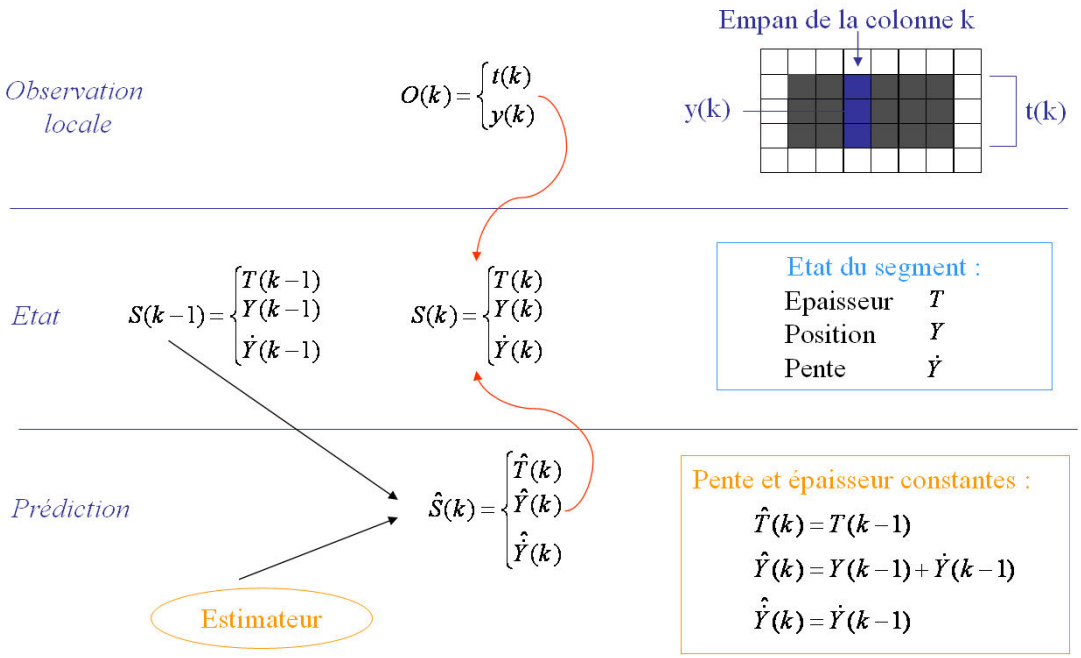


FIG. A.1 – Application du principe de prédiction/vérification

A.2.2 Filtres utilisés

L'extraction des segments est réalisée grâce à deux filtres de Kalman qui décrivent les valeurs estimées.

Le premier filtre représente l'épaisseur T du segment, qui est supposée constante. L'équation associée est :

$$\hat{T}(k) = T(k-1)$$

Le second filtre représente la position Y de la ligne, et sa pente $\dot{Y}$. Puisque nous cherchons de segments de droite, la pente est supposée constante. En conséquence, l'équation est donnée par :

$$\begin{bmatrix} Y(k) \\ \dot{Y}(k) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Y(k-1) \\ \dot{Y}(k-1) \end{bmatrix}$$

Cela signifie que l'ordonnée de la colonne k dépend de l'ordonnée de la colonne $k-1$. La pente est supposée être la même dans les colonnes k et $k-1$.

Même si les modèles de prédiction sont supposés représenter une pente constante et une épaisseur constante, ils peuvent évoluer lentement pendant l'analyse, en fonction des observations. Cette évolution est prise en compte par les paramètres d'erreur, W et N.

A.2.3 Interprétation

La phase d'interprétation consiste à trouver des relations entre les observations successives. Ces relations sont basées sur l'épaisseur et la position. Trois cas principaux sont possibles lors d'une nouvelle observation :

- le point prédit est noir, la position du point milieu et l'épaisseur sont correctes, *correct* étant défini par la matrice de co-variance. Dans ce cas, l'observation est intégrée, ce qui signifie que l'état estimé est mis à jour et que l'analyse continue ;
- le point prédit est noir avec une épaisseur trop large : le segment traverse probablement des objets plus larges (figures 5.4(b) and 5.4(c)), l'état n'est pas mis à jour mais l'analyse continue, en espérant trouver plus tard une autre colonne avec des observations correctes à intégrer ;
- le point prédit est blanc : il peut s'agir de la fin du segment, mais afin de gérer les lignes pointillées (figure 5.4(a)), l'analyse peut continuer pour un nombre donné de colonnes.

La détection des segments s'arrête au bout d'un certain nombre d'observations blanches.

A.2.4 Intérêts de cette méthode

Notre méthode présente plusieurs propriétés permettant de faire face aux difficultés rencontrée pour la reconnaissance de segments dans des documents anciens, manuscrits ou abîmés. Ces propriétés sont les suivantes :

1. existence possible de discontinuités : il est utile de permettre localement une absence de points, due à la qualité ou à la nature des objets extraits (ligne pointillée, défaut de binarisation, bruit) ;
2. prise en compte de l'épaisseur pour chaque point représentatif ;
3. gestion de la taille variable des segments, allant de quelques pixels à plusieurs centaines ;
4. prise en compte de segments qui se croisent ;
5. prise en compte de la courbure dans un segment ;
6. prise en compte du biais.

Ces propriétés utiles sont illustrées dans la partie 5.2.2.

Annexe B

Description grammaticale des courriers manuscrits

Nous présentons dans cet annexes la grammaire complète appliquée pour les courriers manuscrits présentés dans le chapitre 8. Le code présenté est ici complet ; il est directement compilable par la méthode DMOS-P.

B.1 Reconnaissance d'un courrier

La règle globale est la suivante :

```
courrier (courrierReco E Des DFin O Ouvr2 Co2 Sign Ps)::=
  %Remplissage du calque des lignes
  AT_ABS(toutPage) &&
  detecteToutesLignesDeTexte &&
  %Démarrage de la structure
  signat Zbase Sign&&
  ps Zbase Ps &&
  blocDeTexteInv Zbase Ouvr Co &&
  coorExp E1 (expediteurReco Exp1)&&
  dateLieu DPrec@(dateReco ZD1) &&
  coorDes Des DPrec Dfin &&
  objetRef  O (expediteurReco E2) &&
  ‘‘(append Exp1 E2 E).
```

B.2 Signature

Si on arrive à faire réussir sur 2 lignes.

```
signat Z (signatureReco ListeFin) ::=
  AT_ABS(basPage) &&
```

```
  ligneSignature L &&
  AT(auDessusSignature L) &&
  ligneSignature L2 &&
  ‘‘(condPasTropDecalee L2 L) &&
  ‘‘(!,consZoneLignes [L] Z) &&
  ‘‘(consZonesLigneCC [L,L2] ListeFin).
```

Ou sur une ligne :

```
signat Z (signatureReco ListeFin) ::=
  AT_ABS(basPage) &&
  ligneSignature L &&
  ‘‘(!,consZoneLignes [L] Z) &&
  ‘‘(consZonesLigneCC [L] ListeFin).
```

Sinon, réussit sur vide pour ne pas géner le reste de l'analyse.

```
signat zoneRimesVide (signatureReco []) ::= ‘‘(!).
```

B.3 PS et pièce jointe

On ne gère que le PS situé en bas de la page, sous la zone précédemment trouvée `Zbase`.

```
ps Zbase (pspjReco ListeZones) ::=
  AT(sousZoneLarge Zbase) &&
  ligneTexteLongue Ligne &&
  ‘‘(consZonesLigneCC [Ligne] ListeZones) &&
  ‘‘(!).
ps _ (pspjReco []) ::= ‘‘(!).
```

B.4 Bloc de texte et ouverture

Cette règle analyse successivement les lignes de texte du bas vers le haut, en terminant par une ouverture.

```
blocDeTexteInv Zbase (ouvertureReco ZonesOuv) (corpsReco ZonesCorps) ::=
  AT(surZoneLarge Zbase) &&
  ligneTexteBaseBasse L &&
  ‘‘(extremiteDroite L X) &&
  AT(surLigne L) &&
  blocCorps X LesLignesCorps LaLigneOuv &&
  ‘‘(consZonesLigneCC [L|LesLignesCorps] ZonesCorps ,
  consZonesLigneCC [LaLigneOuv] ZonesOuv) &&
  ‘‘(!).
```

Peut réussir sur vide, cela permet de continuer l'analyse.

```
blocDeTexteInv _Zbase (ouvertureReco []) (corpsReco []) ::=
  ‘‘(!).
```

Soit c'est une ligne avec rien au dessus

```
blocCorps  X  [] Ouv ::=
  ligneTexteBasseAssezLongue X Ouv &&
  AT(surLigne Ouv) &&
  rienDuTout &&
  ‘‘(!).
```

Soit c'est une ligne longue et dans ce cas on continue

```
blocCorps X  [L|Autres] Ouv ::=
  ligneTexteBasseAssezLongue X L &&
  ‘‘(!) &&
  ‘‘(majExtremiteDroite L X Y) &&
  AT(surLigne L) &&
  blocCorps Y Autres Ouv.
```

Soit c'est une ligne courte mais avec quelque chose d'intéressant au dessus : on continue

```
blocCorps X [L,R|Autres] Ouv ::=
  ligneTexteBasseCourte X L &&
  AT(surLigne L) &&
  ligneTexteBasseAssezLongue X R &&
  ‘‘(majExtremiteDroite R X Y) &&
  AT(surLigne R) &&
  blocCorps Y Autres Ouv &&
  ‘‘(!).
```

Sinon, deux lignes courtes peuvent réussir si on trouve une lignes longues derrière, c'est signe que le document continue.

```
blocCorps X [C1,C2,L1|Autres] Ouv ::=
  ligneTexteBasseCourte X C1 &&
  AT(surLigne C1) &&
  ligneTexteBasseCourte X C2 &&
  AT(surLigne C2) &&
  ligneTexteBasseAssezLongue X L1 &&
  ‘‘(majExtremiteDroite L1 X Y) &&
  AT(surLigne L1) &&
  blocCorps Y Autres Ouv &&
  ‘‘(!).
```

Fait réussir sur la plus basse des lignes si on n'a plus rien au dessus

```
blocCorps X [] Ouv ::=
  ligneTexteBasseCourte X Ouv &&
  AT(surLigneDroite Ouv) &&
  SELECT(ligneTexteCourteBasseDuHaut X _R) &&
  ‘‘(!).
```

Fait réussir sur la plus basse des lignes si on n'a plus rien au dessus

```
blocCorps X [] Ouv ::=
  ligneTexteBasseCourte X Ouv &&
  AT(surLigneDroite Ouv) &&
  SELECT(ligneTexteCourteBassePasAlignee Ouv X _R)&&
  ‘‘(!).
```

Faire échouer si on a une ligne courte et juste au dessus à droite une ligne courte par risque d'etre dans l'expediteur ou date.

```
blocCorps X [R] _Ouv ::=
  ligneTexteBasseCourte X L &&
  AT(surLigneDroite L) &&
  ligneTexteBasseCourte X R &&
  ‘‘(!,fail).
```

Sinon c'est une ligne courte avec rien de long au dessus et dans ce cas on arrete sur ouv

```
blocCorps X [] Ouv ::=
  ligneTexteBasseCourte X Ouv &&
  ‘‘(!).
```

B.5 Coordonnées expéditeur

5 lignes et la dernière pas trop longue (nom, adresse1, adresse2, téléphone, référence client)

```
coorExp  (expediteurReco ListeZones) ::=
  AT_ABS(hautGauchePage) &&
  ligneTexteBase L1 &&
  AT(sousLigne L1) &&
  ligneTexteBase L2 &&
  AT(sousLigne L2) &&
  ligneTexteBase L3 &&
  AT(sousLigne L3) &&
  ligneTexteBase L4 &&
  %Pour éviter d'aller piocher dans l'objet
  ‘‘(condPasTropLongue L4 L3) &&
  AT(sousLigne L4) &&
```

```
  ligneTexteBase L5 &&
  %Pour éviter d'aller piocher dans l'objet
  ‘‘(condPasTropLongue L5 L4) &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [L1, L2,L3,L4,L5] ListeZones).
```

4 lignes (nom, adresse1, adresse2, téléphone)

```
coorExp  (expediteurReco ListeZones) ::=
  AT_ABS(hautGauchePage) &&
  ligneTexteBase L1 &&
  AT(sousLigne L1) &&
  ligneTexteBase L2 &&
  AT(sousLigne L2) &&
  ligneTexteBase L3 &&
  AT(sousLigne L3) &&
  ligneTexteBase L4 &&
  %Pour éviter d'aller piocher dans l'objet
  ‘‘(condPasTropLongue L4 L3) &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [L1, L2,L3,L4] ListeZones).
```

3 lignes (nom, adresse1, adresse2)

```
coorExp  (expediteurReco ListeZones) ::=
  AT_ABS(hautGauchePage) &&
  ligneTexteBase L1 &&
  AT(sousLigne L1) &&
  ligneTexteBase L2 &&
  AT(sousLigne L2) &&
  ligneTexteBase L3 &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [L1, L2,L3] ListeZones).
```

2 lignes

```
coorExp  (expediteurReco ListeZones) ::=
  AT_ABS(hautGauchePage) &&
  ligneTexteBase L1 &&
  AT(sousLigne L1) &&
  ligneTexteBase L2 &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [L1, L2] ListeZones).
```

1 ligne

```
coorExp  (expediteurReco ListeZones) ::=
  AT_ABS(hautGauchePage) &&
  ligneTexteBase L1 &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [L1] ListeZones).
```

Sinon, pas de coordonnées expediteur à cet endroit

```
coorExp  (expediteurReco []) ::= ‘‘(!).
```

B.6 Date et lieu

On cherche en priorité la date et le lieu sur une ligne en haut à droite de la page.

```
dateLieu (dateReco ListeZones) ::=
  AT_ABS(hautDroitePage) &&
  FIND(ligneDate Ligne) UNTIL(noStopRule) &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [Ligne] ListeZones).
```

Cela peut aussi etre sur deux lignes : lieu et date en dessous

```
dateLieu (dateReco ListeZones) ::=
  AT_ABS(hautDroitePage) &&
  FIND(deuxLignesDate L1 L2 ) UNTIL(noStopRule) &&
  ‘‘(consZonesLigneCC [L1,L2] ListeZones).

dateLieu (dateReco[]) ::= ‘‘(!).
```

Une ligne de date est isolée : rien au dessus ni en dessous.

```
ligneDate L ::=
  ligneTexteBase L &&
  ‘‘(condAssezADroite L) &&
  AT(sousLigneProche L) &&
  rienDuTout &&
  AT(surLigneProche L) &&
  rienDuTout.
```

Il en est de même pour une date sur deux lignes.

```
deuxLignesDate L1 L2 ::=
  ligneTexteBase L1 &&
  ‘‘(condAssezADroite L1) &&
  AT(surLigneProche L1) &&
  rienDuTout &&
  AT(sousLigneProche L1) &&
```

```
  ligneTexteBase L2 &&
  AT(sousLigneProche L2) &&
  rienDuTout &&
  ‘‘(condAlignees L1 L2) &&
  ‘‘(!).
```

B.7 Coordonnées destinataire

Recherche les coordonnées destinataire et éventuellement une date s'il n'en a pas été trouvé avant.

On reconnait la date parce qu'elle est séparée d'un interligne plus grand que les autres champs.

On n'a pas trouvé la date, elle est en premier

```
destDate (destinataireReco ListeZones) (dateReco []) (dateReco ZD)::=
  AT_ABS(hautDroitePage) &&
  coorDes Lignes &&
  ‘‘(Lignes = [L1,L2,L3|R]) &&
  ‘‘(condEspacePlusGrand L2 L3 L1 L2) &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [L1] ZD) &&
  ‘‘(consZonesLigneCC [L2,L3|R] ListeZones).
```

On n'a pas trouvé la date et elle est en dernier

```
destDate (destinataireReco ListeZones) (dateReco []) (dateReco ZD)::=
  AT_ABS(hautDroitePage) &&
  coorDes Lignes &&
  ‘‘(append R [L1,L2,L3] Lignes) &&
  ‘‘(condEspacePlusGrand L1 L2 L2 L3) &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [L3] ZD) &&
  ‘‘(consZonesLigneCC [L1,L2|R] ListeZones).
```

On ne recherche plus la date mais un destinataire simple

```
destDate (destinataireReco ListeZones) _DateOld (dateReco []) ::=
  AT_ABS(hautDroitePage) &&
  coorDes Lignes &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC Lignes ListeZones).
```

Peut aussi réussir aussi au milieu de la page

```
destDate (destinataireReco ListeZones) _DateOld (dateReco []) ::=
  AT_ABS(hautMilieuPage) &&
```

```
  coorDes Lignes &&
  ‘‘(condPasTropLongDestRec Lignes) &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC Lignes ListeZones).
```

Sinon pas de destinataire

```
destDate (destinataireReco []) _DateOld (dateReco []) ::= ‘‘(!).
```

Version 5 lignes

```
coorDes [L1,L2,L3,L4,L5] ::=
  ligneTexteDroite L1 &&
  AT(sousLigneProche L1) &&
  ligneTexteDroite L2 &&
  AT(sousLigne L2) &&
  ligneTexteDroite L3 &&
  AT(sousLigne L3) &&
  ligneTexteDroite L4 &&
  AT(sousLigne L4) &&
  ligneTexteDroite L5 &&
  ‘‘(!).
```

Version 4 lignes

```
coorDes [L1,L2,L3,L4] ::=
  ligneTexteDroite L1 &&
  AT(sousLigneProche L1) &&
  ligneTexteDroite L2 &&
  AT(sousLigne L2) &&
  ligneTexteDroite L3 &&
  AT(sousLigne L3) &&
  ligneTexteDroite L4 &&
  ‘‘(!).
```

Version 3 lignes

```
coorDes [L1,L2,L3] ::=
  ligneTexteDroite L1 &&
  AT(sousLigneProche L1) &&
  ligneTexteDroite L2 &&
  AT(sousLigne L2) &&
  ligneTexteDroite L3 &&
  ‘‘(!).
```

Version 2 lignes

```
coorDes [L1,L2] ::=
  ligneTexteDroite L1 &&
  AT(sousLigneProche L1) &&
  ligneTexteDroite L2 &&
  ‘‘(!).
```

Version 1 ligne

```
coorDes [L1] ::=
  ligneTexteDroite L1 &&
  ‘‘(!).
```

B.8 Objet

Version 3 lignes :

```
objetRef (objReco Lzo) (expediteurReco Lze) ::=
  AT_ABS(gauchePage) &&
  %Obj sur 2 lignes
  ligneTexteBase L1 &&
  AT(sousLigne L1) &&
  ligneTexteBase L2 &&
  ‘‘(condAvecAlinea L1 L2) &&
  AT(sousLigne L2)&&
  %RefClient
  ligneTexteBase L3 &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [L1,L2] Lzo, consZonesLigneCC [L3] Lze).

objetRef  (objReco Lzo) (expediteurReco Lze) ::=
  AT_ABS(gauchePage) &&
  %RefClient
  ligneTexteBase L1 &&
  AT(sousLigne L1) &&
  %Obj sur 2 lignes
  ligneTexteBase L2 &&
  AT(sousLigne L2)&&
  ligneTexteBase L3 &&
  ‘‘(condAvecAlinea L2 L3) &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [L2,L3] Lzo, consZonesLigneCC [L1] Lze).
```

Version 2 lignes :

```
objetRef  (objReco Lzo) (expediteurReco []) ::=
```

```
  AT_ABS(gauchePage) &&
  %Obj sur 2 lignes
  ligneTexteBase L1 &&
  AT(sousLigne L1)&&
  %On trouve un alinéa car l'objet est sur deux lignes
  ligneTexteBase L2 &&
  ``(condAvecAlinea L1 L2) &&
  ``(!) &&
  ``(consZonesLigneCC [L1,L2] Lzo).

objetRef  (objReco Lzo) (expediteurReco Lze) ::=
  AT_ABS(gauchePage) &&
  %Obj sur 1 lignes
  ligneTexteBase L1 &&
  AT(sousLigne L1) &&
  ligneTexteBase L2 &&
  ``(lignePlusLongue L1 L2 L1) &&
  %la plus longue est l'objet
  ``(!) &&
  ``(consZonesLigneCC [L1] Lzo, consZonesLigneCC [L2] Lze).


objetRef  (objReco Lzo) (expediteurReco Lze) ::=
  AT_ABS(gauchePage) &&
  %RefClient
  ligneTexteBase L1 &&
  AT(sousLigne L1) &&
  %Obj sur 1 lignes
  ligneTexteBase L2 &&
  ``(!) &&
  ``(consZonesLigneCC [L2] Lzo, consZonesLigneCC [L1] Lze).
```

Version sur une ligne :

```
%Ou il n'y a qu'un objet
objetRef  (objReco Lz) (expediteurReco []):=
  AT_ABS(gauchePage) &&
  ligneTexteLongue Ligne &&
  ``(!) &&
  ``(consZonesLigneCC [Ligne] Lz).

%Ou il n'y a qu'une ref
objetRef  (objReco []) (expediteurReco Lz) ::=
  AT(gauchePage) &&
```

```
  ligneTexteBase Ligne &&
  ‘‘(!) &&
  ‘‘(consZonesLigneCC [Ligne] Lz).
```

Sinon, il n'y a rien.

```
objetRef (objReco []) (expediteurReco []) ::=  ‘‘(!).
```

B.9 Description des lignes utiles

L'analyse se déroule dans le calque `LignesDeTexte`. La reconnaissance des lignes de texte est donc réalisée par un simple appel aux reconnaisseurs de terminaux.

```
ligneTexteBase L ::=
  TERM_SEG noCondS noCondS ligne L.
ligneTexteBaseBasse L ::=
  TERM_SEG (condGlobaleS condSegPlusBas) noCondS ligne L.
ligneTexteAssezLongue X L ::=
  TERM_SEG (condAssezLongueLigne X) noCondS ligne L.
ligneTexteBasseAssezLongue X L ::=
  TERM_SEG (condGlobaleS condSegPlusBas) (condAssezLongueLigne X) ligne L.
ligneTexteLongue L ::=
  TERM_SEG (condLongueLigne) noCondS ligne L.
ligneTexteCourte X L ::=
  TERM_SEG (condPasLigneLongue X)  noCondS ligne L.
ligneTexteBasseCourte X L ::=
  TERM_SEG (condGlobaleS condSegPlusBas) (condPasLigneLongue X)  ligne L.
ligneTexteDroite L ::=
  TERM_SEG (condAssezADroite) noCondS ligne L.
ligneTexteCourteBasseDuHaut X L ::=
  ligneTexteBasseCourte X L &&
  AT(surLigneProche L) &&
  rienDuTout.
```

Ne rien reconnaître, c'est échouer si on trouve une ligne et réussir sinon :

```
rienDuTout ::=
  TERM_SEG noCondS noCondS ligne _L2 &&
 ‘‘(!,fail).
rienDuTout ::=
‘‘(!).
```

Bibliographie

[AD07] A. Antonacopoulos and A. C. Downton. Special issue on the analysis of historical documents. *International Journal on Document Analysis and Recognition*, 9(2-4) :75–77, April 2007.

[AGF04] Y. Amit, D. Geman, and X. Fan. A coarse-to-fine strategy for multi-class shape detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.

[Ara05] D. W. Arathorn. Memory-driven visual attention : An emergent behavior of map-seeking circuits. *Neurobiology of Attention*, pages 605–609, 2005.

[Bar05] M. Bar. Top-down facilitation of visual object recognition. *Neurobiology of Attention*, pages 140–145, 2005.

[BCLM98] P. Bottoni, L. Cinque, S. Levialdi, and P. Mussio. Matching the resolution level to salient image features. *Pattern Recognition*, 31(1) :89–104, January 1998.

[Ber95] P. Bertolino. *Contribution des pyramides irréguliers en segmentation d'images multiresolution.* Thèse de doctorat, Institut National Polytechnique de Grenoble, France, Novembre 1995.

[Blo91] D.S. Bloomberg. Multiresolution morphological approach to document image analysis. In *ICDAR 1991*, pages 963–971, 1991.

[BPS07] U. Bhattacharya, S. K. Parui, and B. Shaw. A hybrid scheme for recognition of Handwritten Bangla basic characters based on HMM and MLP classifiers. In *Proc. International Conference on Advances in Pattern Recognition*, pages 101–106. World Scientific, 2007.

[BR77] R. Bajcsy and D. A. Rosenthal. What one can see on the earth from different altitudes : A hierarchical control structure in computer vision. In *Pattern Recognition and Image Processing*, pages 108–111, 1977.

[BR80] R. Bajcsy and D. A. Rosenthal. *Visual and Conceptual Focus of Attention*, pages 133–149. Academic Press, 1980.

[Bri92] P. Brisset. *Compilation de λProlog.* PhD thesis, Université de Rennes I, 1992.

[BSME97] A. Baumgartner, C. T. Steger, H. Mayer, and W. Eckstein. Multiresolution, semantic objects, and context for road extraction. In *Semantic*

Modeling for the Acquisition of Topographic Information from Images and Maps, pages 140–156, 1997.

[Bur88] P. J. Burt. Smart sensing with a pyramid vision machine. *Proceedings of the IEEE*, 76 :1006–1015, 1988.

[CB98] F-I. Cheng and C. Bouman. Trainable context model for multiscale segmentation. In *Proc. IEEE International Conf. Image Processing*, volume 1, pages 610–614, 1998.

[CB01] H. Cheng and C. Bouman. Multiscale bayesian segmentation using a trainable context model. *IEEE Transactions on Image Processing*, 10(4) :511–525, April 2001.

[CBA97] H. Cheng, C. Bouman, and J. Allebach. Multiscale document segmentation. In *IS&T 50th Annual Conference*, pages pp. 417–425, May 1997.

[CCL04] B. Coüasnon, J. Camillerapp, and I. Leplumey. Making handwritten archives documents accessible to public with a generic system of document image analysis. In *International Conference on Document Image Analysis for Libraries (DIAL)*, pages 270–277, 2004.

[CCLM97] V. Cantoni, L. Cinque, L. Lombardi, and G. Manzini. Page segmentation using a pyramidal architecture. In *Workshop on Computer Architectures for Machine Perception*, page Session 6, 1997.

[CeS07] A. Costa e Silva. New metrics for evaluating performance in document analysis tasks - application to the table case. In *9th International Conference on Document Analysis and Recognition (ICDAR 2007)*, pages 481–485, 2007.

[CFL+99] L. Cinque, L. Forino, S. Levialdi, L. Lombardi, and S. L. Tanimoto. Understanding the page logical structure. In *1oth International Conference on Image Analysis and Processing (ICIAP 1999)*, pages 1003–1008, 1999.

[Chu05] M. M. Chun. Contextual guidance of visual attention. *Neurobiology of Attention*, 1 :246–250, 2005.

[CLM98] L. Cinque, L. Lombardi, and G. Manzini. A multiresolution approach for page segmentation. *Pattern Recognition Letters*, 19(2) :217–225, 1998.

[Coh00] J. Cohen. L'écran efficace : trois lois fondamentales de la perception visuelle. *Documentaliste - Sciences de l'information*, 37(3-4) :192–198, septembre 2000.

[Coü01] B. Coüasnon. DMOS : A generic document recognition method to application to an automatic generator of musical scores, mathematical formulae and table structures recognition systems. In *Proceedings of International Conference on Document Analysis and Recognition (ICDAR'01)*, pages 215–220, 2001.

[Coü06] B. Coüasnon. DMOS, a generic document recognition method : Application to table structure analysis in a general and in a specific way. *International Journal on Document Analysis and Recognition, IJDAR*, 8(2) :111–122, 2006.

[CT90] S. Cosby and R. Thomas. Irs : a hierarchical knowledge based system for aerial image interpretation. In *IEA/AIE '90 : Proceedings of the 3rd international conference on Industrial and engineering applications of artificial intelligence and expert systems*, pages 207–215, New York, NY, USA, 1990. ACM Press.

[DB94] O. Déforges and D. Barba. A fast multiresolution text-line and non text-line structures extraction. In *International Conference on Image Processing (ICIP)*, pages 134–138, 1994.

[DPVGB95] O. Déforges, P. Piquin, C. Viard-Gaudin, and D. Barba. Segmentation d'images de documents par une approche multirésolution. extraction précise de lignes de texte. *Traitement du signal*, 12-6 :527–539, 1995.

[Dye87] C. R. Dyer. *Multiscale image understanding*, pages 171–213. Academic Press Professional, Inc., San Diego, CA, USA, 1987.

[DZ01] G. Deco and J. Zihl. A neurodynamical model of visual attention : Feedback enhancement of spatial resolution in a hierarchical system. *Journal of Computational Neuroscience*, 10(3) :231–253, 2001.

[EBE99] V. Eglin, S. Bres, and H. Emptoz. Structuration de documents par repérage de zones d'intérêt. *Traitement du Signal*, 16-3 :219, 1999.

[EDC97] K. Etemad, D. Doermann, and R. Chellappa. Multiscale segmentation of unstructured document pages using soft decision integration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(1) :92–96, 1997.

[Egl06] V. Eglin. Approches perceptives et cognitives en analyse automatique d'images de documents. *Revue TSI technique et Sciences Informatiques, numéro spécial "Document Numérique"*, 25/4 :523–551, December 2006.

[FT01] M. Feldbach and K. D. Tonnies. Line detection and segmentation in historical church registers. In *International Conference on Document Analysis and Recognition*, pages 743–747, 2001.

[Gar95] M. D. Garris. Evaluating spatial correspondence of zones in document recognition systems. In *International Conference on Image Processing*, pages 304–307, 1995.

[GGC+06] E. Grosicki, E. Geoffrois, M. Carré, E. Augustin, and F. Prêteux. La campagne d'évaluation rimes pour la reconnaissance de courriers manuscrits. In *Actes du neuvième colloque international francophone sur l'écrit et le document*, 2006.

[GMA01] B. Gatos, S.L. Mantzaris, and A. Antonacopoulos. First international newspaper segmentation contest. In *International Conference on Document Analysis and Recognition (ICDAR'01)*, volume 00, page 1190, Los Alamitos, CA, USA, 2001. IEEE Computer Society.

[GMC+99] B. Gatos, S. L. Mantzaris, K. V. Chandrinos, A. Tsigris, and S. J. Perantonis. Integrated algorithms for newspaper page decomposition and article tracking. In *ICDAR '99 : Proceedings of the Fifth International Conference*

on Document Analysis and Recognition, page 559, Washington, DC, USA, 1999. IEEE Computer Society.

[HD95] Osamu Hori and David S. Doermann. Robust table-form structure analysis based on box-driven reasoning. In *ICDAR*, pages 218–221, 1995.

[HHI01] K. Hadjar, O. Hitz, and R. Ingold. Newspaper page decomposition using a split and merge approach. In *International Conference on Document Analysis and Recognition (ICDAR'01)*, volume 00, page 1186, Los Alamitos, CA, USA, 2001. IEEE Computer Society.

[HJBJ+96] A. Hoover, G. Jean-Baptiste, X. Y. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgof, K. W. Bowyer, D. W. Eggert, A. W. Fitzgibbon, and R. B. Fisher. An experimental comparison of range image segmentation algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(7) :673–689, July 1996.

[IK01] L. Itti and C. Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3) :194–203, Mar 2001.

[IRT05] L. Itti, G. Rees, and J. K. Tsotsos. *Neurobiology of Attention*. Academic Press, December 2005.

[JR94] J.-M. Jolion and A. Rosenfeld. *A Pyramid Framework for Early Vision : Multiresolutional Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA, 1994.

[KIDM98] K. Kise, M. Iwata, A. Dengel, and K. Matsumoto. Text-line extraction as selection of paths in the neighbor graph. In *Document Analysis Systems*, pages 225–239, 1998.

[Kof35] K. Koffka. *Principles of Gestalt Psychology*. Book, 1935.

[LCQ95] I. Leplumey, J. Camillerapp, and C. Queguiner. Kalman filter contributions towards document segmentation. In *Proceedings of International Conference on Document Analysis and Recognition (ICDAR'95)*, pages 765–769, 1995.

[LGGP08] M. Lemaitre, E. Grosicki, E. Geoffrois, and F. Prêteux. Layout analysis of handwritten letters based on textural and spatial information and a 2d markovian approach. In *International Conference on Frontiers in Handwriting Recognition (ICFHR 2008)*, 2008.

[LGPH08] G. Louloudis, B. Gatos, I. Pratikakis, and C. Halatsis. Line and word segmentation of handwritten documents. In *International Conference on Frontiers in Handwriting Recognition (ICFHR 2008)*, 2008.

[LLHY01] F. Liu, Y. Luo, D. Hu, and M. Yoshikawa. A new component based algorithm for newspaper layout analysis. In *International Conference on Document Analysis and Recognition (ICDAR'01)*, volume 00, page 1176, Los Alamitos, CA, USA, 2001. IEEE Computer Society.

[LR01] S. Lee and B. Ryu. Parameter-free geometric document layout analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11) :pp. 1240–1256, nov. 2001.

[LSF94] L. Likforman-Sulem and C. Faure. Extracting text lines in handwritten documents by perceptual grouping. In *Advances in handwriting and drawing : a multidisciplinary approach*, pages 117–135. C. Faure, P. Keuss, G. Lorette, A. Winter, Europia, Paris, 1994.

[LSF95] L. Likforman-Sulem and C. Faure. Une méthode de résolution des conflits d'alignements pour la segmentation de documents manuscrits. *Traitement du signal*, 12-6 :541–549, 1995.

[LSZT07] L. Likforman-Sulem, A. Zahour, and B. Taconet. Text line segmentation of historical documents : a survey. *International Journal on Document Analysis and Recognition*, 9(2-4) :123–138, April 2007.

[LT01] P. K. Loo and C. L. Tan. Detection of word groups based on irregular pyramid. In *6th International Conference on Document Analysis and Recognition (ICDAR 2001)*, pages 200–204, 2001.

[LT02] P. K. Loo and C. L. Tan. Word and sentence extraction using irregular pyramid. In *DAS '02 : Proceedings of the 5th International Workshop on Document Analysis Systems V*, pages 307–318, London, UK, 2002. Springer-Verlag.

[LT03] P. K. Loo and C. L. Tan. Using irregular pyramid for text segmentation and binarization of gray scale images. In *International Conference on Document Analysis and Recognition*, pages 594–598, 2003.

[LZD06] Y. Li, Y. Zheng, and D. S. Doermann. Detecting text lines in handwritten documents. In *International Conference on Pattern Recognition*, volume 2, pages 1030–1033. IEEE Computer Society, 2006.

[ML95] M. Mkaouar and R. Lepage. Extraction of characteristics from an image by analysis with multiple spatial resolutions. In *Canadian Conference on Electrical and Computer Engineering*, volume 2, pages 1176–1179, 1995.

[MLBS97] H. Mayer, I. Laptev, A. Baumgartner, and C. T. Steger. Automatic road extraction based on multi-scale modeling, context, and snakes. *International Archives of Photogrammetry and Remote Sensing*, 13 :106–113, 1997.

[MRK03] S. Mao, A. Rosenfeld, and T. Kanungo. Document structure analysis algorithms : a literature survey. In *Document Recognition and Retreval X, (Proceedings of SPIE/IST)*, volume 5010, Santa Clara, California, January 2003.

[MSB+08] Samia Snoussi Maddouri, Fadoua Bouafif Samoud, Kaouthar Bouriel, Noureddine Ellouze, and Haikal El Abed. Baseline extraction : Comparison of six methods on ifn/enit database. In *International Conference on Frontiers in Handwriting Recognition*, 2008.

[Nei76] U. Neisser. *Cognition and Reality : principles and implicatios of cognitive psychology*. W. H. Freeman and Company, 1976.

[Not70] D. Noton. A theory of visual pattern perception. *Systems Science and Cybernetics, IEEE Transactions on*, 6 :349–357, Oct. 1970.

[NPH04] S. Nicolas, T. Paquet, and L. Heutte. Text line segmentation in handwritten document using a production system. In F. Kimura and H.Fujisawa, editors, *9th International Workshop on Frontiers in Handwriting Recognition (IWFHR)*, pages 245–250. IAPR, 2004.

[OMLL95] J.-M. Ogier, R. Mullot, J. Labiche, and Y. Lecourtier. Interprétation de documents par cylces perceptifs de construction d'objets cohérents. application aux données cadastrales. *Traitement du signal*, 12-6 :627–637, 1995.

[PAhM+97] C. Pieplu, A. Al-hamdi, R. Mullot, J.-M. Ogier, and P. Dumas. Système d'interprétation de documents techniques. In *Seizième colloque GRETSI*, 1997.

[PC04] U. Pal and B. B. Chaudhuri. Indian script character recognition : a survey. *Pattern Recognition*, 37(9) :1887–1899, 2004.

[PK89] G. Paar and W. G. Kropatsch. *Hierarchical cooperation between numerical and symbolic image representations*, pages 113–130. World Scientific, Singapore ; New Jersey, 1989.

[RB06] Y. Rangoni and A. Belaïd. Document logical structure analysis based on perceptive cycles. In *Document Analysis Systems*, pages 117–128, 2006.

[RGG+05] I. A. Rybak, V. I. Gusakova, A. V. Golovan, L. N. Padladchikova, and N. A. Shevtsova. Attention-guided recognition based on "what" and "where" representations : A behavioral model. *Neurobiology of Attention*, 1 :663–670, 2005.

[Ros84] D.A. Rosenthal. Visual and conceptual hierarchy - a paradigm for studies of automated generation of recognition strategies. In A. Rosenfeld, editor, *Multiresolution Image Processing and Analysis*, pages 60–76. Springer–Verlag, 1984.

[RPL08] P. P. Roy, U. Pal, and J. Lladós. Morphology based handwritten line segmentation using foreground and background information. In *International Conference on Frontiers in Handwriting Recognition (ICFHR 2008)*, 2008.

[RVB+05] K. Roy, Szilárd Vajda, Abdel Belaïd, U. Pal, and Bidyut Baran Chaudhuri. A system for indian postal automation. In *Proc. International Conference on Document Analysis and Recognition (ICDAR)*, pages 1060–1064. IEEE Computer Society, 2005.

[RVE98] J.-Y. Ramel, N. Vincent, and H. Emptoz. Interprétation de documents techniques par "cycles perceptifs" à partir d'une perception globale du document. *Traitement du signal*, 15-2 :83–102, 1998.

[SAA01] A. A. Salah, E. Alpaydin, and L. Akarun. A selective attention based method for visual pattern recognition. In J.D.Moore and K.Stenning, editors, *Proc. 23rd Annual Conference of the Cognitive Science Society*, pages pp.881–886, 2001.

[SB93] S. Sarkar and K. L. Boyer. Perceptual organization in computer vision : a review and a proposal for a classificatory structure. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(2) :382–399, 1993.

[SG05] Z. Shi and V. Govindaraju. Multi-scale techniques for document page segmentation. In *ICDAR '05 : Proceedings of the Eighth International Conference on Document Analysis and Recognition*, pages 1020–1024, Washington, DC, USA, 2005. IEEE Computer Society.

[SGS93] V. Shapiro, G. Gluhchev, and V. S. Sgurev. Handwritten document image segmentation and analysis. *Pattern Recognition Letters*, 14(1) :71–78, 1993.

[Sil88] T. M. Silberberg. Multiresolution aerial image interpretation. In *Image Understanding Workshop*, pages 505–511, 1988.

[Sor85] H. W. Sorenson. *Kalman Filtering : Theory and application.* IEEE Press, New York, 1985.

[SVL+04] G. Sanchez, E. Valveny, J. Lladós, J. Mas Romeu, and N. Lozano. A platform to extract knowledge from graphic documents. application to an architectural sketch understanding scenario. In *Workshop on Document Analysis Systems*, pages 389–400, 2004.

[Tre92] A. Treisman. L'attention, les traits et la perception des objets. *Introduction aux sciences cognitives*, 1992.

[VGB91] C. Viard-Gaudin and D. Barba. A multi-resolution approach to extract the address block on flat mail pieces. In *Proceedings of ICASSP91*, pages 2701–2704, 1991.

[WF03] M. Wattenberg and D. Fisher. A model of multi-scale perceptual organization in information graphics. In *Proceedings of IEEE Symposium on Information Visualization*, October 2003.

[Wol05] J. M. Wolfe. Guidance of visual search by preattentive information. *Neurobiolgy of Attention*, 1 :101–104, 2005.

[XL05] D. Xi and S. W. Lee. Extraction of reference lines and items from form document images with complicated background. *Pattern Recognition*, 38(2) :289–305, February 2005.

[YV98] B. A. Yanikoglu and L. Vincent. Pink panther : A complete environment for ground-truthing and benchmarking document page segmentation. *Pattern Recognition*, 31(9) :1191–1204, 1998.

Publications de l'auteur

[1] Aurélie Lemaitre and Jean Camillerapp. Text line extraction in handwritten document with kalman filter applied on low resolution image. In *Document Image Analysis for Libraries (DIAL'06)*, pages 38–45, 2006.

[2] Aurélie Lemaitre, Jean Camillerapp, and Bertrand Coüasnon. Multiresolution cooperation improves document structure recognition. *International Journal on Document Analysis and Recognition (IJDAR)*, 11(2) :97–109, November 2008.

[3] Aurélie Lemaitre, Jean Camillerapp, and Bertrand Coüasnon. Contribution of multiresolution description for archive document structure recognition. In *Proceedings of International Conference on Document Analysis and Recognition (ICDAR'07)*, pages 247–251, 2007.

[4] Aurélie Lemaitre, Jean Camillerapp, and Bertrand Coüasnon. Approche perceptive pour la reconnaissance de filets bruités - application à la structuration de pages de journaux. In *Conférence Internationale sur l'Ecrit et le Document (CIFED'08)*, 2008.

[5] Aurélie Lemaitre, Jean Camillerapp, and Bertrand Coüasnon. A generic method for structure recognition of handwritten mail documents. In *Document Recognition and Retrieval (DRR XV)*, 2008.

[6] Aurélie Lemaitre, Biduyt Baran Chaudhuri, and Bertrand Coüasnon. Perceptive vision for headline localisation in bangla handwritten text recognition. In *Proceedings of International Conference on Document Analysis and Recognition (ICDAR'07)*, pages 614–618, 2007.

[7] Aurélie Lemaitre, Bertrand Coüasnon, and Ivan Leplumey. Using a neighbourhood graph based on Voronoï tessellation with DMOS, a generic method for structured document recognition. In *Proceedings of GREC, Sixth IAPR International Workshop on Graphics Recognition*, pages 260–271, Hong Kong, China, August 2005.

[8] Aurélie Lemaitre, Bertrand Coüasnon, and Ivan Leplumey. Using a neighbourhood graph based on voronoï tessellation with dmos, a generic method for structured document recognition. In *Graphics Recognition, Sixth IAPR International Workshop GREC 2005, Revised Selected Papers*, volume LNCS 3926, pages 267–278, 2006.

Table des figures

Résumé

La vision perceptive humaine combine différents niveaux de perception pour faciliter l'interprétation d'une scène. Les physiologistes la modélisent par le cycle perceptif, guidé par un facteur psychologique, l'attention visuelle.

Ce fonctionnement est à la base de nos travaux sur une méthode générique pour l'analyse de documents structurés. Dans ce contexte, nous proposons le formalisme de calque perceptif ainsi que des outils de multirésolution, pour simuler le cycle perceptif et l'attention visuelle. Le formalisme du calque perceptif permet de fusionner des informations issues de différents niveaux de perception, en étant guidé par des connaissances. Nous aboutissons ainsi à une architecture complète de vision perceptive, DMOS-P, qui est un enrichissement de la méthode DMOS de reconnaissance de documents. Grâce à cette méthode, il devient possible de spécifier simplement des mécanismes complexes de coopération perceptive, adaptés à chaque type de problème, qui améliorent la reconnaissance de la structure de documents.

Nous mettons en évidence un mécanisme de prédiction/vérification lié à la vision perceptive : la vision à basse résolution permet d'émettre des hypothèses sur la structure en utilisant le contexte global ; ces hypothèses sont ensuite vérifiées à plus haute résolution. Ce mécanisme simplifie et améliore la reconnaissance des documents : lorsque les indices visuels sont denses (documents bruités ou à structure complexe), la vision perceptive permet de mieux sélectionner les données structurelles pertinentes ; lorsque l'information structurelle est physiquement diffuse (documents ayant une structure pauvre), la vision perceptive permet de mieux reconstituer la structure du document. Nous avons validé cette approche sur des documents à structure variée (courriers manuscrits, registres d'archives, presse...), à grande échelle (plus de 80 000 images), et de manière industrielle grâce au transfert technologique vers la société Evodia.

Abstract

The human perceptive vision combines several points of view in order to improve the interpretation of a scene. It is modeled by a physiologic component, the perceptive cycle, guided by a psychological aspect, the visual attention.

This mechanism is the base of our work on a generic method for document structure recognition. In this context, we propose the formalism of perceptive layer and some multiresolution tools to simulate the perceptive vision and the visual attention. This produces the perceptive method DMOS-P, which is an improvement of the existing DMOS method. Thanks to this method, it becomes possible to easily specify some complex mechanisms of perceptive cooperation, adapted to each kind of document, and that improve the recognition of the structure.

We point out a mechanism of prediction/verification, linked to the perceptive vision : at low resolution, hypotheses on the contents are proposed, that are verified at a higher resolution. This mechanism simplifies and improves document recognition : for noisy documents, the perceptive vision makes it possible to select only relevant information, whereas for low structured documents, the perceptive vision helps to rebuild the structure. We validated this approach on various kinds of structured documents (incoming mail, archive registers, newspapers...), at a large scale (more than 80,000 images) and thanks to an industrial transfer to Evodia company.