

# Applications

---

## Localisation du bloc adresse sur les objets postaux par une méthode de segmentation ascendante : évaluation et optimisation

---

### *Address Block Location on Flat Mail Pieces by a Bottom-Up Segmentation Method. Evaluation and Optimization*

par Jean-Claude ORIOT, Dominique BARBA, Michel GILLOUX\*

IRESTE – Université de Nantes – SEI/IRESTE

La Chantrerie, CP 3003

F-44087 Nantes cedex 03

\* Service de Recherche Technique de la Poste

10, rue de l'île Mabon

F-44038 Nantes cedex 01

#### Résumé

Nous présentons une méthode de segmentation des zones de texte sur des documents complexes, peu ou pas structurés. Cette méthode est appliquée à la localisation du bloc adresse sur les grands objets postaux. Parmi ces objets, nous avons identifié deux catégories principales et nous présentons une méthode adaptée à chacune. Nous portons enfin une attention particulière sur la phase d'optimisation des traitements et d'évaluation des performances.

**Mots clés :** Analyse d'image, reconnaissance de formes, suivi de route, contours actifs.

#### Abstract

*We present a textual areas segmentation method for complex documents not or few structured. This method is applied to the address block location on large mail pieces. Among these objects, we have identified two main classes and we present a dedicated method for each. We then give more details about the optimization and evaluation of the implemented techniques.*

**Key words :** Document analysis, address block location on mail pieces, segmentation, optimization, valuation.

## 1. Introduction

Nous présentons dans cet article une synthèse d'un travail concernant le domaine de l'analyse de documents. De nombreux développements ont été réalisés depuis le milieu des années 80, cependant, l'essentiel des travaux a porté jusqu'à ce jour sur l'analyse et le traitement de documents fortement structurés tels que les publications scientifiques [DER, 90], les index bibliographiques et autres formulaires [HIG, 86]. Au contraire, l'étude présentée ici s'intéresse au traitement de documents peu ou pas structurés. Cette étude s'inscrit dans le cadre du projet d'automatisation du tri des objets postaux. Plus précisément, pour trier les grands objets plats (journaux, magazines, publicités, etc), il est indispensable de localiser le bloc adresse de destination avant de pouvoir appliquer la reconnaissance de caractères. Nous avons développé une

méthode de localisation de l'adresse sur ce type d'objets en distinguant deux situations selon que les objets sont visuellement peu complexes (on note peu chargés) ou complexes (on note chargés). Pour les objets peu chargés, le traitement fait appel à une analyse ascendante des composantes connexes guidée par un ensemble de règles (établies à partir d'une large étude statistique). Pour les objets chargés, le traitement fait intervenir une pré-segmentation originale fournissant une image binaire de plus haut niveau (rehaussement et filtrage). Notre étude n'incluant pas la phase de lecture, nous avons défini une méthodologie d'évaluation des traitements réalisés. Nous avons également cherché à optimiser le réglage des paramètres ainsi que la phase de classement des blocs candidats.

La première partie de cet article donne une description de l'application. Nous y présentons les objets plats, les objectifs de l'application et nous commentons les travaux précédents. Dans une deuxième partie, nous décrivons la stratégie retenue pour aborder

cette étude ainsi que les traitements réalisés. Cette partie est articulée autour de la distinction précédente en objets peu chargés et chargés. La troisième partie traite de la méthodologie d'évaluation et des optimisations réalisées. Les performances obtenues sont alors commentées.

## 2. Description de l'application

L'application consiste à déterminer la région correspondante au bloc adresse de destination (on note BAD) sur les grands objets postaux afin d'appliquer la reconnaissance de l'écriture sur cette seule région d'intérêt. Seule la phase de localisation est étudiée ici. Nous considérons que la région d'intérêt est d'allure rectangulaire et qu'elle peut être placée de façon horizontale ou verticale vis-à-vis du bord de l'image. Nous excluons les cas où l'adresse est fortement inclinée car ceux-ci correspondent le plus souvent au cas particulier d'une étiquette blanche collée de travers [ORI, 92]. L'objectif est donc de sélectionner dans l'image entière une région d'intérêt. A partir des exemples suivants (figures 1 et 2), on peut apprécier la diversité des situations à traiter.

Cette série d'illustrations rend explicite la distinction des images en objets peu chargés et chargés. De plus, venant renforcer la légitimité de cette classification, l'analyse de bases d'images et des données descriptives [ORI, 92] nous a permis d'associer, à chaque catégorie, des caractéristiques concernant l'image entière mais aussi concernant la nature du bloc adresse. Ainsi, globalement, la localisation du BAD sur les objets plats présente deux problématiques principales. D'abord il faut être capable de détecter des blocs de texte aux caractéristiques extrêmement variables dans un environnement pas très complexe présentant globalement des zones d'information bien séparées spatialement (ex : BAD manuscrit sur fond uniforme, objet peu chargé). Il faut aussi être capable de détecter des blocs de texte de caractéristiques peu variables dans un environnement très complexe (ex : BAD dactylographié sur la couverture d'un magazine, objet chargé).

Nous souhaitons d'une part démontrer la faisabilité de la localisation sur les objets plats non chargés avec un taux de succès important par une méthode assez directe (analyse par représentation symbolique puis phase de classement). Par ailleurs, et dans l'esprit de porter l'essentiel de nos efforts sur les traitements de bas niveau, nous avons développé une chaîne de traitements particulière pour traiter les objets plats chargés.

Cette dernière catégorie n'a presque jamais été étudiée spécifiquement, d'une part parce qu'elle représente un groupe minoritaire et d'autre part parce qu'elle correspond à un ensemble important de difficultés a priori. Or il apparaît que le courrier de type commercial ou publicitaire se développe actuellement de façon importante et que les situations de proximité du BAD et d'informations extra-BAD deviennent plus fréquentes. En outre, cette

catégorie d'images entraîne de nombreuses difficultés d'analyse, parfois même inédites, et représente donc un intérêt particulier.

L'approche que nous allons développer s'appuie sur l'analyse critique de travaux antérieurs intéressant ce problème, travaux que nous résumons ci-après.

La séparation des zones de texte, des zones graphiques et des zones photographiques constitue l'un des thèmes majeurs de l'analyse de documents, et de très nombreuses méthodes ont été proposées pour réaliser ces tâches. [WON, 82] utilise l'algorithme RLSA pour séparer les régions. Ces régions sont ensuite classées à partir de caractéristiques statistiques. [WAN, 89] améliore cette approche en utilisant des caractéristiques texturales plus discriminantes. Toutefois, ces techniques de type global s'avèrent inefficaces lorsque l'image ne présente plus des régions disjointes d'allure rectangulaire contenant un seul type de contenu. Ces situations plus complexes ont donné lieu à des approches locales. [KUB, 83] suggère que le texte présente localement plus de « traits » que le graphique et caractérise la complexité en sommant les inverses des distances de ces traits au pixel courant. [FLE, 88] extrait les groupes de caractères alignés selon une direction quelconque au moyen de la transformée de Hough. [BIX, 88] décrit une méthode de suivi de caractères inspirée des techniques de vectorisation des graphiques. [MEY, 88] propose une segmentation utilisant une stratégie ascendante à partir d'une description des composantes connexes et de leurs principales relations topologiques.

Alors que les études précédentes supposent une résolution spatiale élevée de 200 à 300 dpi (« dot per inch »), des segmentations ascendantes à moyenne ou basse résolution spatiale ont été mises en œuvre pour des applications de localisation du BAD sur les objets postaux. Une résolution moyenne de 130 dpi est utilisée dans [SRI, 89]. Les composantes connexes sont d'abord regroupées en entités de type bloc et c'est à l'intérieur de ces blocs que la segmentation de type ligne est réalisée (ce qui ne nous semble pas adapté aux objets chargés). Dans [ANT, 88], un pré-traitement de l'image binaire (à très basse résolution : 32 dpi) a pour objectif de « consolider » puis de filtrer les régions non texte par regroupement de pixels en s'appuyant sur l'hypothèse que les pixels de texte et les pixels non-texte sont plus distants que ne le sont les pixels non-texte entre eux. Cette hypothèse n'est pas souvent valide pour les images complexes.

## 3. Méthodologie et traitements

Le choix de la résolution spatiale doit permettre de conserver les fréquences spatiales horizontales et verticales qui caractérisent respectivement le nombre de caractères et le nombre de lignes par unité de longueur. C'est pourquoi la très basse résolution de 32 dpi a été écartée. Nous avons retenu une résolution de 75 dpi.

# Applications

Localisation du bloc adresse

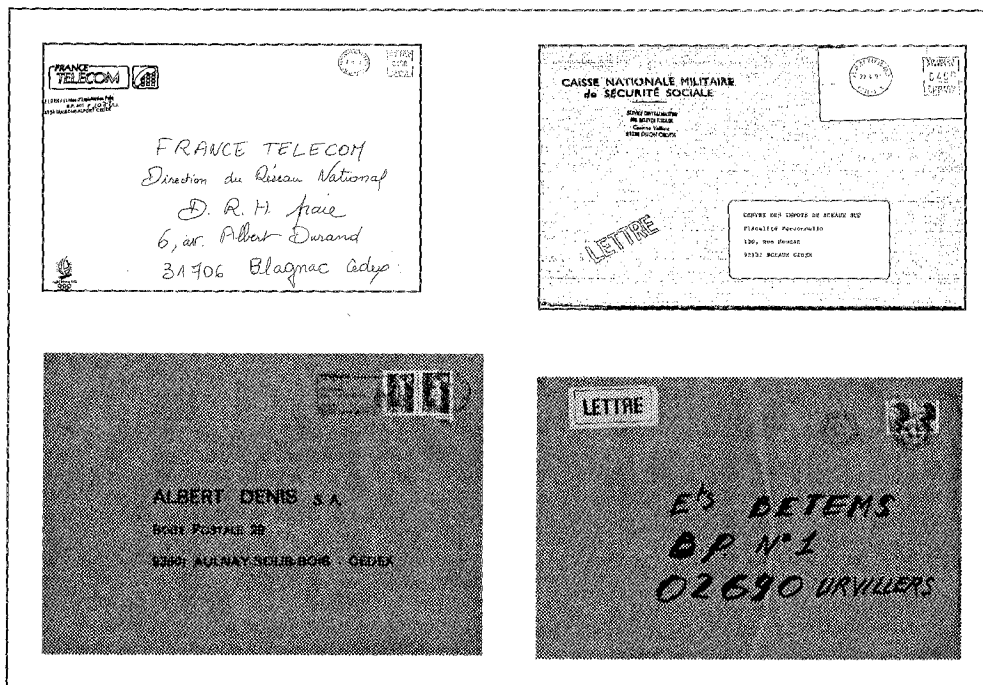


Figure 1. – Exemples d'objets peu chargés.

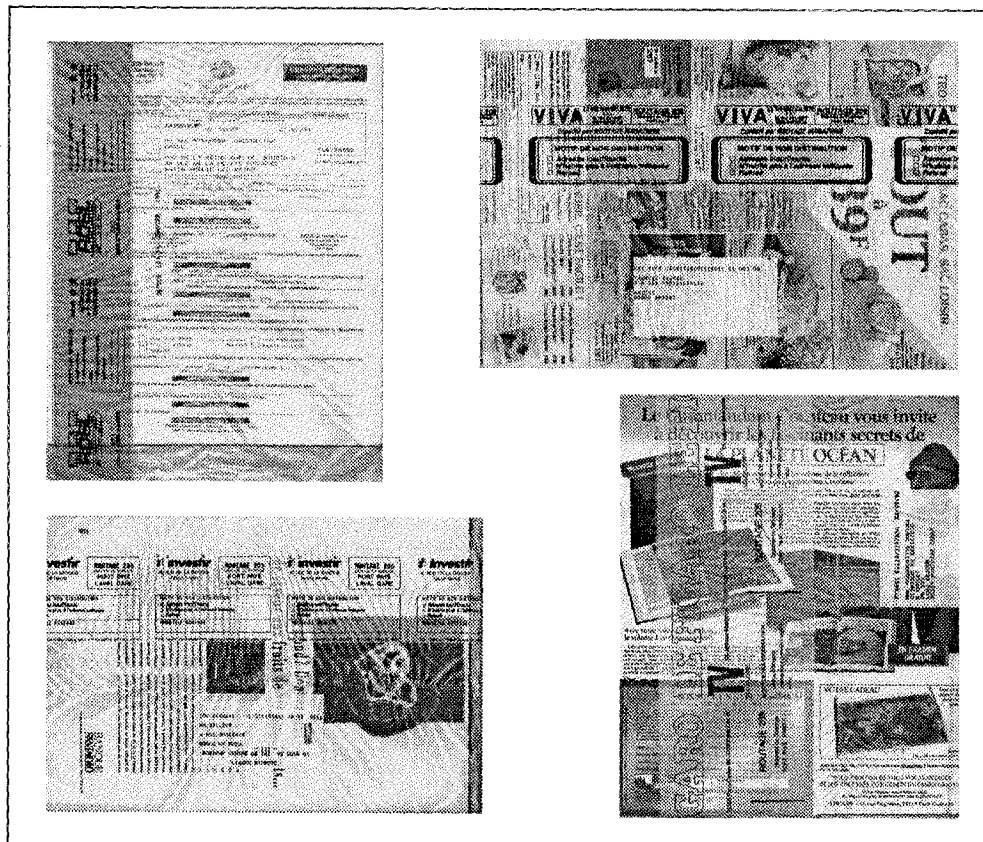


Figure 2. – Exemples d'objets chargés.

### 3.1. SCHEMA GLOBAL DE LA METHODE

Nous présentons sur le schéma de la figure 3 un diagramme des blocs fonctionnels mis en évidence lors de la conception des traitements :

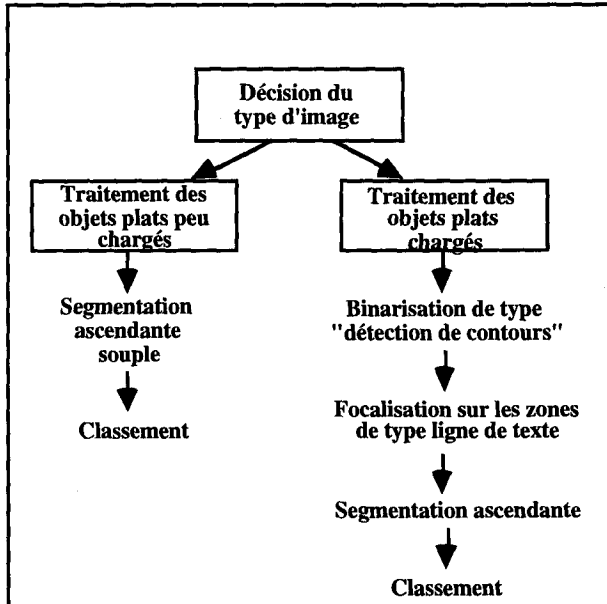


Figure 3. - Diagramme fonctionnel de la méthode.

La décision du type d'image est principalement basée sur la densité de l'image binaire et sur le nombre de composantes connexes. Vis-à-vis des données descriptives disponibles pour chaque image, nous avons pu vérifier un bon accord de notre classifieur de type d'image [ORI, 92]. Les performances obtenues en classification à 3 catégories (images peu, moyennement et très chargées) fournissent un taux d'erreur de l'ordre de 10 % qui proviennent essentiellement de la catégorie moyennement chargée. Elle représente à elle seule 8 % d'erreur. Ces performances sont tout à fait bonnes compte tenu du fait que nous comparons des catégories calculées à des catégories attribuées sur la base d'une appréciation visuelle (subjectivité de l'opérateur humain).

### 3.2. PROBLÉMATIQUE DES OBJETS PEU CHARGÉS ET TRAITEMENTS DÉVELOPPÉS

Les objets peu ou moyennement complexes sont le plus souvent constitués de quelques blocs d'information assez bien différenciés spatialement. Ces objets présentent presque toujours un bloc adresse de destination, un bloc affranchissement et optionnellement un bloc adresse d'expéditeur (adresse de retour). D'autres éléments peuvent apparaître tels que des logos publicitaires ou bien des étiquettes contenant différentes mentions (LETTRE, PLI URGENT, PAR AVION etc...). L'adresse de ces objets peut être

indifféremment manuscrite ou de type machine (dactylographiée, imprimée). Les deux ou trois blocs d'informations principaux respectent des règles de position relative. Par conséquent, les images peu ou moyennement complexes présentent une grande diversité dans les caractéristiques du texte adresse et une disposition relativement structurée de certains blocs d'information.

Devant la grande diversité des types d'écriture de la région d'intérêt et la faible proportion de graphique (densité, proximité), nous avons opté pour une méthode ascendante souple autorisant, entre autres, le découpage de certains regroupements pour gérer les cas de chevauchement des lignes manuscrites. La figure 4 montre le schéma global de la méthode développée pour traiter les objets peu chargés.

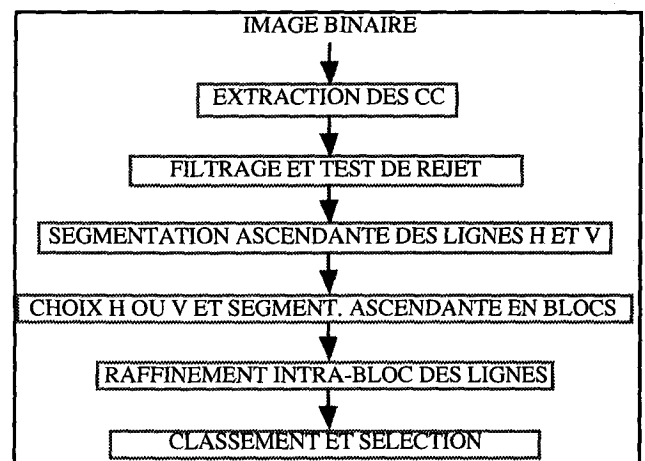


Figure 4. - Schéma global du traitement des objets peu chargés.

La stratégie consiste, comme dans le cas des objets chargés présenté ensuite, à générer une liste de blocs candidats (groupes de lignes spatialement proches et de caractéristiques semblables). Les candidats sont ensuite classés afin de pouvoir retenir le plus vraisemblable. Dans le cas des objets peu chargés, le classement est assez simple puisque généralement, le nombre de candidats est réduit et que la position dans l'image est assez discriminante [ORI, 92].

### 3.3. PROBLÉMATIQUE DES OBJETS CHARGÉS ET TRAITEMENTS DÉVELOPPÉS

Les objets complexes se distinguent par l'absence de structure générique. Les régions de type texte ne sont pas nécessairement majoritaires. On trouve sur ces images des photos, des graphiques, des éléments séparateurs (encadrés, traits pleins ou pointillés), des éléments de texte publicitaire de grande taille (caractères creux ou pleins, parfois en « inverse vidéo »). L'adresse peut se trouver n'importe où dans l'image, cependant, un élément vient réduire la complexité de la situation. En effet, l'adresse sur ce type d'image est presque toujours de type dactylographié ou imprimé. De plus,

les polices utilisées sont de taille moyenne (typiquement entre 10 et 14 points).

La conséquence la plus importante du choix d'une faible résolution (75 dpi) concerne le fait que les composantes connexes issues de zones de texte dactylographié correspondent assez souvent à des groupes de caractères (le faible espace inter-caractères n'est pas suffisant pour donner lieu à un point d'échantillonnage). Les techniques de recherche de composantes alignées sont moins efficaces puisque le nombre de composantes connexes est moindre. Cette situation indique une caractérisation initiale au niveau des groupes de caractères. C'est pourquoi nous amplifions les relations de connexité dans l'image binaire selon la direction du texte, au moyen d'une procédure de pré-segmentation. La figure 5 représente le schéma de principe de la pré-segmentation.

Cette transformation de l'image binaire d'origine a pour objectif de rendre plus compactes les zones de textes répondant aux critères de recherche et de filtrer autant que possible les autres zones. Nous avons développé pour cela une double caractérisation (directe et transverse) des lignes de texte. Tout d'abord la hauteur des lignes recherchées est bornée par la fourchette  $[h_{\min}, h_{\max}]$  (de 5 à 20 pixels). La caractérisation des lignes de texte dans la direction orthogonale au sens de l'écriture (transverse) est basée sur la détection de cette hauteur. Mais lorsqu'on traverse une ligne d'écriture, de multiples « événements » peuvent se produire (on peut rencontrer 0, 1 ou plus d'objets) et il est difficile dans ces conditions d'essayer de détecter la hauteur.

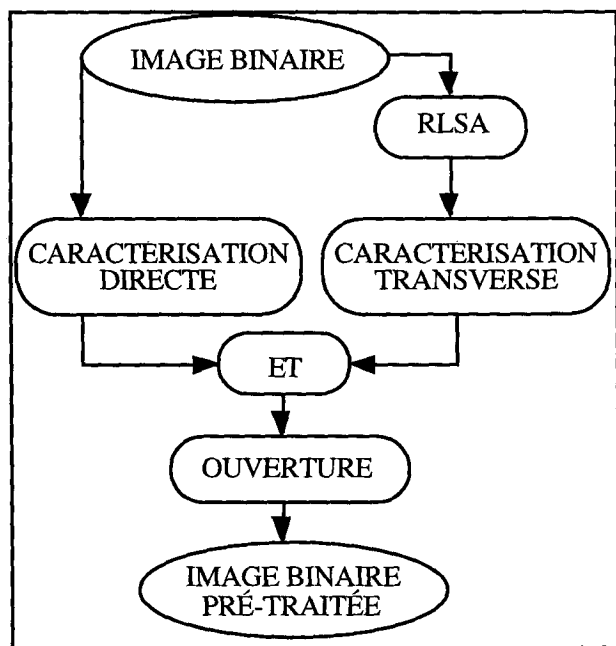


Figure 5. – Principe de la pré-segmentation.

C'est pourquoi nous faisons précéder la détection transverse d'une phase de « bouchage morphologique » au moyen de l'algorithme RLSA [ABE, 81]. L'algorithme est utilisé selon la seule direction du texte avec une distance correspondant environ à la largeur

des plus grands caractères recherchés (environ 10 pixels). Cette préparation étant effectuée, nous pouvons maintenant détecter directement la hauteur au moyen de la distance entre deux transitions fond-objet et objet-fond consécutives. En filtrant ces distances par la fourchette  $[h_{\min}, h_{\max}]$ , nous éliminons les éléments de hauteur différente qui auraient pu être connectés malencontreusement par le bouchage préalable.

Une seconde caractérisation (directe) est menée selon la direction du texte. Cette transformation a également pour objectif de rendre compactes les régions de texte. Le texte recherché présente, lorsqu'on le traverse dans le sens de l'écriture, des paires de transitions fond-objet. Ces distances entre les paires de transitions précédentes peuvent être bornées par une fenêtre  $[d_{\min}, d_{\max}]$  correspondant à la gamme des distances intra-caractères et de l'épaisseur moyenne des traits de caractère pour le texte recherché (environ [2, 17]). De cette façon, un trait plein, même un peu incliné, ne sera pas retenu par cette transformation.

Dans les deux types de caractérisation présentées ci-dessus, les pixels impliqués dans les paires de transitions sont mis à la valeur objet. Ces deux transformations produisent des formes compactes d'allure à peu près rectangulaire au niveau des mots et présentent des effets de filtrage complémentaires [ORI, 92]. Un ET logique entre les deux images transformées combine les différents effets obtenus. L'image résultante présente alors un ensemble bruité de tâches compactes au niveau des mots. Une ouverture morphologique permet de « nettoyer » l'image obtenue. Ceci conclut la phase de pré-segmentation en fournissant une image binaire dans laquelle les mots (et parfois des groupes de mots) sont représentés sous la forme de composantes connexes compactes.

A partir de cette image binaire, nous commençons une analyse des composantes connexes selon une stratégie ascendante guidée par les données. Les composantes connexes sont regroupées selon des critères topologiques en entités de type groupe de mots. Ces nouvelles entités sont caractérisées à partir de mesures sur la région de l'image binaire initiale correspondante (densité de transitions objet-fond/objet-fond, hauteur moyenne, inclinaison, etc) puis regroupées selon des critères topologiques et d'homogénéité de caractéristiques pour former des entités de type bloc de lignes. Par ailleurs, nous avons contraint les étapes de pré-segmentation et de regroupement ascendant avec l'utilisation d'un « masque d'obstacles ». Devant la très grande complexité de certaines images, il est apparu qu'il était possible d'identifier dans l'image binaire initiale des composantes qui ne sont certainement pas des éléments de texte. En considérant ces composantes comme des obstacles à ne pas franchir (pour le RLSA) ou à ne pas rencontrer (pour les regroupements ascendants) nous avons amélioré nettement la méthode au prix, il est vrai, de plus lourds calculs. Les performances seront analysées dans la section suivante. On peut indiquer cependant dès maintenant que les gains en performances sont de l'ordre de 10 %.

Voici (figure 6) le schéma de principe de génération du masque d'obstacles à partir de l'image binaire :

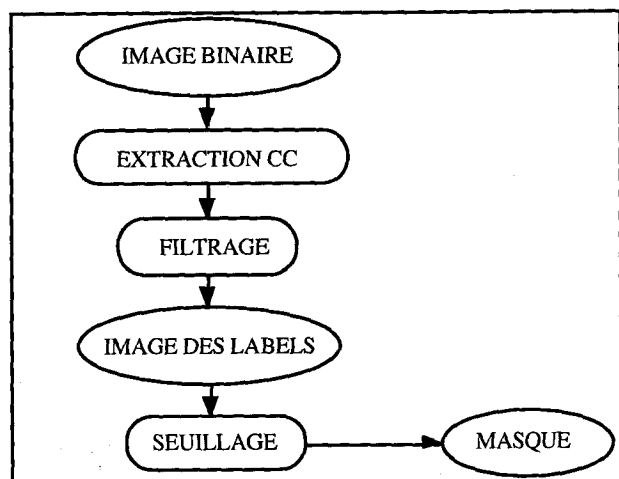


Figure 6. - Génération du masque d'obstacles.

Ce masque intervient lors de la phase de segmentation ascendante, mais il intervient également au niveau de la pré-segmentation. Aussi, avons nous défini une version ternaire de l'algorithme RLSA, dont le principe est illustré sur la figure 7 :

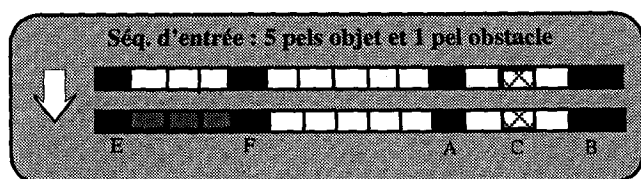


Figure 7. - RLSA contraint par des pixels obstacles, avec un seuil de bouchage de 3 pixel.

Le masque d'obstacles permet, en isolant des régions, de limiter les erreurs de sous-segmentation.

Cette analyse ascendante conduit à l'obtention d'une liste de groupes de lignes qui sont considérés comme des BAD candidats. Un vecteur de caractéristiques est mesuré sur chacun. Ce vecteur contient en particulier une « intégration » des caractéristiques des lignes proportionnellement à la longueur des lignes (la fiabilité de ces grandeurs croît avec la longueur des lignes). Ce vecteur permet alors de calculer une mesure de vraisemblance pour chaque candidat.

## 4. Évaluation et optimisation des traitements

En l'absence de la phase de lecture d'adresse qui aurait permis d'évaluer très simplement les traitements (par comparaison du code postal sélectionné et du code postal réel, par exemple), nous avons dû définir des critères d'appréciation de la qualité de la phase de localisation. Ceci permet de définir une procédure d'évaluation automatique.

Ce type d'évaluation ne permet pas d'atteindre la finesse d'une appréciation visuelle, cependant cela permet d'obtenir des mesures systématiques. Ainsi sur un lot d'images donné, on pourra correctement comparer les mesures obtenues soit à partir de réglages différents soit à partir d'options de traitement différentes. Ceci est très important et nous a permis de vérifier la validité du schéma de pré-segmentation développé et de rechercher un jeu de réglage optimal dans le cas des objets chargés.

### 4.1. ÉVALUATION DU TRAITEMENT DES OBJETS PEU CHARGÉS

La méthode développée pour le traitement des objets peu chargés est basée sur une analyse ascendante relativement simple des composantes connexes. Pour cette catégorie d'images, nous disposons d'une base d'images très importante (plus de 15 000 images) et nous avons réalisé une évaluation automatique en nous basant sur la correspondance entre les rectangles encadrant les deux lignes du bas du bloc adresse sélectionné et les rectangles encadrant deux lignes du bas du vrai bloc adresse (démarche proche du fonctionnement réel d'un dispositif de lecture d'adresse). Au moyen de deux seuils permettant d'apprécier les notions de « bonne correspondance » et « d'intersection suffisante » [ORI, 92], nous avons distingué trois situations : localisation correcte, localisation incertaine et localisation incorrecte. La base d'images a été séparée en quatre catégories. La catégorie A (resp. B) représente les objets plats dont l'adresse est dactylographiée et continue (resp. discontinue). La catégorie C regroupe les objets plats avec adresse imprimée (offset, grande taille), la catégorie D correspond aux objets plats avec une adresse manuscrite. Le tableau suivant présente les résultats obtenus lors de cette évaluation à grande échelle.

$NBI$  : Nb images

$t1$  : taux de rejet = Nb images rejetées/ $NBI$

$t2$  : taux bon = Nb images bonnes/ $(NBI - Nb images rejetées)$

$t3$  : taux incertain = Nb images incertaines/ $(NBI - Nb images rejetées)$

$t4$  : taux mauvais = Nb images mauvaises/ $(NBI - Nb images rejetées)$

$t5$  : taux mini = taux bon

$t6$  : taux maxi = taux bon + taux incertain

lot	NBI	t1	t2	t3	t4	t5	t6
A	3389	3 %	81 %	12 %	7 %	81 %	93 %
B	4786	4 %	69 %	22 %	9 %	69 %	91 %
C	1209	6 %	74 %	16 %	10 %	74 %	90 %
D	5688	3 %	69 %	24 %	7 %	69 %	93 %
$\Sigma$	15072	5 %	72 %	20 %	8 %	72 %	92 %

Tableau 1. - Résultat du traitement sur les objets peu chargés, base de 15 000 images.

Globalement, entre 70 % et 90 % des images sont traitables par cette méthode générique et évolutive de segmentation ascendante, et ce pour des textes de caractéristiques très variables. La taille très importante de la base d'images utilisée pour mener cette évaluation permet d'être confiant sur la validité des résultats obtenus.

## 4.2. ÉVALUATION DU TRAITEMENT DES OBJETS CHARGÉS

Pour les objets chargés, la méthode mise en œuvre est plus complexe et nous avons voulu comparer plusieurs versions (options) au niveau de la pré-segmentation ainsi que plusieurs réglages de paramètres intervenants au cours des différentes étapes du traitement.

Nous rappelons le principe de l'évaluation automatique dans le cas de ces objets, ainsi que les grandeurs caractéristiques utilisées. La comparaison d'un bloc candidat avec le BAD réel fait appel à un coût intégrant deux mesures qui estiment, en terme de surface, la quantité relative « d'information perdue » (*ipr*) et la quantité relative « d'information ajoutée » (*iar*). Ceci permet d'estimer les sur-segmentations (*ipr*) et les sous-segmentations (*iar*). Un autre élément d'évaluation consiste à vérifier si le bloc code postal est présent dans le bloc sélectionné (on mesure aussi une quantité relative d'information perdue (*ipr\_cp*)). La comparaison est basée sur des intersections de surfaces. Les notations (*Scand*, *Scomm1*, *Scomm2*, *Sréf1* et *Sréf2*) représentent des surfaces et sont définies sur la figure 8.

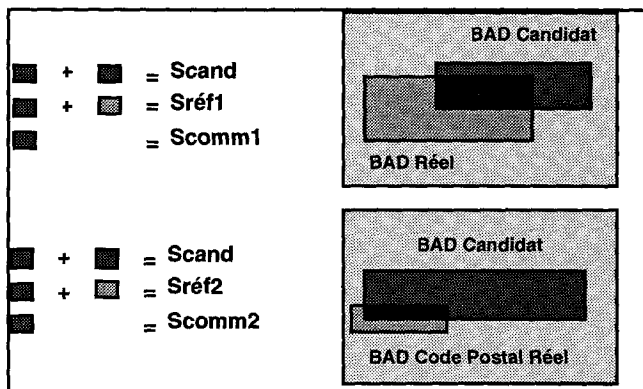


Figure 8. - Grandeurs utilisées pour établir la comparaison entre un candidat et le BAD réel.

Pour comparer chaque bloc candidat avec le bloc de référence, avec les notations de la figure 8, nous définissons les grandeurs caractéristiques suivantes :

$$iar = (Scand - Scomm1) / Sréf1$$

$$ipr = Sréf1 - Scomm1$$

$$ipr\_cp = (Sréf2 - Scomm2) / Sréf2$$

$$coût = iar + ipr$$

Pour chaque image, nous pouvons définir un candidat de coût minimum noté  $C_{min}$ . Nous notons par ailleurs  $C1$  et  $C2$  les

deux meilleurs candidats d'après la vraisemblance qui leur a été attribuée. En rapport avec les grandeurs *iar*, *ipr* et *ipr\_cp* précédentes, nous introduisons trois seuils *tiar*, *tivr* et *tivr\_cp*. Pour chaque image, nous décrivons la qualité de la localisation du bloc adresse au moyen des quatre booléens  $B_{cmin}$ ,  $B_{cmin'}$ ,  $B_{c1}$  et  $B_{c2}$  définis de la manière suivante :

$$B_{cmin} = (iar(C_{min}) < tiar \text{ ET } ipr(C_{min}) < tivr)$$

$$B_{cmin'} = (ipr\_cp(C_{min}) < tivr\_cp)$$

$$B_{c1} = (B_{cmin} = 1 \text{ ET } C1 = C_{min})$$

$$B_{c2} = (B_{cmin} = 1 \text{ ET } C2 = C_{min} \text{ ET } C2 \neq C1)$$

Utilisation de la procédure :

Cette procédure automatique nous a permis tout d'abord de valider le schéma de pré-segmentation. La figure 9 suivante représente ce schéma et les différentes options que nous avons mises en compétition.

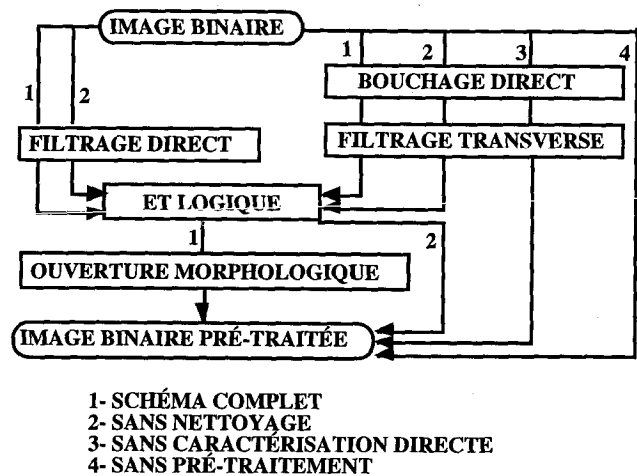


Figure 9. - Options de présegmentation.

La variante du schéma complet est nommée *V1*, la variante sans nettoyage morphologique est nommée *V2*, la variante *V3* n'utilise pas de caractérisation directe et la variante *V4* n'utilise aucun pré-traitement. Le tableau 2 (page suivante) contient les mesures relevées pour chaque option. Ces mesures ont été obtenues en utilisant la moitié environ d'un lot de 135 images d'objets chargés noté *lot\_map* (*lot* servant aux réglages et à la mise au point de la méthode). Nous nous sommes limités à 70 images pour le lot d'apprentissage car la réalisation d'essais de différents réglages pour différents paramètres est une opération très coûteuse en temps de calcul. Les grandeurs *m<sub>iar</sub>*, *m<sub>ipr</sub>* et *s<sub>iar</sub>*, *s<sub>ipr</sub>* représentent respectivement les valeurs moyennes et les écarts types de *iar* et *ipr*. La grandeur *T<sub>m</sub>* représente le temps d'exécution moyen, en seconde. Les seuils *t<sub>iar</sub>*, *t<sub>ivr</sub>* et *t<sub>ivr\_cp</sub>* précédents ont été positionnés (de façon moyenne, à partir du *lot\_map*) aux valeurs suivantes : 2.0, 0.3 et 0.25.

On remarque tout d'abord que la variante *V4*, qui ne fait pas appel à la pré-segmentation, obtient des performances nettement inférieures à celles obtenues avec la variante complète (*V1*). On peut également remarquer que les variantes *V2* et *V3* sont aussi



Options	Bcmin	Bc1	Bc2	Bcmin'	miar	siar	mipr	sipr	Tm
V1	80.0	58.6	11.4	84.3	0.67	0.88	0.16	0.23	22.3
V2	74.3	54.3	7.1	85.7	0.73	0.91	0.15	0.20	23.8
V3	75.7	52.9	10.0	85.7	0.79	0.99	0.16	0.22	25.4
V4	67.1	55.7	7.1	71.4	0.72	1.00	0.22	0.27	76.8

Tableau 2. – Comparaison des options de pré-segmentation (70 images).

en assez net retrait vis-à-vis de la variante V1. Par ailleurs, on a pu mesurer que, en moyenne, l'utilisation de la pré-segmentation permet de réduire d'un facteur trois le nombre de composantes connexes. Le temps d'exécution très important de la variante V4 illustre cet effet, en raison d'une plus grande quantité de composantes connexes à traiter. Ces remarques permettent de valider le schéma de pré-segmentation qui a été conçu. Ainsi, le rehaussement et le filtrage obtenu permettent de sélectionner très rapidement les informations les plus pertinentes.

La deuxième utilisation de la procédure d'évaluation automatique concerne l'analyse de la sensibilité des paramètres (étude de 8 paramètres parmi les plus significatifs). Cette étude a permis de montrer que, pour la plupart des paramètres, une petite variation de leur valeur n'entraîne pas de modifications importantes des performances. Ceci indique une bonne robustesse globale de la méthode. Nous avons, en particulier, mis en évidence l'importance du facteur de regroupement des lignes en bloc, intervenant dans la phase de segmentation ascendante. De tous les paramètres étudiés, c'est probablement le plus sensible. Dans sa version initiale, le système ne pouvait pas supporter une valeur trop élevée de ce paramètre au risque de générer trop d'erreurs de sous-segmentations. L'introduction d'un masque d'obstacles (cf section 3) a permis de presque doubler cette valeur ce qui rend possible une bonne segmentation des situations présentant des distances inter-lignes variées (ce qui est assez fréquent) tout en limitant les erreurs de sous-segmentation.

Les mesures obtenues après optimisation du réglage des paramètres, sur la moitié du *lot\_map* sont représentées dans le tableau 3 :

Cmin	C1	C2	Cmin'	Tm
85.7	62.9	10.0	88.6	22.6
miar	siar	mipr	sipr	
0.80	0.79	0.12	0.19	

Tableau 3. – Mesures après optimisation du réglage, demi lot-map.

Ces performances sont nettement meilleures que celles obtenues avant optimisation du réglage (Cf tableau 2, variante V1). Pour vérifier la stabilité des résultats, nous avons étendu les mesures à l'ensemble du *lot\_map*. Nous avons constaté un retrait des performances (prévisible) dans des proportions très acceptables (tableau 4).

Bcmin	Bc1	Bc2	Bcmin'	Tm
81.5	63.7	7.4	86.7	24.4
miar	siar	mipr	sipr	
0.80	0.92	0.14	0.19	

Tableau 4. – Mesures après optimisation du réglage étendues au lot-map entier (135 images).

Grâce à ce travail, nous avons observé une amélioration des performances de la segmentation en examinant la sensibilité des paramètres et en choisissant un jeu de paramètres optimisé. La procédure automatique nous a également fourni la possibilité de valider (par comparaison de plusieurs options) le schéma de pré-segmentation.

### 4.3. OPTIMISATION DE LA PHASE DE CLASSEMENT POUR LES OBJETS CHARGÉS

Le travail précédent constitue un premier type d'optimisation. Une deuxième phase importante du traitement pouvant faire l'objet d'une optimisation concerne le classement des candidats sur la base du vecteur de caractéristiques extrait pour chacun des blocs de texte. Cette étape, dans sa version de base, intègre de façon empirique chaque caractéristique interprétée et normalisée. On peut alors rechercher une séquence de pondération optimale qui prendrait mieux en compte l'importance relative de chaque caractéristique. Ceci peut être réalisé au moyen d'un filtre adaptatif, comme illustré sur la figure 10.

Avec ce type de filtre, on associe usuellement l'algorithme des moindres carrés moyens [BEL, 89]. De façon classique, et avec les notations de la figure précédente, le critère à optimiser s'écrit :

$$J = \sum_{p=1}^N J_p \quad \text{avec} \quad J_p = \frac{1}{2} E(p)^2$$

où  $p$  est l'indice d'une image servant à l'apprentissage.

Pour chaque image analysée lors de cette phase d'optimisation, le signal d'erreur est élevé au carré et divisé par deux. On obtient alors le critère à optimiser. Le poids  $W_j$  de chaque caractéristique est décrétement d'un terme proportionnel à la dérivée partielle du critère vis-à-vis de ce poids. La formule de mise à jour des poids



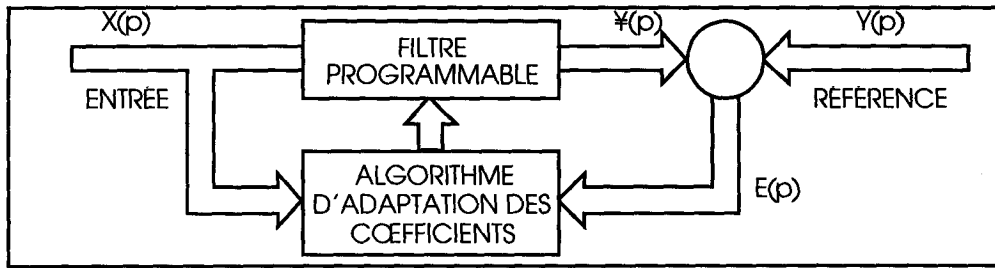


Figure 10. – Principe d'un filtre adaptatif.

est donc :

$$W_j^{k+1} = W_j^k - \varepsilon \times \frac{\partial}{\partial W_j} (J_p)$$

où  $\varepsilon$  est un pas d'ajustement ( $\varepsilon > 0$  et  $\varepsilon \ll 1$ )

Après optimisation du classement des candidats, voici les mesures réalisées sur le *lot\_map* précédent :

Bcmin	Bc1	Bc2	Bcmin'	Tm
81.5	66.7	11.1	86.7	20.7
miar	siar	mipr	sipr	
0.80	0.92	0.14	0.19	

Tableau 5. – Mesures après optimisation du classement, *lot-map*.

Vis-à-vis de la situation avant optimisation de la phase de classement, on peut remarquer que l'écart entre  $(Bc1 + Bc2)$  et  $Bcmin$  est passé de 10.4 % à 3.7 %, ce qui démontre l'intérêt de cette procédure d'optimisation.

Nous devons maintenant vérifier la stabilité (et donc la fiabilité) des mesures de performance en appliquant la méthode à un autre lot d'images, le *lot\_test*. Les images de ce lot n'ont jamais été observées et utilisées autrement que par l'opérateur de saisie lors de la génération de la base de données associée à la base d'image. Ce lot représente des images sélectionnées au hasard parmi les objets plats chargés (129 images). Nous rappelons que les images collectées l'ont été sur la base des difficultés a priori qu'elles présentaient. Ces images (*lot\_map* et *lot\_test*) ne représentent donc pas un ensemble type d'objets plats mais plutôt une sélection des situations a priori les plus délicates. Voici les mesures réalisées à partir du *lot\_test* :

Bcmin	Bc1	Bc2	Bcmin'	Tm
82.2	62.8	15.5	84.5	20.1
miar	siar	mipr	sipr	
0.73	0.79	0.15	0.21	

Tableau 6. – Mesures comparatives sur le *lot-test* (129 images).

Nous pouvons remarquer que, globalement les performances sont très proches du niveau atteint pour le *lot\_map*. Cette stabilité des

performances indique que les principes mis en œuvre ainsi que les grandeurs mesurées et interprétées ne sont pas seulement valables sur un lot d'entraînement et de mise au point, mais se révèlent pertinentes également sur une population d'images chargées non analysées au préalable. Par conséquent, cette dernière expérience valide de manière très satisfaisante la méthode développée.

#### 4.4. ÉVALUATION VISUELLE DU TRAITEMENT DES OBJETS CHARGÉS

L'évaluation automatique présentée précédemment contient des imprécisions et des approximations. Cependant, ces erreurs n'étaient pas réellement gênantes pour mener une analyse comparative entre différentes options ou réglages sur un même lot d'images. Par contre, pour avoir une bonne appréciation de la méthode (performances au sens plus absolu) nous devons recourir à une évaluation visuelle.

Pour cela, nous avons défini plusieurs situations (codes A à E) en fonction de la qualité de la segmentation et de la phase de classement. L'évaluation est, ici encore, basée sur l'analyse des deux lignes du bas du bloc adresse.

Le code A correspond aux cas où le code postal réel se trouve au début d'une des deux lignes du bas du candidat le plus proche du BAD réel.

Le code B correspond aux cas de type A mais pour lesquels il y a erreur d'orientation (ainsi les lignes du bas sont en haut).

Le code C (resp. C') correspond aux cas où le code postal réel se trouve sur la troisième (resp. la quatrième) ligne à partir du bas du candidat le plus proche du BAD réel.

Les codes D et E sont indépendants des rangs de classement et représentent respectivement les cas où la ligne code postal est absente du candidat le plus proche du BAD réel (c'est une sur-segmentation très gênante) et les cas où aucun BAD candidat ne correspond au BAD réel (c'est un échec total de la segmentation).

Le tableau 7 représente la répartition des différentes situations rencontrées, sur les images du *lot\_map*, pour la segmentation (codes A à C') en fonction du rang de classement du candidat le plus proche du BAD réel (R1 à R6 : rangs 1 à 6 et R7+ : rangs 7 et plus).

En considérant que certaines erreurs (de segmentation et de classement) sont plus ou moins aisément récupérables, nous suggérons le regroupement suivant en trois catégories (succès, semi-échec et échec) :

1- Succès :  $A(R1)+A(R2)+B(R1)= 80.7 \%$

2- Semi-échec :  $A(R3)+A(R4)+C(R1)+C(R2)+C'(R1)=10.4 \%$

3- Échec :  $A(R7+)+D(R1)+E(R1)= 8.9 \%$

Compte tenu de la complexité des images à traiter, ces résultats sont d'un niveau très convenable.

code	R1	R2	R3	R4	R7+
A	88 65.2%	16 11.8%	4 3.0%	3 2.2%	3 2.2%
B	5 3.7%				
C	5 3.7%	1 0.7%			
C'	1 0.7%				
D	4 3.0%				
E	5 3.7%				

Tableau 7. – Évaluation visuelle sur 135 images chargées (lot-map).

## 5. Conclusion

Le tri des grands objets postaux nécessite une phase de localisation de l'adresse suivie d'une phase de reconnaissance de caractères. Or notre étude n'englobe pas la phase de lecture. Nous avons donc dû définir des caractéristiques permettant d'apprécier certains aspects de la qualité de la localisation. Ceci rend la tâche d'évaluation plus complexe mais fournit aussi l'opportunité de réaliser une évaluation automatique sur la base de critères qui décrivent la seule phase de localisation. En effet, une évaluation au niveau d'un système complet (localisation et lecture) ne permettrait pas de bien savoir attribuer à l'une ou l'autre des deux étapes la responsabilité d'un échec.

Le fait de disposer d'une procédure automatique d'évaluation, même si elle comporte un certain nombre d'imprécisions, permet de comparer de façon objective différentes options ou différents réglages sur un même lot d'images.

L'évaluation menée montre plusieurs choses :

1. Il est possible de détecter des objets peu chargés de façon assez fiable. Le taux de confusion est aisément maintenu à de faibles valeurs.

2. Un traitement assez générique, de type analyse des composantes connexes, permet de s'adapter à la grande diversité des types d'écriture sur les objets peu chargés. Un tel traitement conduit à des performances d'un très bon niveau (entre 72 % et 92 %) à un coût de calcul très réduit. La marge d'amélioration de cette méthode semble assez importante.

3. Moyennant une analyse plus complexe, les objets plats très chargés peuvent être traités automatiquement avec des performances assez correctes (au moins 80 % et très probablement entre 85 % et 90 %). Les situations mal traitées par cette méthode représentent un ensemble de situations atypiques qu'il semble difficile de prendre en considération sans alourdir de façon très importante la méthode.

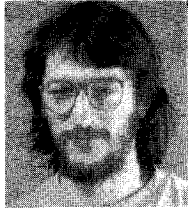
## BIBLIOGRAPHIE

- [ABE, 81] L. Abele, F. Wahl and W. Scherl « Procedures for an Automatic Segmentation of Text, Graphic and Half Tone Regions in Documents », *Proc. of Scandinavian Conf. on Image Analysis*, Vol. 2, 1981, pp. 177-182.
- [ANT, 88] T. Antognini, L. Turnbaugh, « An expert System for Location of Destination Address in Mail of Arbitrary Complexity », *USPS Advanc. Techno. Conf. #3*, 1988, pp. 279-292.
- [BEL, 89] M. Bellanger, « Analyse des signaux et filtrage numérique adaptatif », *CNET-ENST, Coll. Techn. et scient. des Télécommunications*, Ed. Masson, 1989.
- [BIX, 88] J.P. Bixler, « Tracking Text in Mixed-mode Document », *ACM Conf. on Document Processing Systems*, 1988, pp. 177-185.
- [DER, 90] D. Derrien Peden, « Analyse des structures de documents : une approche objet », *Thèse de l'Université de Rennes-I*, 1990.
- [FLE, 88] L.A. Fletcher and R. Kasturi, « A Robust Algorithm for Text String Separation from Mixed Text/Graphics Images », *IEEE PAMI*, Vol. 10, n°6, 1988, pp. 910-918.
- [HIG, 86] J. Higashino et al, « A knowledge based segmentation method for document understanding », *ICPR*, Vol. 8, 1986, pp. 745-748.
- [KUB, 83] K. Kubota, O. Iwaki, H. Arakawa, « Image Segmentation Techniques for Document Processing », *Int. Conf. on Text Processing with a Large Character Set*, 1983, pp. 73-78.
- [MEY, 88] E. Meynieux, « Un Système d'Analyse de Documents Multimédia », *Thèse de doctorat de l'École Centrale de PARIS*, 1988.
- [ORI, 92] J.C. Oriot, « Analyse d'images de documents à structures variées, application à la localisation du bloc adresse sur les objets postaux », *Thèse de l'Université de Nantes*, 1992.
- [SRI, 89] S.N. Srihari, « Features extraction for locating address blocks on mail pieces », in *From Pixels to Features*, JC Simon, Elsevier Science Publisher, 1989, pp. 261-273.
- [WAN, 89] D. Wang and S.N. Srihari, « Classification of Newspaper Image Blocks Using Texture Analysis », *Computer Vision Graphics and Image Processing*, Vol. 47, 1989, pp. 327-352.
- [WON, 82] K.Y. Wong, R.G. Casey and F.M. Wahl, « Document Analysis System », *IBM Journal of Research and Development*, Vol. 26, n°6, 1982, pp. 647-656.

Manuscrit reçu le 6 Septembre 1994.

## LES AUTEURS

### Jean-Claude ORIOT



Jean-Claude Oriot est né en 1964. Il est titulaire d'un DEA d'Automatique et Informatique Industrielle de l'Université de Nantes (1989) et d'un Doctorat de l'Université de Nantes et de l'École Centrale de Nantes en Sciences pour l'Ingénieur dans la spécialité Automatique et Informatique Appliquée (1992). Sa thèse de doctorat s'est effectuée sur un sujet lié à l'analyse d'images de documents à structures variées avec application à la localisation du bloc adresse sur les objets postaux.

### Dominique BARBA



Dominique Barba est né en France en 1944. Il est titulaire d'un doctorat de 3ème cycle en Télécommunications de l'Université de RENNES I en 1969 et un doctorat es Sciences Mathématiques (mention Informatique) de l'Université de PARIS VI sur un sujet lié au traitement numérique des images avec critère psychovisuel de qualité. Il est actuellement professeur à l'IRESTE (Université de Nantes) et responsable d'une équipe de recherche en Analyse et Traitement des images. Ses centres d'intérêt concernent le traitement numérique des images fixes et animées, spécialement en vue de l'analyse ou du codage des images.

### Michel GILLOUX



Michel Gilloux est ancien élève de l'École Polytechnique et de l'École Nationale Supérieure des Télécommunications. De 1982 à 1991, il a été ingénieur de recherche au Centre National d'Études des Télécommunications où ses travaux ont porté sur les systèmes experts, la reconnaissance de la parole et l'informatique linguistique. Depuis 1991, il est responsable du département Reconnaissance-Vision-Architecture au Service de Recherche Technique de la Poste où il dirige différents projets concernant les applications postales de la lecture automatique.