HAL open science

# LearnPos : un nouvel outil pour l'apprentissage interactif de positionnement

Cérès Carton, Aurélie Lemaitre, Bertrand B. Coüasnon

**To cite this version:**

Cérès Carton, Aurélie Lemaitre, Bertrand B. Coüasnon. LearnPos : un nouvel outil pour l'apprentissage interactif de positionnement. Conférence Internationale Francophone sur l'Ecrit et le Document, 2014, Nancy, France. pp.325-340. hal-01088810

**HAL Id: hal-01088810**

**https://inria.hal.science/hal-01088810v1**

Submitted on 28 Nov 2014

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

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

# LearnPos : un nouvel outil pour l'apprentissage interactif de positionnement

**Cérès$_1$ Carton$_1$* — Aurélie$_2$ Lemaitre$_2$** — Bertrand Coüasnon$_3$***

* *INSA - Irisa - UEB, Campus de Beaulieu, 35043 Rennes, France*
** *Université Rennes 2 - Irisa - UEB, Campus de Beaulieu, 35043 Rennes, France*

RÉSUMÉ. *L'analyse de documents structurés nécessite la localisation de données à l'intérieur du document lors du processus de reconnaissance. Dans cet article, nous présentons LearnPos, un nouvel outil générique, indépendant de tout système de reconnaissance de document. LearnPos modélise et évalue le positionnement à partir d'un échantillon d'apprentissage de documents. L'utilisateur est ainsi assisté dans la définition de la structure physique du document. LearnPos peut fournir des positionnements absolus et relatifs, en interagissant avec l'utilisateur. Notre méthode permet de traiter des relations spatiales composées de plusieurs zones distinctes et est capable de fournir un ordre de parcours des différentes zones ainsi qu'un sens de lecture pour chacune des zones afin de minimiser les erreurs. Nous montrons que les opérateurs de position obtenus peuvent être utilisés pour la reconnaissance de documents structurés, tout en réduisant l'exploration manuelle du corpus de documents.*

ABSTRACT. *The analysis of 2D structured documents often requires localizing data inside of a document during the recognition process. In this paper we present LearnPos a new generic tool, independent of any document recognition system. LearnPos models and evaluates positioning from a learning set of documents. Thanks to LearnPos, the user is helped to define the physical structure of the document. LearnPos is able to furnish spatial information for both absolute and relative spatial relations, in interaction with the user. Our method can handle spatial relations compose of distinct zones and is able to furnish appropriate order and point of view to minimize errors. We prove that resulting models can be successfully used for structured document recognition, while reducing the manual exploration of the data set of documents.*

MOTS-CLÉS : *opérateur de position$_1$, reconnaissance de la structure de document$_2$, interaction utilisateur$_3$, apprentissage de positionnement$_4$.*

KEYWORDS: *position operator$_1$, document structure recognition$_2$, user interaction$_3$, position learning$_4$.*

## 1. Introduction

L'analyse de documents structurés consiste à définir la structure logique et la structure physique du document. Pour l'analyse de documents structurés, il est utile de modéliser et évaluer la position des différents composants entre eux et au sein de la page. Le positionnement de ces composants est utilisé durant l'analyse de la structure pour définir l'orientation et l'ordre d'analyse. La qualité des informations spatiales affecte donc les performances de l'analyse.

Les zones 2D qui sont nécessaires pour guider l'analyse sont difficiles à définir. Actuellement, elles sont souvent définies manuellement ce qui présentent des inconvénients. Un opérateur humain doit observer les images de documents pour définir une zone appropriée d'après ces exemples. De plus, les cas rares ne sont pas observés et l'analyse des erreurs doit alors être faite afin d'ajuster les zones définies. L'ajustement des paramètres est chronophage et n'est pas une tâche facile.

Dans cet article, nous présentons LearnPos, un nouvel outil pour la détermination automatique des opérateurs de position. En utilisant LearnPos, l'utilisateur est assisté dans la définition de la structure physique du document sans avoir besoin d'explorer manuellement le corpus de document. LearnPos procède par l'analyse d'un échantillon d'apprentissage annoté. LearnPos permet l'exploration exhaustive et interactive d'un corpus de documents. Les cas rares sont alors facilement détectés, contrairement à ce qui est actuellement possible avec l'analyse manuelle.

Dans la section 2, nous présentons des travaux existants sur la modélisation des relations spatiales et sur les limites qui ont motivé nos travaux. Dans la section 3, nous présentons comment LearnPos procède pour définir les opérateurs de position. Dans la section 4, nous présentons un nouvel indice, l'indice de confusion, qui permet de minimiser les erreurs en ordonnant les différentes zones obtenues pour un opérateur de position et en choisissant le meilleur point de vue pour chaque zone. Dans la section 5, nous évaluons LearnPos sur le corpus public RIMES, corpus de lettres manuscrites en français. L'évaluation sur 1250 lettres manuscrites montre que les opérateurs de position définis automatiquement obtiennent des performances comparables à celles obtenues avec des opérateurs de position définis manuellement, tout en diminuant significativement le nombre de paramètres définis manuellement. Le temps passé à cette tâche est également fortement réduit grâce à l'analyse automatique du corpus.

## 2. État de l'art

En analyse de la structure de document, nous pouvons considérer deux approches principales : les méthodes statistiques et les méthodes syntaxiques. Les méthodes statistiques permettent l'apprentissage des caractéristiques de chaque type d'éléments mais ne permettent pas de transmettre la structure hiérarchique d'un document. Les méthodes syntaxiques segmentent l'image en primitives et construisent un arbre de règles qui décrit comment assembler ces primitives. La définition des opérateurs de position fait partie de la définition de la structure physique des documents. Par

exemple, les méthodes syntaxiques nécessitent la définition d'opérateurs de position pour les règles définies. En général, les paramètres des opérateurs de position sont définis manuellement. Cette définition manuelle est fastidieuse et non exhaustive.

L'expression des relations entre les objets est un problème rencontré dans de nombreux domaines comme la reconnaissance de l'écriture manuscrite, la reconnaissance de la structure de document et plus généralement en analyse d'image. Les performances d'une méthode dépendent de la manière de modéliser les objets et de la précision utilisée pour le positionnement.

Différentes manières de représenter les objets ont été utilisées. En analyse de documents, les objets sont souvent réduits à un ensemble de points représentatifs (méthodes des centroïdes, boîtes englobantes, etc.). Comme l'ont démontré (Bloch et Ralescu, 2003), cette représentation des objets n'est pas adaptée à toutes les problématiques et notamment dans le cas de la reconnaissance de l'écriture manuscrite. La reconnaissance de l'écriture manuscrite traite d'objets complexes, où une représentation plus précise des objets est nécessaire. Par exemple, pour la reconnaissance des caractères asiatiques, les méthodes doivent être capables de gérer correctement les objets présentant des concavités.

En ce qui concerne le positionnement spatial, deux approches peuvent être distinguées : le positionnement absolu et le positionnement relatif. Pour la modélisation de positionnement relatif, de nombreuses méthodes ont été proposés, avec des degrés différents de précision. Nous présentons ici les méthodes dites qualitatives. Dans le cas de ces méthodes, les auteurs définissent a priori, selon le domaine étudié, un découpage de l'espace. Les relations spatiales qui peuvent être représentées par les méthodes qualitatives sont de deux types : directionnelles et topologiques. Dans ces approches, les objets sont représentés par des rectangles, des cercles et des formes irrégulières.

Les relations directionnelles décrivent l'ordre des objets dans un espace défini (par exemple « nord », « sud »). Le modèle défini par (Papadias et Theodoridis, 1997) est composé par exemple de 9 relations considérées comme suffisantes pour exprimer n'importe quelle relation directionnelle entre deux régions, ou deux régions sont modélisées par des rectangles. Les relations topologiques décrivent les concepts de voisinage, incidence et chevauchement et sont invariantes à des transformations telles que des modifications d'échelle et des rotations. Dans le modèle proposé par (Egenhofer et Herring, 1994), 8 relations topologiques entre deux objets ont été définies : *disjoint, touch, overlap, cover, covered by, contain, inside et equal*. Cette approche est appliquée à des objets de formes arbitraires. En général, les approches qualitatives manquent de précision. Par exemple, les relations directionnelles ne donnent aucune information sur la distance existant entre les objets. L'introduction d'une distance flexible permet d'augmenter l'expressivité et la précision d'un modèle. Cette approche a été proposée par (Maredj *et al.*, 2008) Pour se faire, ils ont introduit deux nouveaux concepts : "au plus proche" et "au plus éloigné" entre deux régions rectangulaires.

Les méthodes basées sur les CRF (champs aléatoires conditionnels) sont aussi connues pour être capable d'analyser et de modéliser la disposition du document.

Elles sont utilisées pour modéliser les interdépendances spatiales entre les différentes régions des documents. Cependant, les interdépendances spatiales modélisées avec les CRF sont limitées à des petites parties de l'espace. Par exemple, (Shetty *et al.*, 2007) modélisent des relations spatiales entre des "patchs" voisins, où un patch est approximativement de la taille d'un mot. En reconnaissance de la structure de document, nous souhaitons modéliser des relations spatiales entre des blocs de texte. Une modélisation au niveau du mot est alors trop limitée pour cette tâche.

Dans la méthode DMOS-P (Coüasnon, 2006), le sens de lecture est manuellement défini par l'utilisateur à l'aide du curseur. Cette approche donne de bons résultats mais nécessite une intervention humaine. Dans (Bloch et Ralescu, 2003), des paysages flous ont été utilisés. Dans cette méthode, un ensemble flou est calculé qui assigne à chaque point de l'espace un degré d'appartenance permettant de mesurer la concordance à la relation spatiale considérée. (Delaye *et al.*, 2009) développe l'idée présentée par Bloch. Ils introduisent une nouvelle définition générale de modèle de positionnement qui autorise des variabilités dans les degrés d'appartenance pour chaque relation directionnelle. Ces approches utilisant des paysages flous sont particulièrement adaptées pour modéliser des relations complexes comme celles qui existent dans la reconnaissance d'écriture manuscrite. Cependant, ce degré de précision n'est pas adapté pour la reconnaissance de la structure de document où la vérité terrain contient généralement la boîte englobante de chaque élément.

Dans cet article, nous introduisons un apprentissage automatique des opérateurs de position basé sur un ensemble d'apprentissage annoté. Dans cette méthode, les objets sont représentés par leur boîte englobante. Nous voulons définir des zones d'intérêt où nous allons rechercher un type d'élément. Les frontières de cette zone ne doivent pas être utilisées pour identifier la classe d'un élément. Le but de notre méthode de positionnement est de permettre un bon rappel tout en limitant la combinatoire. La précision est déterminée par la qualité des règles et des conditions définies. Les aspects directionnels sont introduits dans notre analyse à travers la détermination automatique de la meilleure direction pour analyser la zone.

## 3. Apprentissage interactif de positionnements

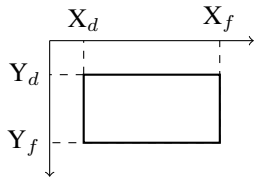



**Figure 1.** *Représentation des objets*

Dans cet article, nous présentons un nouvel outil, LearnPos, capable d'analyser un ensemble d'apprentissage annoté de documents pour apprendre des opérateurs de

position. Par exemple, nous possédons un corpus de courriers manuscrits et nous souhaitons apprendre la position des « date/lieu » dans le corpus. Dans notre méthode, chaque composant d'une page est représenté sous la forme de sa boîte englobante. Cette représentation est souvent utilisée dans les méthodes existantes, comme nous l'avons présenté dans la section précédente. Nous tenons alors compte de deux point pour représenter l'objet : les coordonnées du coin haut gauche et les coordonnées du coin bas droit, comme illustré dans la figure 1. Pour chaque document du corpus d'apprentissage, la boîte englobante de chaque élément est connue. En utilisant la vérité terrain, LearnPos calcule un opérateur de position que l'utilisateur peut intégrer directement dans son code simplement en le copiant-collant.

Un opérateur de position peut être composé de plusieurs zones séparées. Par exemple, dans notre corpus de lettres manuscrites, une catégorie regroupe les post-scriptum et les pièces jointes (appelée "PS/PJ"). La figure 2(b) représente une vue de toutes les occurrences de "PS/PJ" présentes dans l'ensemble d'apprentissage. Deux groupes distincts existent. Si nous ne tenons pas compte de ces groupes pour le calcul de l'opérateur de position, comme c'est fait dans la figure 2(c), nous obtenons alors pratiquement toute la page comme zone d'intérêt. Au contraire, quand les deux groupes sont pris en compte et une zone produite pour chaque groupe (cf. figure 2(d)), alors l'opérateur de position obtenu est bien plus précis.

L'utilisateur définit donc la structure logique du document et il peut utiliser LearnPos pour être assisté dans la définition de la structure physique. LearnPos est découpé en plusieurs étapes d'analyse automatique :

1) L'utilisateur demande explicitement l'opérateur de position P d'un élément

2) Calcul de P (précédé par une détection de groupes et la détection des valeurs extrêmes) : une ou plusieurs frontières de zone(s) sont produites

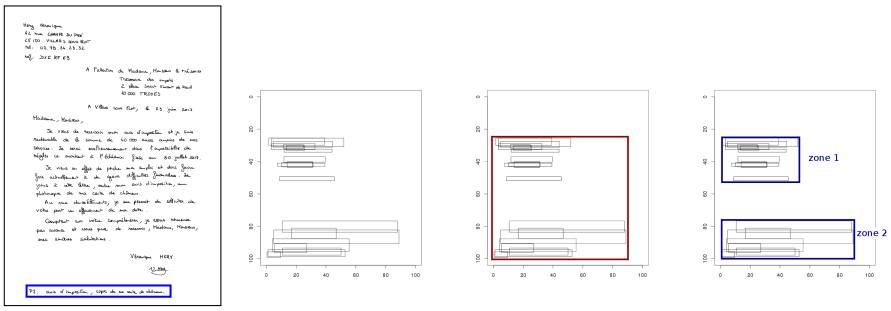
3) Détermination du meilleur point de vue pour chaque zone

4) Détermination de l'ordre d'analyse des zones

5) Introduction de l'opérateur de position P dans la structure logique définie par l'utilisateur. L'utilisateur peut revenir à la première étape pour un autre élément.

Comme c'est fréquemment le cas dans la littérature, nous utilisons deux modes de positionnement dans un document : positionnement absolu et positionnement relatif. Un positionnement absolu consiste à décrire la position de chaque élément par rapport à un élément fixe, ici la page entière, indépendamment de la position des autres composants. Cette approche est bien adaptée pour des éléments dont la position est stable dans la page. Le positionnement relatif est plus associé à notre perception des similarités dans les arrangements spatiaux. Par exemple, dans le cas des lettres manuscrites, nous sommes capables d'exprimer des relations telles que "la date/le lieu sont au dessus des coordonnées destinataires". Le positionnement relatif s'intéresse à la position des éléments entre eux et non pas à leur position absolue dans la page.

Dans cette partie, nous allons d'abord détailler le rôle de l'utilisateur dans notre méthode. Nous présenterons ensuite la différence de traitement entre le positionne-




(a) Exemple de lettre contenant PS/PJ

(b) Position des PS/PJ pour un corpus de 300 pages

(c) Calcul d'une zone unique pour tous les éléments

(d) Calcul d'une zone par groupe

**Figure 2.** *Exemple de calcul de l'opérateur de position pour les éléments "PS/PJ". La boîte englobante de chaque occurence est représentée par un rectangle dans une page normalisée.*

ment absolu et le positionnement relatif. Les différentes étapes de la chaîne de traitement seront présentés : détection des groupes, détection et traitement des valeurs extrêmes et enfin le calcul des frontières de chaque zone.

### 3.1. *Rôle de l'utilisateur*

Pour utiliser LearnPos l'utilisateur doit seulement fournir la vérité terrain sur laquelle sera effectué l'apprentissage ainsi que l'élément pour lequel il souhaite obtenir un opérateur de position.

La vérité terrain doit contenir pour chaque document les boîtes englobantes de chaque élément. Un document peut contenir plusieurs éléments appartenant à la même classe. Par exemple, un document peut contenir deux signatures. Le traitement de cette multiplicité dépend du type de positionnement désiré par l'utilisateur.

L'utilisateur doit spécifier quel élément doit être positionné. Il doit également indiquer s'il désire un positionnement absolu ou un positionnement relatif. Dans le cas d'un positionnement relatif, l'utilisateur doit également indiquer quel élément doit être positionné en fonction de quel autre.

Après l'analyse, LearnPos produit un opérateur de position que l'utilisateur peut simplement copier-coller dans son code. Les opérateurs de position produits par LearnPos sont écrits de la même manière que ceux produits par un opérateur humain. Cela permet de la flexibilité, l'utilisateur peut modifier les valeurs obtenues s'il a des informations complémentaires à ajouter par exemple.

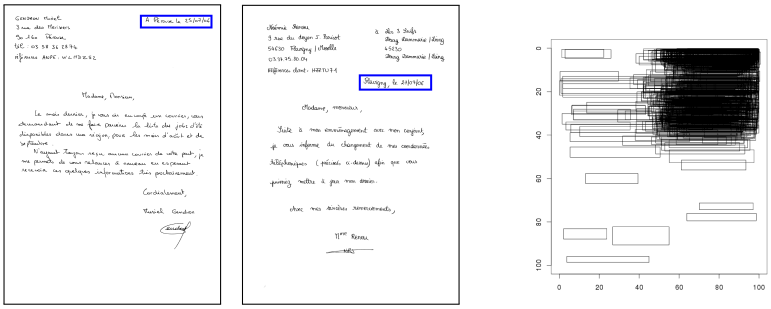
### 3.2. *Positionnement absolu et relatif*

Comme nous l'avons présenté, LearnPos permet d'obtenir un opérateur de position selon deux modes de positionnements : absolu et relatif. Dans cette partie, nous détaillons les différences de traitement entre ces deux modes de positionnement.

#### 3.2.1. *Positionnement absolu*

Le positionnement absolu consiste à décrire la position d'un élément dans la page entière indépendamment de la position des autres éléments. Cette approche est particulièrement utile pour démarrer l'analyse, lorsqu'aucun autre élément n'a été trouvé. Par exemple, l'utilisateur peut utiliser un positionnement absolu pour l'élément "date/lieu" des courriers manuscrits dont la position est présentée dans la figure 3.

Le positionnement absolu ne nécessite pas de prétraitement de la vérité terrain. Dans chaque document, l'élément recherché peut être présent ou absent. LearnPos utilise seulement les documents où l'élément est présent pour calculer l'opérateur de position. Cependant, le nombre de documents où l'élément est absent est fourni à l'utilisateur qui peut vouloir modifier sa description de la structure logique en fonction de cette information.




(a) Exemple de position de "date/lieu" (b) Exemple de position de "date/lieu" (c) Boîtes englobantes des "date/lieu" dans une page normalisée

**Figure 3.** *Représentation de la position des boîtes englobantes des "date/lieu" du corpus d'apprentissage des lettres manuscrites. Chaque rectangle représente une boîte englobante.*

#### 3.2.2. *Positionnement relatif*

Le positionnement relatif consiste en la description de la position d'un élément en fonction d'un autre. Dans cet article, on appelle *relatif* l'élément que l'on cherche à positionner en fonction d'un autre qui est lui appelé *référence*. Par exemple, l'utilisateur peut utiliser un positionnement relatif pour situer l'élément "coordonnées destinataire" en fonction de l'élément "date/lieu" dans des courriers manuscrits.

Dans le cas du positionnement relatif, la vérité terrain nécessite un prétraitement afin de produire l'opérateur de position. Tout d'abord, les documents ne comportent

pas forcément un relatif et une référence. Si plusieurs références et relatifs sont présents dans le document, une identification des différentes paires est nécessaire comme nous l'avons présenté dans la section précédente. Nous dénombrons tous les cas apparaissant dans notre vérité terrain. L'utilisateur est informé des effectifs dans chaque catégorie. Dans le cas où le document présente un relatif mais pas de référence, un opérateur de position *absolu* est calculé.

Nous traitons ensuite le cas où nous avons des paires de références et de relatifs. La référence est réduite à son coin haut gauche ($X_d$, $Y_d$). Une translation du système de coordonnées est effectuée. Les coordonnées des éléments relatifs sont modifiées pour que l'origine se situe à ce point ($X_d$, $Y_d$). En utilisant ces nouvelles coordonnées pour la boîte englobante de l'élément relatif, nous pouvons alors procéder comme dans le cas du positionnement absolu.

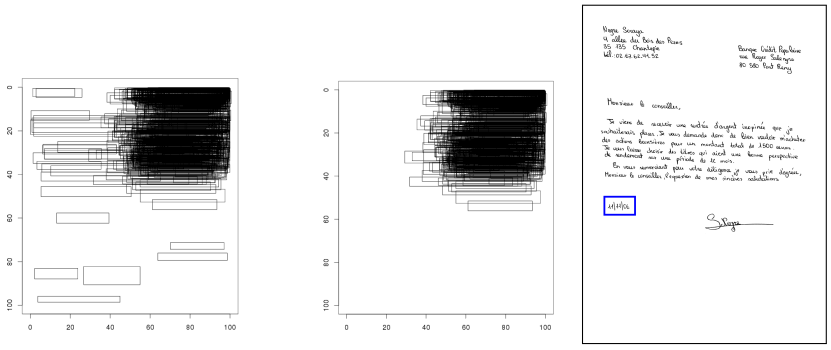
### 3.2.3. *Détection automatique des groupes*

Comme nous l'avons présenté dans la figure 2, il est nécessaire d'avoir des opérateurs de position pouvant être composés de plusieurs zones. Afin de conserver une méthode automatique, nous devons être capables de détecter automatiquement si une ou plusieurs zones existent. LearnPos ne peut donc pas s'appuyer sur une prédétermination manuelle du nombre de zones d'intérêt. LearnPos trouve automatiquement le nombre de groupes en s'appuyant sur la détection des maxima locaux dans un histogramme. Le nombre total de maximum locaux correspond au nombre de zones différentes dans la page. Cette détection est faite via une méthode de recherche séquentielle, présentée par Lerddaradsamee (Lerddararadsamee et Jiraraksopakun, 2012). Un paramètre est fixé qui détermine la capacité de la méthode à trouver les petits groupes en termes d'effectif dans l'ensemble d'apprentissage.

L'utilisation de cette technique nous amène à nous interroger sur le problème de la stabilité des maxima locaux détectés. Pour résoudre ce problème, nous pouvons utiliser différents lissages de la densité et conserver les maxima locaux qui sont présents avec tous les lissages testés. Comme référence, nous utilisons le sélecteur de Sheather et Jones(Sheather et Jones, 1991), qui est un sélecteur se basant sur les données présentes. Cette technique nous permet de détecter les groupes distincts qui nécessitent des zones distinctes dans l'opérateur de position.

### 3.2.4. *Détection automatique des valeurs extrêmes*

Dans notre analyse, nous ne souhaitons pas modéliser les valeurs extrêmes qui regroupent deux types d'éléments : les éléments atypiques et les éléments aberrants, provenant par exemple d'une erreur dans la vérité terrain. Pour la détection de ces éléments, nous nous appuyons sur une détection classique à l'aide de l'écart-type. Nous considérons que tout point n'appartenant pas à $[moyenne - t \times SD; moyenne + t \times SD]$ est une valeurs extrême, où SD désigne l'écart-type de la distribution. Si le nombre d'observations est supérieur à 80 ; $t = 3$, sinon $t = 2,5$. La détection des valeurs extrêmes est effectuée après la détection des groupes, cette détection étant faite dans chaque groupe et non dans la population totale.



(a) Position des boîtes englobantes des "date/lieu" dans une page normalisée pour 300 documents.

(b) Suppression des valeurs extrêmes

(c) Exemple de "date/lieu" avec une position atypique détectée

**Figure 4.** *Représentation de la position des boîtes englobantes des "date/lieu" du corpus d'apprentissage des lettres manuscrites. Chaque rectangle représente une boîte englobante.*

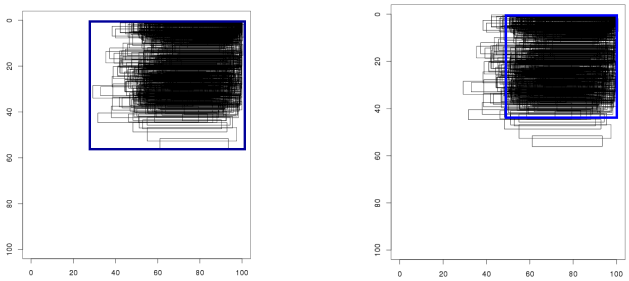
Nous faisons ici une détection des valeurs extrêmes variable par variable, pour les positions $X_d$, $Y_d$, $X_f$ et $Y_f$ ainsi que pour la largeur et la hauteur de la boîte englobante. Toutes les zones détectées comme des valeurs extrêmes possibles ne sont pas utilisées pour le calcul de l'opérateur de position.

La figure 4 présente les positionnements de l'élément "date/lieu" dans des lettres manuscrites. Comme nous pouvons l'observer certains éléments sont détectés comme valeurs extrêmes et supprimer de l'analyse globale. Cependant, ces valeurs extrêmes peuvent elles-mêmes être séparées en deux groupes. Un groupe est composé des valeurs extrêmes présentes dans la partie haute gauche de la page, l'autre groupe étant composé des éléments se situant en bas de la page. Le groupe "haut gauche" présente suffisamment d'exemples pour que nous souhaitions définir une zone afin de reconnaître ces éléments. Le groupe "bas" ne présente pas suffisamment d'éléments pour que nous choisissions de les modéliser. Un exemple d'élément "date/lieu" se situant en bas de la page est présenté dans la figure .

### 3.2.5. *Méthode basée sur la densité pour le calcul des frontières de zones*

Lorsque les groupes ont été détectés et les valeurs extrêmes supprimées, les frontières de chaque zone sont alors définies. La figure 5 représente la boîte englobante des positions des "date/lieu" dans les lettres manuscrites, une fois les valeurs extrêmes supprimées comme indiqué dans la section précédente. Si nous prenons simplement la zone englobant toutes les boîtes englobantes, la zone risque d'être trop large et d'introduire des erreurs dans l'analyse.

Nous utilisons alors la notion de densité et de grille afin de réduire les frontières de la zone. Pour cela, nous divisons d'abord l'espace en une grille où la largeur d'une




(a) Zone obtenue en incluant toutes les boîtes englobantes

(b) Zone obtenue avec notre méthode basée sur la densité

**Figure 5.** *Exemple d'ajustement de frontières de la zone avec une méthode basée sur la densité*

case représente un pourcent de la largeur de la page et la hauteur d'une case représente un pourcent de la hauteur de la page. Pour chaque case, nous comptons le nombre de boîte englobante ayant une intersection non vide avec cette cellule. Cette information est utilisée comme une approximation de la densité. Puis, si une case contient moins d'éléments qu'un seuil fixé, alors nous fixons l'effectif de cette case à zéro. Toutes les cases qui ont une densité non nulle sont ensuite utilisées pour définir l'opérateur de position. Les zones sont des unions d'unités non vides de l'espace. La forme finale de la zone peut alors être variée. Dans LearnPos, nous utilisons pour l'instant des zones rectangulaires. Chaque opérateur de position peut être composé d'une ou plusieurs de ces zones rectangulaires.

## 4. Comment gérer la confusion

Après la chaîne de traitement présentée dans la section précédente, notre outil a déterminé une ou plusieurs zones pour l'opérateur de position demandé. Comme nous l'avons indiqué, les zones ne sont pas des frontières strictes pour la détermination d'un type d'élément. Même si la zone contient effectivement l'élément recherché, elle peut également contenir de nombreux autres éléments qui apportent de la confusion dans notre analyse. En effet, ces autres éléments peuvent être sélectionnés à la place de l'élément recherché.

Une manière de gérer ces risques de confusion et de définir des règles et des conditions permettant de déterminer quel élément est celui que nous recherchons parmi les autres. Cependant, les caractéristiques peuvent ne pas différer suffisamment pour que nous puissions choisir facilement le bon élément. Une autre manière de gérer ces risques de confusion est présentée ici. Elle consiste à créer des opérateurs de position avec une notion d'ordre et de point de vue. Dans ce but, nous introduisons un nouvel indice dans notre analyse, l'*indice de confusion*. Cette notion est introduite en deux points de notre analyse. Une première fois, nous l'utilisons pour ordonner les diffé-

rentes zones trouvées. Puis, nous l'utilisons pour savoir dans quel ordre parcourir les éléments dans la zone.

*L'indice de confusion*

L'indice de confusion est un indicateur qui permet de savoir combien d'éléments d'un autre type que celui recherché sont présents dans la zone. Il est défini comme : $confusion(list, zone) = \sum_{i=1}^{n} list[i] \cap zone \neq \emptyset$. $list$ contient tous les éléments à l'exception de ceux du type recherché. Dans le cas de positionnement relatif, nous devons exclure à la fois les références et les relatifs pour obtenir les meilleurs résultats possibles.

*Ordre des zones*

Quand plus d'une zone a été trouvée, nous analysons les éléments en parcourant chacune des zones tant que nous n'avons pas trouvé l'élément recherché. Pour minimiser les erreurs, nous devons commencer par les zones qui présentent l'indice de confusion le plus petit. En procédant de cette manière, nous limitons les possibilités de choisir un élément d'un autre type que celui recherché. Nous calculons donc l'indice de confusion dans chacune des zones trouvées pour l'opérateur de position. Les zones sont ensuite analysées par valeur croissante d'indice de confusion.

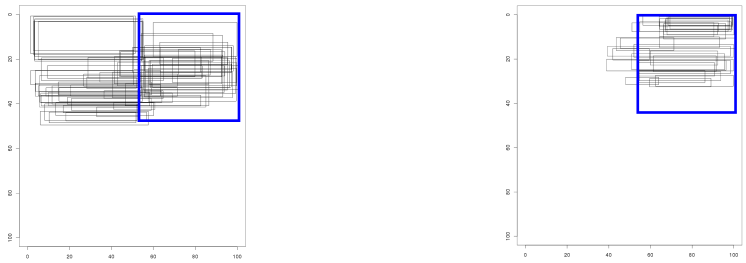
Par exemple, LearnPos détecte deux zones différentes pour les "coordonnées destinataires en fonction de "date/lieu". Une zone (a) correspond aux "coordonnées destinataire" au dessus de "date/lieu". L'autre zone (b) correspond aux "coordonnées destinataire" sous "date/lieu". L'analyse commencera par la zone (a) car peu de confusions sont possibles tandis que la zone (b) peut contenir l'ouverture ou le corps de texte de la lettre qui vont apporter de la confusion dans l'analyse.

*Choisir le meilleur point de vue*

Lorsque nous cherchons à reconnaître un élément dans un document, nous ne devons pas seulement savoir dans quelle partie de l'image le rechercher. Nous devons également définir dans quel ordre nous allons parcourir les éléments de cette zone, c'est-à-dire déterminer le point de vue utilisé. Les éléments seront alors parcourus dans la zone du plus proche au plus éloigné. Nous choisissons ici de déterminer le meilleur point de vue automatiquement, à l'aide de l'indice de confusion afin de minimiser les risques d'erreurs. Pour définir le meilleur point de vue, nous calculons l'indice de confusion pour un ensemble prédéterminé de points de vue. Nous choisissons alors le point de vue qui minimise l'indice de confusion.

Dans l'exemple de "date/lieu" dans les lettres manuscrites, la figure 6(a) présente une représentation synthétique de tous les boîtes englobantes des éléments pouvant apporter de la confusion. Pour augmenter la lisibilité de la figure, l'exemple se base sur un sous-ensemble de 30 lettres. La figure 6(b) représente elle les éléments "date/lieu" que nous recherchons pour ces mêmes 30 lettres. Le point de vue est choisi afin de maximiser les chances de trouver l'élément recherché tout en minimisant les risques

de sélectionner un autre élément de la page. Dans ce but, nous choisissons ici de parcourir les éléments de la zone du haut vers le bas.



(a) Représentation de tous les éléments pouvant apporter de la confusion

(b) Représentation des éléments recherchés ("date/lieu")

**Figure 6.** *L'opérateur de position de "date/lieu" est analysé du haut vers le bas pour maximiser les chances de sélection l'élément "date/lieu" tout en minimisant les risques de sélectionner un autre élément*

## 5. Évaluation et résultats

Pour l'évaluation, nous avons utilisé LearnPos afin d'obtenir des opérateurs de position pour la méthode grammaticale DMOS-P. Deux aspects ont été évalués. D'abord, nous avons comparé les résultats obtenus par la même grammaire avec des opérateurs de position définis manuellement et les opérateurs de position définis par LearnPos. Puis, nous avons validé l'intérêt de l'utilisation de l'indice de confusion en comparant les résultats obtenus en choisissant le point de vue grâce à l'indice de confusion par rapport à un point de vue fixe. Ces deux aspects ont été évalués sur un corpus de lettres manuscrites. Nous allons d'abord présenter le contexte des expérimentations puis nous détaillerons les résultats obtenus.

### 5.1. *Contexte des expérimentations*

***RIMES*** La campagne d'évaluation RIMES[1], Recherche et Indexation de données Manuscrites et fac-similES (Grosicki *et al.*, 2009), a créé une base publique contenant des lettres manuscrites et des fax. Ces documents peuvent être utilisés dans différentes tâches liées à la reconnaissance de document : analyse de la structure, identification de l'auteur, reconnaissance d'écriture manuscrite, etc. Les images sont scannées en niveaux de gris à 300 dpi. Toutes les images ont été manuellement annotées pour obtenir la vérité terrain.

1. RIMES : *Reconnaissance et Indexation de données Manuscrites et fac-similES*, `http ://www.rimes-database.fr/doku.php`

La reconnaissance de la structure des documents dans RIMES consiste à identifier jusqu'à huit types de contenus distincts : coordonnées expéditeur, coordonnées destinataire, date/lieu, objet, ouverture, corps de texte, signature et PS/PJ. La taux de reconnaissance pour cette tâche est défini comme le rappel par classe : $Rappel = \frac{\text{nombre de pixels noirs affectés}}{\text{nombre de pixels noirs attendus}}$. Seuls les pixels noirs sont utilisés afin de ne pas inclure le fond.

***Méthode DMOS-P*** La méthode DMOS (Description and MOdification of the Segmentation) (Coüasnon, 2006) est une méthode grammaticale pour la reconnaissance de documents structurés. Cette méthode est utilisée ici pour illustrer l'efficacité de LearnPos. LearnPos est un outil indépendant de DMOS et peut être utilisé avec un autre système de reconnaissance. L'analyse est guidée par les opérateurs de position. Un opérateur de position est composé des zones à analyser (coordonnées des points haut gauche et bas droit) et de la position du curseur. Le curseur est utilisé pour parcourir les composants dans la zone dans un certain ordre, du plus proche du curseur au plus éloigné. Pour chaque opérateur de position de la grammaire, six paramètres doivent donc être définis. Dans la méthode DMOS standard, ces paramètres doivent être définis manuellement par l'utilisateur.

### 5.2. *Évaluation des opérateurs de position définis avec LearnPos*

Une grammaire a été créée en 2008 par (Lemaitre *et al.*, 2008) pour la campagne d'évaluation RIMES. Dans cette grammaire, les opérateurs de position étaient définis manuellement. Nous avons généré avec LearnPos des opérateurs de position pour cette grammaire et nous avons introduit les opérateurs de position inférés dans la grammaire déterministe existante. L'objectif est de vérifier que les opérateurs de position ainsi obtenus sont corrects et d'automatiser une partie de la conception d'une grammaire.

L'échantillon d'apprentissage est composé des mêmes 300 images que celui qui a servi pour la définition manuelle des opérateurs. Dans le cas des opérateurs de position définis manuellement, les 300 images ont été observées manuellement par un opérateur humain afin de déterminer les meilleurs paramètres possibles. L'observation manuelle de 300 images est évidemment une tâche coûteuse et la synthèse de ces observations pour produire un opérateur de position n'est pas une tâche aisée.

Nous avons utilisé LearnPos pour calculer des opérateurs de position absolus pour "coordonnées expéditeur", "date/lieu", "signature" et "objet". LearnPos a également été utilisé pour calculer un opérateur de position relatif pour positionner "coordonnées destinataire" en fonction de "date/lieu".

*Conservation du taux de reconnaissance*

Après avoir obtenu les opérateurs de position grâce à LearnPos, nous les avons intégrés dans la grammaire existante. Nous avons alors pu comparer les taux de reconnaissance obtenus sur la base de validation. Le tableau 1 indique le taux de re-

connaissance pour chaque classe. L'utilisation des opérateurs de position définis manuellement et de ceux définis automatiquement par LearnPos donne des résultats similaires. Quelques éléments voient leur rappel amélioré. C'est le cas par exemple de la "signature" ce qui a de plus un impact sur la détection du corps de texte et de l'ouverture. Pour certains éléments, le rappel est légèrement inférieur mais nous avons globalement une amélioration des résultats. De plus, LearnPos a permis une réduction importante du temps nécessaire à la détermination des opérateurs de positions. L'utilisateur doit seulement spécifier pour quel élément il souhaite obtenir un opérateur de position puis il peut simplement copier-coller le résultat obtenu grâce à LearnPos.

| Classe | (1) Opérateurs de position manuels | (2) Opérateurs de position inférés avec LearnPos |
|---|---|---|
| **Corps de texte** | **91.4** | **93.9** |
| Expéditeur | 91.3 | 91.3 |
| Destinataire | 85.1 | 83.6 |
| **Signature** | **88.1** | **91.4** |
| Objet | 66.8 | 64.5 |
| **Date/lieu** | **75.6** | **76.7** |
| **Ouverture** | **77.1** | **80.5** |
| PS/PJ | 9.6 | 9.6 |
| Total | 88.6 | 90.2 |

**Tableau 1.** *Comparaison des taux de rappel obtenus sur 950 documents avec deux grammaires : (1) la version existante, avec des opérateurs de position définis manuellement et (2) notre version, avec des opérateurs de position inférés avec LearnPos.*

Avec cette technique, nous avons réduit le nombre de paramètres définis manuellement comme nous pouvons le voir dans le tableau 2 : 48 paramètres sont maintenant appris de manière automatique avec LearnPos. Le nombre total de paramètres est modifié car certains opérateurs de position sont maintenant composés de plusieurs zones, ce qui n'était pas le cas avant. Cependant, certains opérateurs de position utilisés dans la grammaire ne peuvent pas être appris automatiquement avec LearnPos. C'est le cas des opérateurs de position définissant des relations entre les lignes (au dessus d'une ligne, sous une ligne, etc.). Nous ne pouvons pas utiliser LearnPos pour ces opérateurs de positions car la vérité terrain ne contient pas d'information au niveau des lignes. Avec une vérité terrain adaptée, LearnPos aurait pu être utilisé pour tous les opérateurs de position de la grammaire.

| | Paramètres manuels | Paramètres automatiques |
|---|---|---|
| Grammaire existante | 102 | 0 |
| Notre grammaire | 66 | 48 |

**Tableau 2.** *LearnPos permet une diminution du nombre de paramètres fixés manuellement. Les 66 paramètres qui restent fixés manuellement n'ont pas de vérité terrain disponibles.*

*Sens de lecture*

Nous souhaitons maintenant montrer l'importance du choix du point de vue, en comparant les résultats obtenus en plaçant le curseur au milieu de la zone et en choisissant le curseur proposé par LearnPos. Nous avons limité nos expérimentations à l'élément "date/lieu" car utilisant une grammaire existante, un changement de point de vue pour tous les éléments aurait nécessité des changements trop profonds de la grammaire. Dans les deux expérimentations, les frontières des zones et l'ordre d'analyse des zones sont identiques, obtenus avec LearnPos.

Comme nous pouvons le voir dans le tableau 3, le rappel est amélioré lorsque nous utilisons le point de vue déterminé avec LearnPos de 75,6% à 76,7%. Les erreurs en nombre de pixels sont réduites de 4,5% avec ce point de vue inféré avec LearnPos.

| Classe | (1) Point de vue au centre | (2) Point de vue inféré |
|---|---|---|
| Corps de texte | 93,9 | 93,9 |
| Expéditeur | 91,3 | 91,3 |
| Destinataire | 83,7 | 83,6 |
| Signature | 91,4 | 91,4 |
| Objet | 64,7 | 64,5 |
| **Date/lieu** | **75,6** | **76,7** |
| Ouverture | 80,5 | 80,5 |
| PS/PJ | 9,6 | 9,6 |
| Total | 90,1 | 90,2 |

**Tableau 3.** *Taux de rappel évalué sur 950 documents selon deux méthodes : (1) le point de vue est positionné au centre de la zone de (2) le point de vue est déterminé par LearnPos.*

## 6. Conclusion

Dans cet article, nous avons présenté un nouvel outil pour le positionnement interactif pour les systèmes d'analyse de documents. LearnPos est un outil générique indépendant de tout système de reconnaissance. La structure logique est définie par l'utilisateur et LearnPos est utilisé par l'utilisateur pour définir la structure physique, sans avoir besoin d'explorer manuellement le corpus d'apprentissage. LearnPos infère automatiquement des opérateurs de position qui peuvent être intégrés directement dans le système de reconnaissance utilisé. Cette analyse est dirigée par l'utilisateur qui demande explicitement les opérateurs de position qui l'intéressent. Les opérateurs de position inférés sont du même type que ceux définis manuellement, ils sont donc facilement compréhensibles et peuvent être modifiés par l'utilisateur. Le concept de confusion permet de déterminer le meilleur ordre de parcours des différentes zones et le meilleur point de vue à adopter pour chaque zone.

LearnPos permet la simplification de l'écriture d'une grammaire pour l'utilisateur et la réduction du temps nécessaire à cette écriture. LearnPos analyse de manière ex-

haustive l'échantillon d'apprentissage ce qui permet une détermination des cas généraux comme des cas rares. Les expérimentations sur des lettres manuscrites on montré que notre outil obtient des résultats comparables à ceux obtenus avec des opérateurs de positions ajustés manuellement.

## 7. Bibliographie


Bloch I., Ralescu A. L., « Directional relative position between objects in image processing : a comparison between fuzzy approaches », *Pattern Recognition*, vol. 36, n° 7, p. 1563-1582, 2003.

Coüasnon B., « 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*, vol. 8, n° 2, p. 111-122, June, 2006.

Delaye A., Macé S., Anquetil E., « Modeling Relative Positioning of Handwritten Patterns », *14th Biennial Conference of the International Graphonomics Society*, p. 122-127, September, 2009.

Egenhofer M. J., Herring J., Categorizing binary topological relations between regions, lines, and points in geographic databases, Technical report, University of Maine, 1994.

Grosicki E., Carree M., Brodin J.-M., Geoffrois E., « Results of the RIMES Evaluation Campaign for Handwritten Mail Processing », *Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on*, p. 941-945, 2009.

Lemaitre A., Camillerapp J., Coüasnon B., « A generic method for structure recognition of handwritten mail documents », *Document Recognition and Retrieval DRR XV*, San Jose, États-Unis, 2008.

Lerddararadsamee T., Jiraraksopakun Y., « Local maximum detection for fully automatic classification of EM algorithm », *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2012 9th International Conference on*, p. 1-4, 2012.

Maredj A.-E., Nourreddine T., Sadallah M., Alimazighi Z., « A flexible distance for the spatial placement in a multimedia document », *Information and Communication Technologies : From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*, p. 1-4, 2008.

Papadias D., Theodoridis Y., « Spatial Relations, Minimum Bounding Rectangles, and Spatial Data Structures », *International Journal of Geographic Information Science*, vol. 11, p. 111-138, 1997.

Sheather S. J., Jones M. C., « A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation », *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 53, n° 3, p. 683-690, 1991.

Shetty S., Srinivasan H., Srihari S., Shetty S., Srinivasan H., Beal M., Srihari S., « Segmentation and labeling of documents using Conditional Random Fields », *Document Recognition and Retrieval DRR XIV*, 2007.