Structuration de documents par repérage de zones d'intérêt

Document segmentation by interest areas detection

par Véronique EGLIN, Stéphane BRES et Hubert EMPTOZ

Laboratoire de Reconnaissance de Formes et Vision RFV
INSA de Lyon
20, avenue Albert Einstein 69621 VILLEURBANNE CEDEX
Phone : (33) 04 72 43 60 54 Fax : (33) 04 72 43 80 97
E-mail : eglin@rfv.insa-lyon.fr

résumé et mots clés

Cette étude présente une nouvelle approche de la structuration de documents imprimés basée sur l'exploitation de la dynamique du regard dans le repérage de l'information. Le système qui a été mis en place nous permet d'obtenir une représentation du document segmenté en faisant appel à des procédures d'extraction de primitives géométriques simples (traitements de bas niveau) relevant de la prise en compte de certains comportements caractéristiques chez l'homme dans l'extraction d'information. Il utilise une série de représentations de type multirésolution du document où la nature du sous-échantillonnage est une fonction de la position du regard. Cette approche est basée sur la recherche des zones de focalisation de l'attention permettant de conserver une description précise des éléments dans les zones de fixation, tout en résumant les régions présentant un « intérêt » moindre. La simulation du parcours de l'œil sur le document que nous avons retenue traduit la segmentation que ferait un lecteur qui aborde le document sans a priori sur ce qu'il veut trouver. Pour cela, nous nous sommes inspirés d'une stratégie exploratoire particulière : *le survol*. Celle-ci s'appuie essentiellement sur l'aspect visuel du document, c'est-à-dire sur les caractéristiques visuelles de bas niveau de l'image. Elle permet en outre une perception équilibrée des données en privilégiant l'organisation globale du document. La technique mise en œuvre s'appuie sur un partitionnement *évolutif* de l'espace, en zones centrées aux points de fixation successifs. C'est sur la base de ce partitionnement, que la description des différentes régions *ciblées* du document évolue et converge vers une représentation segmentée.

Structuration de documents, traitements de bas niveau, perception visuelle, multirésolution, intégration de données visuelles.

abstract and key words

This paper presents a new approach of document structuring by the description of a foveated vision system implied in extracting visual and eye-catching information of a document. The simulation system is based on psycho-perceptive rules for visual data capturing. It allows us to obtain a representation of segmented document by using simple low-level processing. The low-level process is based on a *visual integrative memory* which displays the unequal importance of information in the visual field. The resulting segmentation enhances the fact that the access of information is directly linked to the search of attractive areas.
The technical approach of the segmentation (using a *space-variant geometry* and a *multiresolution process*) lays a sound basis for elaborating the kinetic of the ocular displacement on a document. It provides not only a document representation in blocks, but shows a unified view corresponding to the integration of time-variant representations of the same visual field. The resulting blocks (text, graphs, image) are determined and localized all the better, such that the number of fixation points increases and yields a more complete and detailed description of components.

Document structuring, low-level processing, visual perception, multiresolution, visual data integration.

1. introduction : présentation des objectifs

La détermination de la structure des documents composites en blocs homogènes (photographies, graphiques, textes, sous blocs écrits avec des polices différentes...) est un problème très complexe auquel on a souvent tenté de donner des solutions dans des cas bien particuliers (adresses postales, partitions musicales, plans cadastraux...). L'organisation des constituants en blocs séparables pour la description du document constitue la structure *physique*. La *reconnaissance* de cette organisation et l'étiquetage des blocs en diverses catégories constituent la phase d'*identification* de la structure. Elle est connue sous le nom de structuration *logique* du document. Le travail que nous présentons est une partie d'un grand projet de structuration fonctionnelle des documents à la frontière entre les deux types d'analyse physique et logique. La partie qui est présentée dans cet article ne traite que de la partie physique.

De manière générale, pour obtenir la segmentation physique du document, on a recours à des méthodes, généralement très spécialisées, traitant l'image uniformément sur toute sa surface. Jusqu'ici, les réponses qui ont été proposées correspondent à des besoins spécifiques, notamment en analyse de documents administratifs, de chèques, ou de formulaires. Nous pouvons citer quelques travaux qui retracent cette problématique : [Baird92], [O'Gorman93], [Watanabe]. Ces réponses conduisent généralement à la définition de modèles descriptifs des documents (pour une catégorie donnée). On constate que l'intérêt actuel se déporte davantage vers de nouvelles approches, où l'on cherche à s'affranchir de ces modèles en introduisant notamment des notions de perception visuelle humaine, [Déforges], [Likforman], [Ogier]. La question que l'on se pose alors est : « *Comment introduire l'homme dans le processus de structuration automatique de documents*. Une réponse possible, celle que nous avons choisie, consiste à proposer une méthodologie de structuration automatique de documents s'inspirant du mode d'exploration humain. Plus précisément encore, nous avons choisi parmi les différentes stratégies d'exploration visuelle, une stratégie particulière : le survol.

Il est important de rappeler que la segmentation naturelle, faite par l'homme, dépend implicitement et pour une grande part d'un objectif de recherche ou d'une consigne qui aurait été donnée (recherche d'une information particulière...). Il paraît donc assez naturel de s'intéresser de plus près à une information *pertinente* pour l'application, en privilégiant certaines zones informatives du document, et en cherchant avec un « œil attentif » à localiser les composantes qui présentent un intérêt par rapport à l'objectif ou à la consigne. Il en est de même lors de la lecture d'un texte ou du parcours d'un document où l'on cherche à créer un lien continu entre la structuration physique des éléments et le sens de leur organisation. En ce sens, le document est plus qu'une simple image de pixels que l'on pourrait traiter indépendamment du message que l'auteur a voulu faire passer au lecteur. Il faut ainsi pouvoir prendre en compte la présence de l'homme aux différents stades du cycle du vie du document (de sa conception à sa lecture), où l'information de *fond* liée au message que l'auteur veut transmettre s'exprime indirectement par une mise en *forme* particulière des données. Certains auteurs, tels que Nagy, Doermann, Tang et Suen dans [Nagy], [Doermann], [Tang] ont proposé de formaliser ce contexte d'étude en retraduisant l'intervention de l'homme aux différents stades de vie du document.

Sur la base de l'analyse des mécanismes de capture d'informations chez l'homme, nous avons cherché à obtenir une description du document directement liée aux déplacements successifs de l'attention, mettant l'accent sur les zones informatives et attractives. Nous avons choisi de valoriser la stratégie exploratoire la plus générale et qui nous paraissait la première à être exploitée. Il s'agit du *survol* du document qui correspond à un parcours rapide et complet des données. Ce choix était motivé par nos objectifs de segmentation, où nous souhaitions mettre en évidence l'organisation complète du document, étroitement liée à une perception équilibrée de l'ensemble des données, sans utiliser de connaissances a priori sur la mise en forme ou le contenu du document. Il s'agit d'une perception sans intention a priori pour guider la recherche d'informations. La recherche des régions d'intérêt est décrite par le trajet oculaire du regard sur le document issu de la simulation de ce que nous pensons correspondre au mieux au comportement oculomoteur de l'homme dans un contexte de survol. Le système que nous avons mis en place cherche donc à imiter ce survol. Il permet d'établir des liens entre les différentes zones d'un document sur lesquelles, à un instant donné, le regard pourrait être attiré. L'image est donc perçue à des résolutions variables où la région fovéale (correspondant au foyer d'attention) est privilégiée. Nous présentons dans cet article une proposition de segmentation basée sur une représentation multirésolution du document et sur l'introduction d'une mémoire simulant l'intégration des données perçues. Nous aborderons dans les perspectives la problématique liée à la prise en compte des connaissances requises pour la reconnaissance des éléments.

2. les différentes approches de la structuration de document

Dans le domaine de l'analyse de documents, nous pouvons identifier deux types de recherches : l'analyse de composition et l'interprétation du document. Ces deux systèmes de traitements per-

mettent de faire la distinction entre une information *physique* (correspondant aux objets physiques présents dans le document) et une information *logique* (liée à l'interprétation de l'organisation des objets du document). Le premier niveau de données accessibles au système d'analyse est la structure physique du document. Il concerne la répartition spatiale de l'information du document. La structure logique se rapporte au sens de cette organisation. La connaissance de la structure physique permet de déduire la structure logique si les règles de présentation et de composition sont claires et connues.

2.1. les techniques traditionnelles d'analyse de la structure physique

Il existe à l'heure actuelle essentiellement deux familles de traitements dédiées à la recherche de telles structures : les méthodes dites « ascendantes » et « descendantes ». Les premières permettent d'extraire des blocs de textes par la fusion de petites composantes (généralement des composantes connexes) jusqu'à obtenir des blocs plus larges. Les secondes permettent de segmenter l'image par des découpages successifs de grandes composantes (blocs de grandes tailles) en composantes plus petites.

Les méthodes descendantes

Ces méthodes sont performantes dans les situations où l'on connaît la structure a priori du document à analyser. C'est la raison pour laquelle, elles s'appliquent essentiellement à des documents très spécifiques et très hiérarchisés, tels les documents administratifs et scientifiques, [Baird90], [Pavlidis]. Elles sont, dans l'ensemble, relativement rapides car elles traitent les données dans leur globalité (à la différence des approches ascendantes). De plus, si l'image à analyser est de grande taille (cas des pages de journaux), on peut facilement réduire sa taille par des opérations de sous-échantillonnage, faisant ainsi apparaître la fusion des composantes les plus petites, [Bloomberg], [O'Gorman91], [Déforges]. Une des principales limitations des méthodes descendantes est liée au format des images traitées. L'environnement doit être très fortement normalisé pour garantir le maximum d'efficacité. Une représentation du document en blocs polygonaux ne pourrait pas être analysée par une telle approche, [Akindele]. Nous verrons par la suite que la mise en place de méthodes hybrides (mi-descendantes, mi-ascendantes) est une solution aux problèmes des documents de structures complexes.

Les méthodes ascendantes

La stratégie générale de la segmentation ascendante est fondée sur l'analyse de composantes connexes. Elle consiste à fusionner les morceaux jusqu'à l'assemblage complet de la page du document. La technique la plus répandue est la technique de lissage directionnel ; elle a été proposée par Wong *et al.* dans [Wong]. Des variantes statistiques ont été introduites dans [Normand], [Amamoto], [Wong], [Likforman], [Tsujimoto] pour limiter les effets du choix parfois arbitraire des seuils. Le texte est généralement considéré à pleine résolution et les réponses proposées sont souvent très spécifiques à un type de documents donné. Par ailleurs, ces méthodes nécessitent des connaissances a priori très fortes sur les caractéristiques typographiques des textes. Pour éviter ces limitations ainsi que celles qui sont liées aux méthodes descendantes, des approches *hybrides* ont été proposées par de nombreux auteurs. Cette nouvelle classe de méthodes constitue une alternative efficace aux méthodes traditionnelles.

Les méthodes hybrides

Comme il a été bien souvent évoqué dans [Pavlidis] et [Baird90], les méthodes hybrides (mi-ascendantes, mi-descendantes) sont plus efficaces que les méthodes exclusivement ascendantes et descendantes. Dans [Tsujimoto] et dans [Wieser], les auteurs ont mis à contribution cette mixité pour reformer à partir de la fusion d'objets (caractères, symboles) des composantes complètes de lignes et de paragraphes. On peut également citer parmi les méthodes les plus connues : l'approche par profils de projection, l'approche par transformée de Hough dans [Fletcher] et l'approche par classification par plus proches voisins dans [O'Gorman93]. Les méthodes hybrides permettent généralement de se dégager des fortes connaissances nécessaires à l'analyse (de type descendante) des documents, en évitant dans la plupart des cas de manipuler sur l'ensemble de l'image des données à pleine résolution responsable de la « lenteur » de certains algorithmes. De plus, les combinaisons ascendantes et descendantes offrent la possibilité de traiter des documents non contraints (documents contenant des éléments graphiques, des images, des portions de texte d'orientations variées...) et parviennent parfois à extraire du texte mixé à d'autres éléments informatifs, [Fletcher].

De plus en plus, l'intérêt est déporté vers de nouvelles approches, liées notamment à l'introduction de systèmes d'analyse multiagents, [Ishitani], ainsi que vers des notions de perception visuelle humaine simulant la capacité de l'homme à chercher, à voir et à comprendre, [Déforges], [Likforman]. C'est sur cette dernière notion liant l'aspect perceptif et cognitif de notre mode d'exploration que nous avons choisi de fonder notre approche. Celle-ci est basée sur la recherche de zones d'intérêt du document.

2.2. les approches liées à la capture d'information

Comme nous l'avons évoqué dans l'introduction, le lien qui existe entre la mise en page du document et la nature du message transmis est à la base de notre compréhension du contenu, [Barbara]. C'est la raison pour laquelle une partie du travail de conception du rédacteur se porte sur la mise en forme du document pour sa compréhension. Et c'est à travers cette mise en page et les outils typographiques utilisés que ce transfert de connaissances peut

se réaliser dans les meilleures conditions pour le lecteur. Celui-ci dispose de plusieurs stratégies pour aborder le document : la *lecture complète* (ou lecture mot à mot), l'*inspection* (ou recherche précise dans une région du document) et enfin le *survol* (ou aperçu du document) illustré à la figure 1. Cette figure présente des résultats expérimentaux de mesures oculométriques obtenues sur des observateurs humains. Ces mesures ont été prises au laboratoire CLIPS-IMAG de Grenoble. L'outil de mesure permettant de récupérer le parcours de l'œil sur les documents est un *eyetracker* (ou *eyeputer*). Les mesures obtenues ont été échantillonnées avec une fréquence de 60 points par seconde. Sur les images de la figure 1, nous n'avons représenté que les fixations qui correspondaient à un arrêt du regard d'au moins une demi-seconde. Les autres n'ont pas été conservées.

La tâche de survol à première vue très simple met en jeu des mécanismes de repérage d'informations très complexes qui soulignent le fait qu'il y a chez l'homme une émergence de l'information par zones d'intérêt, matérialisées sur ces résultats par des zones de fixations. Plus précisément encore, on s'aperçoit sur un grand nombre de résultats que les fixations sont localisées pour la plupart dans les zones d'images (grandes illustrations ou icônes), de titres et de sous-titres, c'est-à-dire dans les zones que le rédacteur du document aura particulièrement soignées et mises en évidence (taille des caractères, graisses, espaces inter-lignes, couleur...).

Les premiers constats faits sur la nature de l'exploration visuelle humaine nous ont conduit à reconsidérer l'importance de la mise en forme d'une page de document, [Barbara]. Le lien très étroit qui existe entre le rédacteur d'un document (celui qui met en forme) et le lecteur est à la base de notre proposition. Désormais, il ne s'agit plus de parcourir aveuglément le document en cherchant à en reconstituer la structure, mais au contraire, toute la problématique s'oriente à présent vers une recherche *guidée* par des régions présentant un *intérêt* visuel et sémantique dans

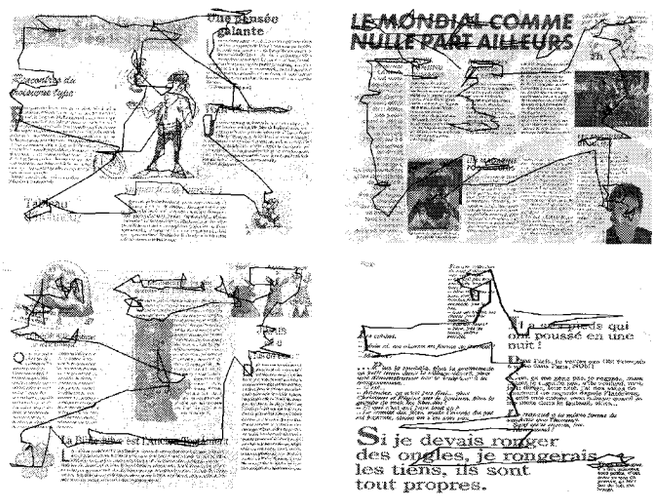


Figure 1. – Exemples de résultats de mesures oculométriques d'observateurs humains dans un contexte de survol du document.

la capture d'informations (ce dernier point sera abordé dans les perspectives). La découverte pas à pas des éléments *riches de sens* et d'informations nous conduit à une re-construction du document dans un ordre lié à l'intérêt que l'on porte aux régions explorées.

2.3. notre approche

Les Hypothèses fondamentales

A l'issue de cette analyse préliminaire sur le comportement de l'homme dans la recherche d'informations, nous avons retenu quelques hypothèses, à la base de notre approche :

- Toute l'information du document n'a pas le même poids visuel ni le même poids sémantique
- La segmentation faite par l'homme est liée à un ordre logique de parcours
- L'approche la plus adaptée que nous avons retenue consiste à imiter la stratégie de survol

Les deux premières hypothèses soulignent le fait que la recherche d'informations chez l'homme n'est pas aléatoire; elle ne se fait pas au hasard, au contraire, elle est implicitement liée à l'organisation des données sur le document. La dernière hypothèse traduit la solution que nous avons apportée et qui consiste à imiter le survol pour le repérage de l'information. En s'appuyant directement sur l'aspect visuel du document, le survol ne nécessite pas, dans une première version de notre système, l'utilisation de connaissances a priori. Il permet également de mettre en évidence très rapidement l'organisation globale du document et permet ainsi un perception équilibrée des données.

Exploitation des hypothèses

C'est donc sur ces hypothèses que nous avons basé la recherche de l'information de bas niveau. En guidant cette recherche, nous allons obtenir une description du document qui met en évidence des zones de fixation, à chaque fois différentes. La segmentation du document qui dérive de cette extraction de données mettra en évidence l'importance inégale des composantes informatives du document. Chaque point de fixation sera donc considéré comme un pointeur sur des données exploitables. Pour mener cette étude, nous sommes partis du constat que les méthodes relevant de la classe traditionnelle et évoquées précédemment sont généralement menées séquentiellement sur des images dont la *résolution* est uniforme partout. Dans la majorité des cas, elles demandent un échantillonnage très fin des données, qui n'est pas forcément nécessaire et qui nécessite, en outre, de longs temps de traitement. Il nous a donc paru utile, et même indispensable pour certaines applications, de reconsidérer l'image à des *résolutions variables*, pour ne privilégier une information à haute résolution que ponctuellement, permettant ainsi d'effectuer des traitements plus grossiers et plus rapides sur des régions de *moindre intérêt* (avec une résolution plus faible). La technique utilisée pour extraire les blocs informatifs n'est plus typiquement « ascendante »

ou « descendante » mais relève d'une analyse de type multirésolution de l'information. La re-construction de la structure du document peut alors être obtenue par le rassemblement (ou la *fusion*) des informations obtenues à partir de l'observation du document en différents points. Les techniques permettant de sélectionner les *zones informatives* sont présentées dans la suite. Dans ce contexte, nous pouvons citer les travaux de Yamamoto dans [Yamamoto] qui aborde la description des scènes avec une approche de type multirésolution utilisant l'intégration de plusieurs vues pour la reconstruction finale de l'image. D'autres travaux basés sur des modèles de la rétine, tels ceux proposés par Manzanera dans [Manzanera] ont souligné l'importance de la multirésolution pour le codage des images par régions d'intérêts et l'analyse et l'interprétation du mouvement sur des séquences vidéo.

3. notre contribution

Loin de tenter d'essayer de construire ce que nous pourrions appeler une « machine voyante », et qui ferait l'objet d'une démarche bien ambitieuse, il nous a semblé intéressant de choisir dans la complexité des processus de vision, les éléments qui pourraient faire l'objet d'une simulation informatique. Nous avons ainsi défini des règles de calcul et des algorithmes susceptibles de rendre compte de l'aspect *modulaire* et *progressif* des opérations perceptives de bas niveau. En particulier, nous avons tenté de nous dégager d'un échelonnement *séquentiel* entre une vision *préattentive* et *attentive*, et avons organisé notre travail davantage autour des notions complémentaires de *cycle perceptif*. Il s'agit donc de mettre en évidence l'existence de processus parallèles impliquant les domaines physiologiques, psychologiques (et cognitifs, ceux-ci n'ont pas été pris en compte dans cet article).

3.1. modèles et simulation

Les bases physiologiques de notre approche

Dans cette partie, nous allons faire le lien entre des modèles de l'*organisation fonctionnelle de la prise d'informations chez l'homme*, et l'usage que nous en faisons.

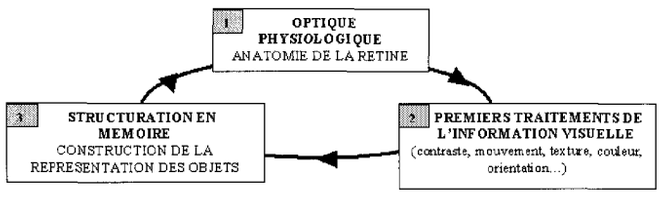


Figure 2. – Processus cyclique de repérage et de structuration de l'information en mémoire.

Les travaux sur lesquels nous avons basé notre étude sont ceux de chercheurs en psychologie, physiologie et sciences cognitives sur le processus de capture d'informations par l'homme. Treisman et Kosslyn dans [Treisman] et [Kosslyn] ont proposé des modèles de cette capture et ont ainsi mis en évidence les différentes étapes qui étaient impliquées : de l'œil... à la structuration de l'information en mémoire. La figure 2 illustre ce cheminement de l'information.

Exploitation du modèle

La première étape recensée sur le schéma concerne la structure de l'œil, plus précisément celle de la rétine. L'aire visuelle (rétinienne) se découpe en deux grands domaines : l'aire fovéale (où l'information est à haute résolution) et l'aire périphérique (qui possède un pouvoir d'attraction particulier avec une information à très faible résolution). Ce domaine de la périphérie qui apparaît très *flou* et très peu détaillé est essentiellement dédié à la perception du mouvement et des formes ponctuelles, [Tsotsos]. Nous simulerons pour cela un découpage irrégulier de l'espace visuel qui va ainsi nous permettre de mettre en évidence l'information attractive à haute résolution et l'information périphérique à basse résolution. Viennent ensuite les premiers traitements de l'information visuelle correspondant à la deuxième étape. A ce niveau, les premières primitives de bas niveau sont extraites. On recense généralement parmi ces primitives les informations de contours, de couleur, de texture et de mouvement. A notre niveau, nous définirons des facteurs de saisie des figures liés, dans un premier temps à l'information de contours (les traitements de la texture et de la couleur seront évoqués en perspective). Enfin la troisième étape de structuration de l'information en mémoire (mémoire associative) permet à l'observateur de s'approprier le champ visuel par la construction de la représentation des objets perçus. Nous simulerons cette étape par la reconstruction pas à pas du document : reconstruction progressive liée aux positionnements successifs des points de fixation. La représentation cyclique qui est illustrée sur le schéma traduit l'aspect cyclique de la prise d'informations chez l'homme : la reconstruction des données perçues ne peut se réaliser qu'à partir d'un ensemble de fixations impliquant l'intervention systématique de la première étape. La cycle est donc parcouru autant de fois que nécessaire pour capturer l'information souhaitée. La figure 3 présente les étapes que nous avons choisies de faire intervenir pour simuler la capture d'information par l'homme. Ce schéma est directement dérivé de celui de la figure 2.

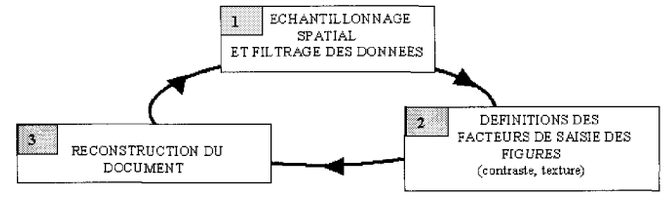


Figure 3. – Processus cyclique de structuration des documents.

Les bases psychologiques de notre approche

Le second constat de cette analyse est de nature *psychologique*. C'est lui qui va nous permettre de définir un ensemble de *règles pour la sélection* des zones de focalisation de l'attention. Ces règles concernent cette fois, non plus les caractéristiques visuelles de bas niveau de l'image (*points anguleux, à forte courbure et à fort contraste*, tels que les débuts et fins de lignes, les gros caractères, et les extrémités des formes, [Adelson], [Bonnet]), mais leur *organisation*, mettant ainsi l'accent sur les propriétés des figures : leur symétrie, leur régularité, leur compacité, leur proximité... en somme des propriétés qui ont été longuement étudiées par les adeptes de la théorie Gestaltiste et reprises par de nombreux psychologues, tels Bonnet, Lecas, Lévy-Schoen, dans [Bonnet], [Lecas], [Lévy]. Une liste plus complète des règles d'organisation a été synthétisée dans [Eglin97]. A l'issue de cette double analyse physiologique et psychologique, dont nous en présentons ici que les grandes lignes, on peut former le tableau de la figure 4 qui dresse le bilan des attributs qu'il va nous falloir prendre en compte pour définir le choix des fixations successives.

Domaines impliqués	Attributs génériques	Attributs du document
Physiologique	Contraste (contours, fins de lignes, points anguleux à forte courbure), couleur, texture, mouvement, orientation...	Contour, texture
Psychologique	Organisation, prégnance, symétrie, régularité, proximité, bonne forme...	Facteur de forme Mise en forme matérielle

Figure 4. – Attributs pris en compte dans notre simulation.

3.2. le modèle global de simulation

Le schéma de la prise d'information que nous proposons s'organise autour des trois étapes traitées au paragraphe 3.1. Ces trois étapes sont appelées respectivement dans le schéma suivant : Prétraitement, Choix des zones de fixation, Reconstruction du document. (voir figure 5).

Pour la partie de *Choix de zones de fixation*, nous ne présentons dans cet article que la définition des attributs géométriques liés à l'information de contours. De cette étape, on déduit le point de fixation suivant. Celui-ci alimente une *mémoire* qui fait évoluer la description du document au fur et à mesure de l'exploration. Celle-ci converge vers une représentation segmentée du document. Concrètement, le processus peut être décrit de la manière suivante. Nous entrerons dans le détail de mise en place des procédures dans la suite de l'article.

Dans l'ordre, le principe consiste à extraire les contours de l'image à partir de la première représentation centrée autour du premier

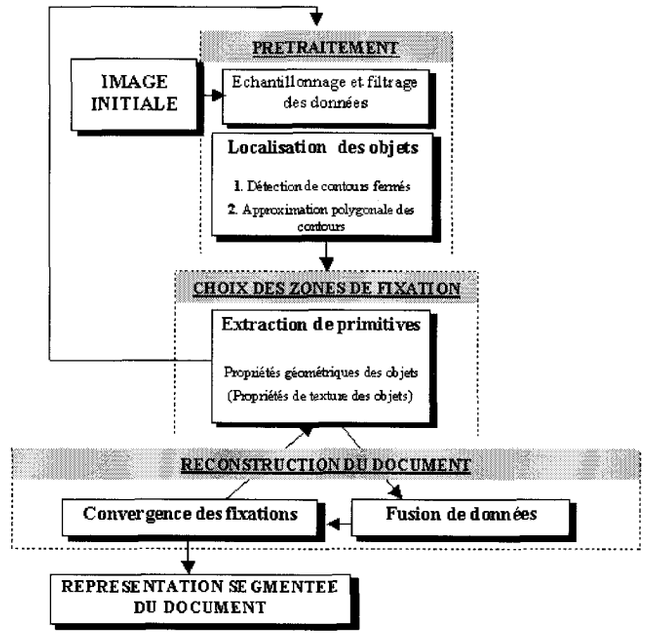


Figure 5. – Schéma fonctionnel de la prise d'information visuelle sur un document.

point de fixation. Cette représentation est de type multirésolution : à haute résolution au centre de fixation et à basse résolution en périphérie. De cette représentation en contours, nous extrayons une *approximation* polygonale (que nous appellerons *forme* dans la suite) qui simplifie la description du contour et facilite le calcul des attributs géométriques. Ces attributs ont été choisis en fonction de leur forte connotation *perceptive*, [Bonnet]. Le choix de ces primitives géométriques nous permet d'abord de sélectionner sur le document un ensemble de portions intéressantes (ou zones d'intérêt) où l'information est fortement concentrée celles-ci sont directement associées aux formes concernées sur lesquelles on peut alors calculer les attributs de surface, de symétrie (qui renseignent sur le caractère *prégnant* du bloc, [Wertheimer]), de compacité (qui renseignent sur la complexité de la forme). On part ainsi d'un ensemble de zones d'intérêt pour finir par un ensemble de points candidats. Il reste alors à choisir le point qui sera véritablement choisi. A chaque nouveau point est associée une nouvelle description que nous faisons fusionner avec la précédente pour obtenir deux représentations complémentaires : celle où apparaît le maximum de détails (*intégration à haute résolution*) ou au contraire celle où apparaît le moins de détail (*intégration à basse résolution*). Ce processus est répété jusqu'à ce que la description issue de la fusion des représentations successives ne puisse plus évoluer : on dira qu'elle converge vers la représentation la plus satisfaisante. A l'issue de cette étape se dégage ainsi une structure du document que l'on peut qualifier de physique. Elle renseigne sur la localisation des formes représentées à l'aide de boîtes englobantes (c'est-à-dire à l'aide des rectangles minimaux qui incluent chaque forme). Nous pouvons considérer cette première représentation comme une *esquisse évolutive de*

segmentation physique. Le terme *évolutif* traduit le fait que la segmentation se compose pas à pas, et que d'un point de fixation au suivant on accède à une étape supplémentaire dans l'évolution de la représentation finale.

4. aspects techniques du système

4.1. les opérations de prétraitement

Description fonctionnelle du prétraitement

Les opérations successives de prétraitement que nous allons présenter dans ce chapitre se synthétisent par le diagramme suivant, voir figure 6. Elles se composent successivement d'opérations de filtrage spatial atténuant les informations visuelles périphériques et d'opérations de détection de contours rendant compte des zones de forts gradients. Ces opérations sont basées sur un découpage (ou un pavage) irrégulier de l'image, correspondant à une simulation des mécanismes physiologiques de l'œil.

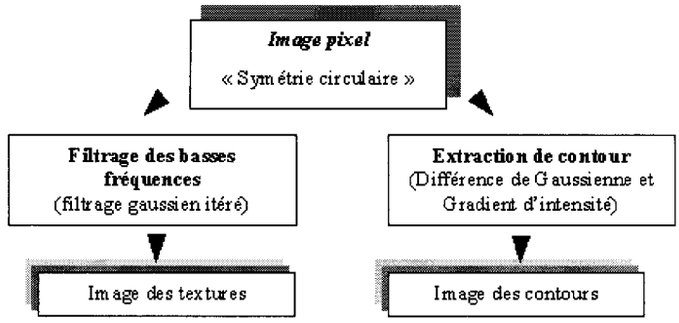


Figure 6. – Description fonctionnelle du prétraitement.

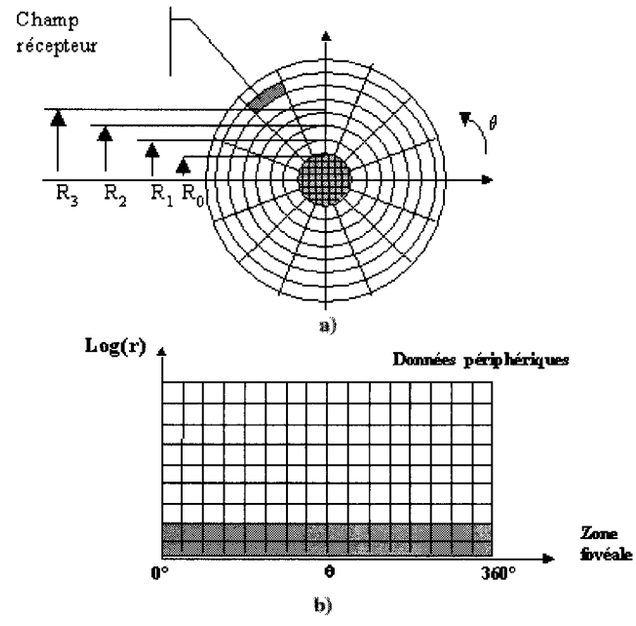


Figure 7. – Pavage irrégulier de l'espace (7.a) et représentation de sa projection log-polaire (7.b).

Modèle de l'échantillonnage spatial

Pour rester le plus fidèle possible à l'aspect biologique de l'œil humain (de la rétine), nous avons choisi une organisation des champs récepteurs variable dans l'espace. Elle permet de rendre compte de la sensibilité de l'œil au contraste et en particulier au contraste en vision périphérique. La première étape de ce processus consiste à définir une répartition ordonnée des *niveaux concentriques de perception*. Compte tenu du fait que notre perception n'est pas homogène sur toute sa surface, nous avons basé la représentation de l'image rétinienne sur la construction de *niveaux de perception*. Il s'agit d'anneaux concentriques centrés en un point appelé point de fixation et ayant comme rayon la valeur d'une fonction croissante de la distance à ce point. Nous utiliserons une croissance *exponentielle* des niveaux de perception qui garantit une croissance *linéaire* des surfaces des champs récepteurs. Les notions de niveaux de perception et de champs récepteurs sont illustrées sur le schéma de la figure 7.a. La construction des niveaux, appelés encore anneaux concentriques dans [Yamamoto], ou réseaux multicouches [Crettez] est née de la classification des champs récepteurs en sous-ensembles selon leurs dimensions, leur orientation et leur complexité. La notion purement formelle à notre niveau de *champ récepteur* sera utilisée pour quadriller l'espace de représentation, voir figure 7.a et 7.b. Pour cela, nous définissons un ensemble de secteurs angulaires réguliers variant de 0 à 2π et partitionnant l'espace à partir du point de fixation courant. Sur la figure n'ont été représentés que 9 niveaux concentriques et 16 secteurs angulaires.

Les rayons concentriques (R_i) suivent une progression géométrique définie par la suite suivante : $R_i = R_0 \times q_i$, où R_0 représente le rayon fovéal, et $i \in \{1, N\}$. N est le nombre maximal de rayons autorisés et q est la raison de la suite. Elle est définie comme une fonction de la distance du point de fixation courant au bord le plus éloigné de l'image et de telle sorte à ce que l'on conserve un nombre constant de rayons quelle que soit la position du point de fixation. Ceci nous garantit ainsi une description homogène du champ visuel par projection log-polaire, conservant ainsi une même structure : même nombre de secteurs et même nombre de rayons.

On a ainsi :

$$q = (d_{\mathrm{Max}}/R_0)^{1/N},$$

avec d_{Max} la distance maximale à un des bords de l'image.

D'un point de fixation à l'autre, c'est la valeur de d_{Max} qui change. Finalement, nous pouvons exprimer la suite croissante des rayons comme une fonction exponentielle de l'excentricité. Cette définition exponentielle va nous permettre de passer à un

nouvel espace de représentation connu sous le nom d'*appariement log-polaire*. Il a été introduit de manière similaire par Wilson dans [Wilson] et Yamamoto dans [Yamamoto] dans un contexte de vision active. Finalement, en exprimant les R_i par la relation suivante :

$$\log(R_i/R_0) = \log(q^i) \Rightarrow R_i = R_0 \times \exp^{(i \times \log(q))}$$

et avec $q = (d_{\text{Max}}/R_0)^{1/N}$ on obtient :

$$R_i = R_0 \times \exp^{(i/N \times \log(d_{\text{Max}}/R_0))}$$

pour $i \in \{1, N\}$.

Cette nouvelle définition des R_i permet d'introduire le pavage irrégulier de l'espace par un changement de repère lié à l'utilisation d'une échelle logarithmique. Le changement d'espace s'effectue en prenant le logarithme des coordonnées de chaque point défini dans l'espace polaire. On définit ainsi la fonction $F_{\log}$ comme le passage de l'espace polaire à l'espace logarithmique, soit :

$$F_{\log} : r.e^{i\theta} \longrightarrow \log(r) + i \cdot \theta$$

Nous avons choisi d'implémenter le système avec 32 rayons concentriques ($N = 32$) au maximum. Les raisons de ce choix sont les suivantes : avec 32 niveaux et notre définition des anneaux concentriques, le rapport entre la largeur d'un champ récepteur (Lc) et son éloignement au point de fixation (sa distance au centre dc) est compris entre 0.1 et 0.2, ceci quelque soit la taille de l'image considérée (de 128*128 à 2048*2048), voir figure 8. Des mesures physiologiques ont permis de conclure que chez l'homme ce rapport était compris entre 0.1 et 0.3, [Shah].

Notre représentation bien que très largement simplifiée est cependant qualitativement compatible avec les données physiologiques concernant la répartition des photorécepteurs à la surface de la rétine. Rappelons-nous à ce propos, que la répartition des cellules ganglionnaires et des champs récepteurs est directement à l'origine de notre perception dégradée des scènes. Voyons comment cette dégradation va nous permettre de traiter plus efficacement l'image.

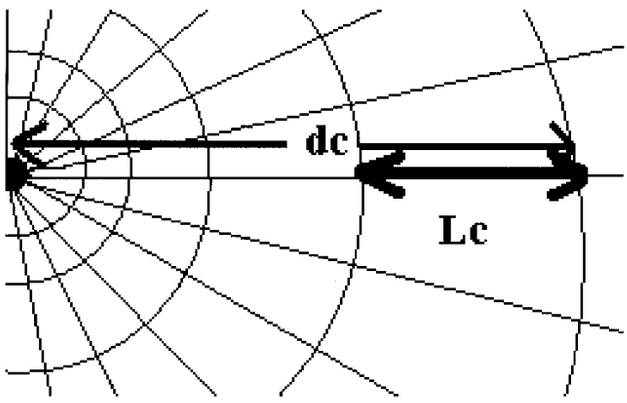


Figure 8. – Représentation du rapport Lc/dc à partir d'un pavage centré en un point de fixation quelconque.

Filtrage périphérique des données

La deuxième étape de notre système consiste à rendre compte de l'aspect dégradé des formes en périphérie, en respectant le découpage irrégulier de l'espace liée au pavage défini précédemment. Le principe consiste à dégrader l'information du document en supprimant les hautes fréquences en périphérie. Ceci nous permet de considérer les régions très peu filtrées (régions centrales) avec leur pleine résolution d'origine, et les régions périphériques (très filtrées) avec des résolutions variables proportionnelles à leur éloignement au centre. Pour cela, nous sommes partis de résultats d'expériences (faites par des physiologistes et reprises par des théoriciens, tels Marr, et Bruce, dans [Marr], [Bruce]) qui montrent que l'organisation des champs récepteurs d'un certain nombre de cellules ganglionnaires est de type *gaussien* (et plus généralement de type Gabor correspondant aux traitements de bas niveau du cortex sensibles à l'*orientation* spatiale de l'information) et ressemble de très près au LoG (laplacien d'une gaussienne). Les filtres utilisés constituent une approximation convenable en terme de complexité calculatoire et de durée de traitements. Il n'est pas sans importance de rappeler ici que l'objectif d'un système de vision n'est pas de reproduire le comportement visuel humain dans les moindres détails ou de tenter d'en trouver une équivalence, mais au contraire de s'en inspirer en le simplifiant et l'adaptant à nos besoins.

Le principe du filtrage consiste donc réaliser des filtrages passe-bas de plus en plus importants au fur et à mesure que l'on s'éloigne du point de fixation choisi. En pratique, on construit une suite d'images $I_i(x, y)$, avec i variant de 0 à n. $I_0(x, y)$ est l'image d'origine, $I_n(x, y)$ est l'image résultat. L'image $I_{i+1}(x, y)$ s'obtient par produit de convolution de l'image $I_i(x, y)$ avec la fonction gaussienne définie ci-dessous (dans le domaine fréquentiel) dans l'intervalle $[R_{n-i}, R_n]$:

$$g(u, v) = \mathrm{e}^{\frac{(u^2+v^2)}{2\sigma^2}}$$

Ce produit de convolution est réalisé par transformée de Fourier inverse du produit simple de la transformée de Fourier de l'image et de la fonction gaussienne $g(u, v)$. C'est donc toute l'image qui sera filtrée à chaque itération i pour donner une image $I_i(x, y)$, voir figure 9. L'opération effectuée consiste donc à appliquer $(n-1)$ fois la fonction de lissage gaussien sur l'intervalle $[R_{n-1}, R_n]$, $(n-2)$ fois sur l'intervalle $[R_{n-2}, R_{n-1}]$, et finalement aucune fois sur l'intervalle $[R_0, R_1]$. D'intervalle en intervalle, on parvient ainsi à une décroissance de la « netteté » (ou de la précision) de la fovéa à la périphérie.

Ceci revient à filtrer l'image de départ en utilisant une série de fonctions gaussiennes $g_k(u, v), k \in \{1, \ldots, n\}$, équivalentes aux k produits simples de g avec elle-même :

$$g_k(u, v) = \left[\mathrm{e}^{-\frac{(u^2+v^2}{2\sigma^2}}\right] = \mathrm{e}^{-\frac{(u^2+v^2)\cdot k}{2\sigma^2}} = \mathrm{e}^{-\frac{(u^2+v^2)}{2\cdot\frac{\sigma^2}{2}}}$$

L'image finale correspond ainsi à l'application successive de plusieurs fonctions de transfert $G_{n,i}$ sur l'image d'origine.

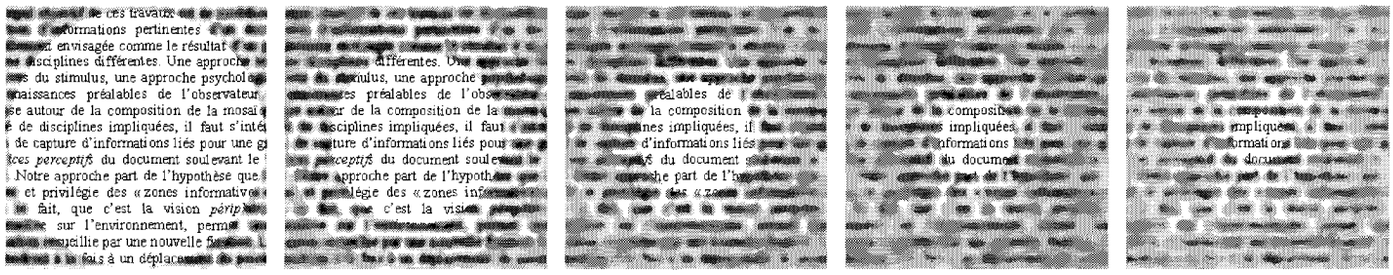


Figure 9. – Représentation simplifiée (avec seulement cinq niveaux concentriques) des résultats pas à pas du filtrage itératif. De la première image à la dernière, on peut constater la croissance du flou en périphérie. La zone de foyer est au centre de l'image.

L'utilisation du filtrage fréquentiel tel que nous l'avons présenté ci-dessus permet de réduire les effets de bords qui apparaissent sur l'image lors d'un filtrage spatial par masque de convolution. De plus, cela permet de réaliser un vrai filtrage gaussien avec l'atténuation théorique attendue sur tout le domaine fréquentiel, et non pas une approximation comme on peut l'avoir avec l'utilisation des masques de convolution. Les résultats sont donc plus conformes à nos attentes.

Extraction de contours : coopération DoG/ gradient

Nous allons tirer profit du filtrage gaussien précédent et extraire à l'intérieur des différents domaines les changements d'intensité (les hautes fréquences). Puisque le filtrage passe-bas croît avec la distance au centre (point de fixation), les *hautes* fréquences restantes correspondent à des contours de plus en plus grossiers en périphérie. Pour mettre en évidence cette dégradation périphérique, nous avons réalisé un *détecteur de contours* directement basé sur des différences de gaussiennes successives (DoG), appelées également DOLP (Difference of Low-pass transforms) par Crowley dans [Crowley]. Ces différences sont calculés sur l'ensemble du champ à l'intérieur de chaque domaine $[R_i, R_{i+1}]$ de l'image filtrée, voir figure 10. Selon cette définition, la fovéa ne subit aucun filtrage; elle reste donc inchangée. Dans la pratique, le filtrage passe-bas itératif et l'extraction de contours des objets sont calculés simultanément.

Parallèlement à l'évaluation de DoG, nous appliquons à l'image un opérateur gradient avec un seuil global élevé pour délimiter les grandes régions du document (paragraphes de textes, images...) et supprimer les petites composantes dans la périphérie de la zone de fixation. Le seuil global est fixé *a priori* pour l'ensemble des images testées (la valeur est 80). Ceci est rendu possible par le fait que les images traitées sont constituées essentiellement de texte et présentent des niveaux de résolutions identiques (du centre à la périphérie, nous avons fixé à 32 le nombre de niveaux concentriques, ce qui renseigne sur les caractéristiques des portions de textes filtrées à chaque niveau).

Une manière simple d'éliminer les contours de faibles gradients revient à effectuer une simple opération logique (ET) entre les deux images de contours binaires DoG et gradient. Les contours résultants sont ainsi d'épaisseur unitaire (car les contours DoG le sont), et sont représentatifs des fortes discontinuités en périphérie sans tenir compte des variations locales à l'intérieur des paragraphes. Les contours ne sont cependant pas continus (pas connectés). Pour garantir la continuité des contours périphériques

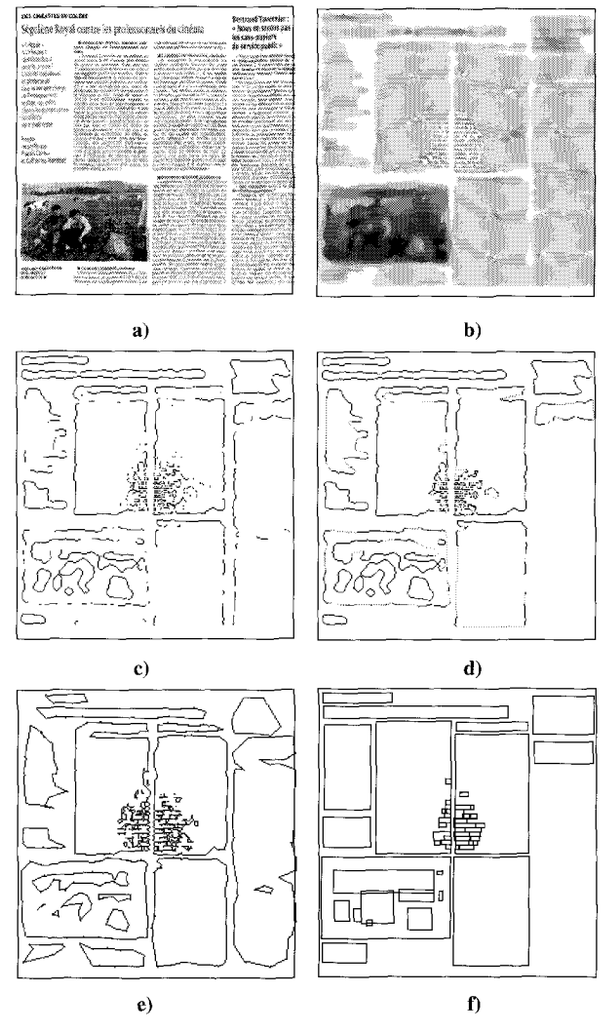


Figure 10. – a) Image originale. b) Résultat du filtrage à partir d'un point de fixation central. c) Résultat de l'opération de mise en correspondance des contours DoG et gradient (par l'opération logique ET). d) Représentation de contours fermés par couplage des extrémités de segments. e) Représentation de l'approximation polygonale des formes. f) Représentation des rectangles circonscrits aux blocs.

obtenus par l'opérateur gradient, nous proposons une approche basée sur la connexion *des extrémités de segments*. Il s'agit pratiquement de reconstituer le contour complet des blocs en cherchant à joindre les extrémités libres les plus proches.

Couplage des extrémités de segments

L'objectif de cette procédure est de reconstituer les contours continus des blocs. L'algorithme que nous proposons est le suivant. Un premier balayage de l'image permet de localiser puis de marquer l'ensemble des points extrémités (*nœuds*). On traite toujours successivement les deux extrémités d'un même segment de manière à ne pas reconstituer en même temps le contour de deux blocs différents : on s'assure de cette manière de la bonne connexité des contours. On cherche, dans le voisinage de chaque nœud, les cinq extrémités les plus proches non encore connectées (la taille du voisinage est liée à la localisation de ces cinq points). On calcule ensuite la distance moyenne des cinq points au nœud courant. Puis, on évalue sur les cinq segments ainsi formés la valeur moyenne des gradients de façon à ne privilégier que le segment de gradient maximal. Lorsqu'on ne peut pas conclure (plusieurs moyennes identiques, ou gradients non significatifs), on privilégie simplement le point le plus proche, lorsque la distance au point courant est très inférieure à la moyenne des distances. Sinon, dans un cas plus général, on privilégie les directions de segments horizontales et verticales. Ceci constitue la base du couplage des extrémités. Il faut, dans tous les cas, s'assurer qu'un nœud ne peut être connecté à plus de deux extrémités. Une fois connecté, un nœud est supprimé de la liste des candidats. Par ailleurs, deux segments ne doivent jamais se croiser. S'en assurer permet de rejeter des situations d'erreur. Certaines erreurs fréquentes de couplage peuvent néanmoins survenir. On note en particulier les situations où un nœud ne trouve pas d' « associé ». Dans ce cas, il faut modifier dans le voisinage de ce nœud certaines liaisons ayant conduit à ce résultat. En revanche, si un segment est complètement isolé après la procédure de couplage, alors on le supprime. Dans ce processus, les points isolés ne sont jamais négligés, car ils permettent de préciser la direction générale d'une ligne.

Il est donc possible que plusieurs solutions soient obtenues par cette approche. Ceci n'est pas gênant, car la segmentation finale résulte d'une intégration d'une succession de descriptions : on parvient généralement à des découpages cohérents. Le résultat de l'application de cette procédure est illustré à la figure 10. Le découpage en rectangles est obtenu par la mise en place d'une procédure simple de suivi de contours. On conserve ensuite pour chaque contour fermé, les points les plus extrêmes, et on trace le rectangle passant par ces points. C'est finalement sur l'approximation polygonale des contours, correspondant à l'intersection entre les maillons du pavage irrégulier et les contours que portera la caractérisation géométrique des formes que nous allons aborder. Dans les exemples que nous avons traités, nous n'avons pas pris en compte des documents de mise en forme particulière contenant des données textuelles inclinées, ou des images non rectangulaires. Ce cas de figure est envisagé dans les perspectives. Des solutions sont actuellement à l'étude.

4.2. choix des zones de fixation

Principe général

Les attributs que nous avons retenus vont intervenir de manière séquentielle. Nous les avons classés à la figure 11 du plus global au plus local. Nous sélectionnons ainsi sur toute l'image des régions entières possédant une forte distribution des contours. Ces régions correspondent à des formes fermées qui sont choisies en fonction de leur surface, symétrie, compacité. On ne retient finalement que les points de forte courbure restants.

La première étape consiste au repérage des zones à forte densité de contours. On conserve les blocs correspondant à ces régions. Cette étape conduit généralement à éliminer les zones très peu chargées d'informations. Le deuxième étape consiste ensuite à sélectionner parmi les blocs précédents, ceux dont la surface est supérieure à la moyenne des surfaces des blocs. Cette règle met en évidence des blocs de grande taille, et supprime les petits blocs recouvrant les composantes connexes du texte ou le bruit. Elle permet de ne pas sélectionner systématiquement les régions très proches de la fovéa, généralement très riches en contours. On mesure ensuite pour les blocs restants (issus des deux premières étapes) la compacité puis le facteur de symétrie (étape 3 et 4). Pour la compacité et la symétrie, on ne garde que les blocs présentant une valeur très faible (inférieure à la moyenne de valeurs obtenues sur les blocs restants). Il ne reste alors qu'un ensemble très réduit de blocs de configurations *relativement* stables (présentant un fort relief structural et une symétrie maximale). Enfin, à l'étape 5, on sélectionne *le* point qui sera le candidat unique à la fixation suivante. Pour cela, on évalue sur les blocs issus des quatre premières étapes un ensemble de *points d'intérêt* localisés sur les contours des blocs correspondant aux angles à fortes courbures. Sur le bloc le plus éloigné du point de fixation courant (lorsque plusieurs blocs sont candidats), on conserve alors le point de courbure maximale situé sur le contour de l'objet.

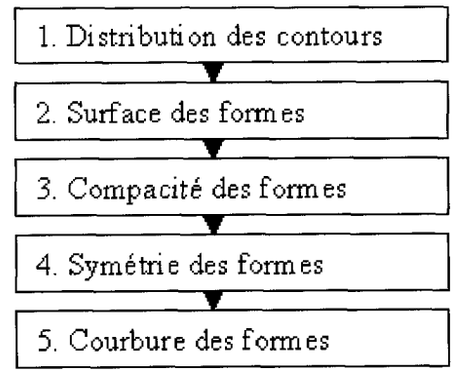


Figure 11. – Processus séquentiel du choix des zones de fixation.

La distribution des contours

Nous avons repris certains constats liés à la psychologie de la perception qui montrent que plus une figure est structurellement complexe, plus elle attire le regard, [Treisman]. Nous avons choisi de mesurer la « complexité des formes » du document, par l'analyse de la distribution des contours des objets répartis sur l'ensemble du champ de vision. La distribution qui va nous intéresser est une distribution des contours par maillon du pavage. Plus spécifiquement, nous avons calculé pour chaque *champ récepteur* (portion comprise entre deux anneaux concentriques et deux secteurs angulaires successifs), le nombre d'intersections entre les contours des régions et le maillage circulaire. Ce nombre nous permet d'évaluer *localement* la densité des contours répartis sur le document à partir du point de fixation courant. La fonction densité utilisée est appelée $CB_r(\theta_{k+1} - \theta_k)$, concentration des bords des objets pour chaque rayon r et chaque secteur angulaire d'amplitude constante $(\theta_{k+1} - \theta_k)$, avec θ_k valant $2k\pi/N_\theta$, pour k variant de 0 à $N_\theta - 1$ (N_θ est le nombre total de secteur angulaire). On ne conserve alors que l'ensemble des *maillons* dont la densité est supérieure à la distribution moyenne des densités sur l'ensemble du champ. Ainsi, cet ensemble, noté $\Theta_{\theta,r}$, doit vérifier la relation suivante :

$$\Theta_{\theta,r} = \left\{ (\theta, r) \left[CB_r(\theta_{k+1} - \theta_k)_{k \in \{0,2N_\theta - 1\}} > \frac{1}{N_\theta} \sum_{k=0}^{n_\theta - 1} CB_r(\theta_{k+1} - \theta_k) \right] \right\}$$

Cette analyse par *maillon* du pavage nous renseigne sur la localisation spatiale des zones denses fortement traversées par des contours. On peut ainsi représenter l'image du document sous la forme d'une carte des distributions où les maillons les plus denses constituent des zones privilégiées de fixation, voir figure 12.

Sur l'exemple de la figure 12, nous avons représenté l'ensemble des intersections entre le maillage et les contours autour d'un point de fixation. Celles-ci sont caractéristiques des zones des transitions, notamment au niveau des frontières des blocs et des images dont les contours sont généralement plus complexes et plus variés que sur le reste du document.

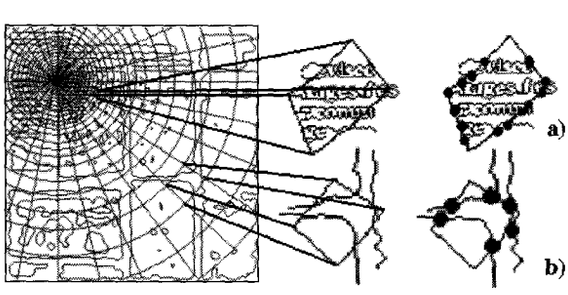


Figure 12. – Représentation des intersections entre les contours issus de la représentation autour d'un point de fixation et le maillage centré en ce point. 12.1.) Zoom sur un maillon de la zone fovéale de texte. 12.2.) Zoom sur une zone de transition entre plusieurs paragraphes.

Surfaces relatives des formes

Lorsque l'on connaît les objets par leur contour, le théorème de Pick permet de calculer leur surface en nombre d'unités de tesselles du maillage (théorème démontré dans [Sankar]). Le contour est représenté par un polygone (que nous avons appelé *approximation polygonale*) dont les sommets sont des points du maillage. Pour calculer la surface, on fait l'approximation d'un point situé sur une arête du maillage par le point du maillage le plus proche. On désigne par I le nombre de points discrets situés à l'intérieur de la ligne polygonale composant le contour et par B le nombre de points discrets appartenant au contour polygonal (appelés points bords). Cette classe inclut les sommets de la ligne polygonale et éventuellement d'autres points discrets situés sur les côtés de la ligne polygonale, voir figure 13.

Les points extérieurs approximés correspondent aux points du maillage les plus proches des intersections. La formule de Pick, qui évalue la surface de la composante connexe délimitée par la ligne polygonale est donnée par la formule suivante : $S = I + B/2 - 1$. Sur l'exemple de la figure 13, la surface de la forme est de 50 unités de tessellation $(37 + 28/2 - 1)$. On peut facilement vérifier le résultat en comptant les unités grisées du maillage. Finalement, nous avons choisi d'appliquer ce calcul de surface au pavage irrégulier de l'espace rendant compte de la décroissance d'une surface avec l'éloignement à la fovéa, voir figure 14. Ainsi, plus un objet est éloigné du centre, moins il existe de points d'intersection entre son contour et le maillage circulaire (simplifié ici pour faciliter les calculs).

En utilisant la méthode de calcul de surface proposée dans ce paragraphe, nous pouvons constater que la surface d'un même objet positionné à différents endroits du champ visuel ne conserve pas une valeur constante, mais qu'elle décroît avec l'excentricité. Cette décroissance de surface est en ce sens « proportionnelle » à l'acuité visuelle (courbe hyperbolique décroissante). Utiliser la surface comme facteur de forme présente donc l'intérêt de donner une importance *relative* aux objets selon leur positionnement dans le champ visuel et de ce fait, de ne privilégier un objet que s'il présente une surface suffisante pour être *perçu*. En particulier, un objet disparaît du champ visuel s'il n'existe plus d'intersection entre ses contours et le maillage (l'objet est dans ce cas complètement inclus à l'intérieur d'un maillon).

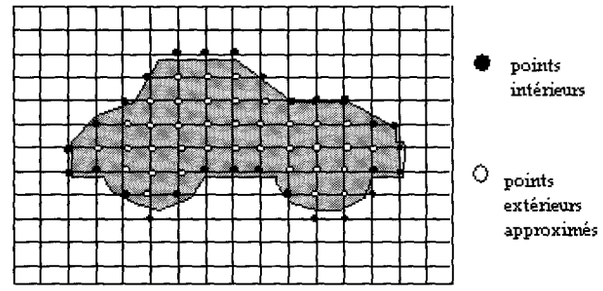
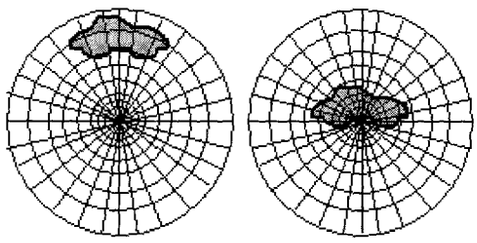


Figure 13. – Points discrets intérieurs et points discrets du contour approximés d'une forme polygonale. Le pavage utilisé est un pavage carré régulier.



Bloc	Nombre de points intérieurs I	Nombre de points du bords B	Surface S
B1	4	11	**8**
B2	42	30	**56**

Figure 14. – Résultats du comptage de points d'intersection entre le contour d'un objet (positionné à deux endroits différents du champ visuel) et le maillage circulaire (simplifié ici pour faciliter les calculs).

La mesure de symétrie

En reliant les points d'intersection du contour de l'objet avec le maillage irrégulier, on obtient une approximation polygonale du contour de l'objet, plus facile à manipuler que le contour complet. La mesure et la localisation des différents segments reflètent le caractère symétrique des formes. Par ailleurs, le centre de gravité des objets détermine une orientation possible d'un éventuel *axe de symétrie* (passant par le centre de fixation). Ainsi, pour chaque cercle concentrique, la *fonction de symétrie* (définie ci-dessous) permet de mesurer les différences de part et d'autre de cet axe de symétrie. Cette différence est évaluée comme un coût d'édition entre deux chaînes, [Hacisalih]. Pour chaque rayon j, on évalue les facteurs $\delta_{j,r}$ et $\delta_{j,l}$ correspondant respectivement aux nombres de points d'intersection entre des formes polygonales avec le maillage à droite (notée r) et à gauche (notée l) de l'axe pris pour référence de symétrie. On pondère ensuite les différences $(\delta_{j,r} - \delta_{j,1})$ avec un facteur $(1 - 1/r_j)$ ([0,1], qui augmente avec l'excentricité (r_j correspond à la valeur du $j^{\text{ème}}$ rayon mesuré à partir de l'origine P_f). En effet, les différences entre les deux « côtés » d'un objet sont moins nombreuses en périphérie mais correspondent à des changements plus importants dans la forme globale. L'expression finale, à la différence d'un coût d'édition a été normalisée. Finalement, l'expression de la fonction de symétrie est la suivante :

$$\Delta_{\text{Sym}} = \left(\frac{\sum_{j=r_1}^{r_p} (\delta_{j,l} - \delta_{j,r})^* \left(1 - \frac{1}{r_j}\right)}{\sum_{j=r_1}^{r_p} [\Gamma_j - \Gamma_r](j)^* \left(1 - \frac{1}{r_j}\right)} \right), \quad \Delta_{\text{Sym}} \in \{0,1\}$$

où $\Gamma_r + \Gamma_l$ est le nombre total de points d'intersection impliqués dans l'approximation polygonale du contour pour chaque rayon j.

Si l'objet est symétrique, Δ_{Sym} est proche de 0, sinon Δ_{sym} peut prendre un ensemble de valeurs proches de 1. La mesure de symétrie dans notre cas va être évaluée globalement sur les différents blocs de l'image. La valeur moyenne des résultats obtenus nous permet de fixer un seuil au delà duquel les blocs ne sont plus considérés. Nous pouvons illustrer sur un exemple simple le calcul de la valeur de symétrie obtenue sur un bloc de texte situé en périphérie du point de fixation central, voir figure 15. Les points noirs représentent les intersections entre le contour et le maillage circulaire. Le graphique de droite illustre la représentation *log-polaire* du bloc. La déformation est liée au changement de repère mais ne change en rien le résultat de symétrie. Le centre G correspond au centre de gravité du bloc. Ainsi, en reprenant les notations précédentes, on obtient pour Δ_{Sym} une valeur de 0.12 proche de 0, voir figure 15. Les points grisés correspondent aux points non symétriques.

Il faut noter également que compte tenu du grand nombre de secteurs angulaires utilisés dans l'évaluation des paramètres géométriques des blocs (nous avons implémenté notre système avec 32 niveaux), les blocs « réellement » symétriques conservent leur propriété de symétrie (avec une valeur de Δ_{Sym} significative), et ceci même avec une approximation polygonale obtenue à l'aide d'un pavage irrégulier de l'espace. Dans tous les cas, si le bloc n'est pas trop éloigné de la zone fovéale, c'est-à-dire si les intersections entre son contour et le maillage sont représentatives

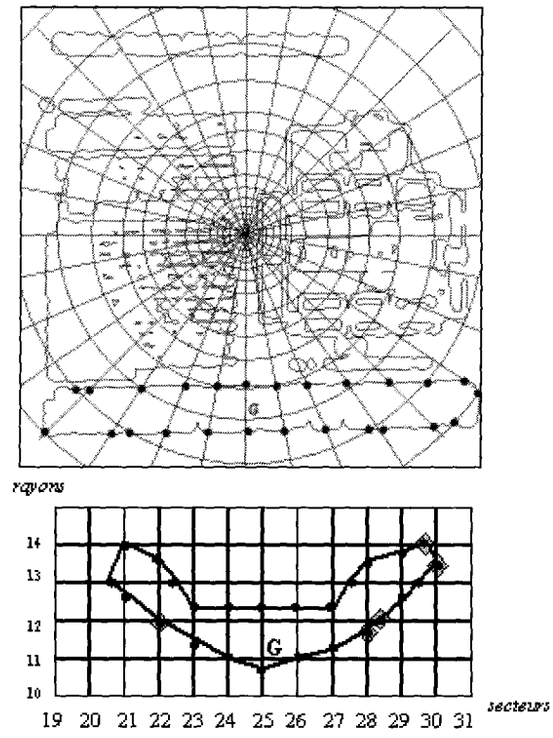


Figure 15. – Représentation de la projection d'un bloc dans le repère log-polaire pour le calcul de la symétrie.

de la forme, la propriété de symétrie est vérifiée. Les cas particuliers de non-symétrie correspondent aux blocs très éloignés du centre et donc très peu recouverts par le maillage. Dans ce cas, la facteur de symétrie n'est plus déterminant dans le choix de la fixation. Signalons en ce sens qu'un bloc très éloigné du centre de fixation courant n'est choisi pour la fixation suivante que s'il accumule tous les paramètres de symétrie, courbure, compacité et distribution de contours.

Compacité

La compacité C d'un objet s'exprime comme le rapport entre la surface d'un objet et son périmètre au carré, soit C =surface/(périmètre)^2 . Cette mesure est par définition sans dimension; elle s'adapte donc très bien aux changements d'échelle imposé dans notre étude par l'irrégularité du pavage. Pour chaque bloc détecté sur le document, nous avons choisi de calculer ce facteur de forme car il renseigne sur la relative « complexité » du bloc. Ainsi, plus le rapport C est faible (proche de 0), plus la forme est globalement complexe, en particulier elle possède un contour très découpé. En effet, le rapport surface/périmètre^2 diminue d'autant plus que le périmètre est important, ce qui est le cas pour des objets dont les contours ont un très important *relief*. D'après les règles que nous avons retenues, ce type de contour sera considéré comme ayant un fort pouvoir attractif.

La fonction de courbure

Chaque objet étant représenté par un contour discret, il est intuitif de définir comme points critiques les points à forte courbure, c'est-à-dire les sommets des approximations polygonales. En géométrie discrète, la courbure calculée en un point utilise des points de la courbe plus éloignés que ses simples points adjacents. Mais plus on s'éloigne de la fovéa, moins l'objet dispose de points pour sa description, donc moins on peut considérer les points éloignés pour le calcul de la courbure. Nous avons donc défini une fonction qui attribue, en fonction de l'éloignement à la fovéa, un ordre k, correspondant à l'éloignement des points P_{i-k} et P_{i+k} à prendre en compte autour du point de courbure P_i. Plus le point P_i est proche de la fovéa, plus il est entouré de points de contours nombreux. En revanche, plus on s'éloigne de la fovéa, plus les points de contours se réduisent au voisinage de P_i. Ainsi, plus on s'éloigne de la fovéa, plus la distance entre deux points P_i

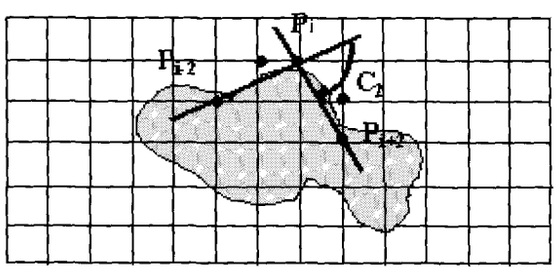


Figure 16. – Représentation de la courbure en un point d'une courbe à l'ordre 2.

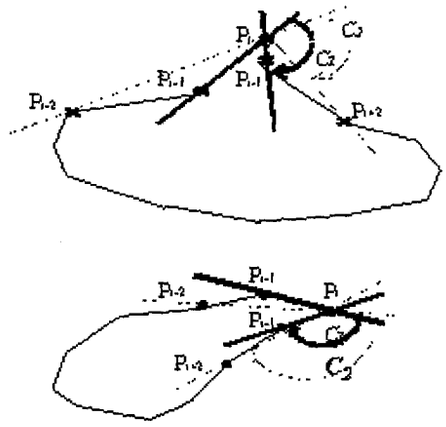


Figure 17. – Représentation du voisinage de deux points et de la mesure de courbure en chacun de ces points. 17.1. Point P_i localement saillant, peu stable ($C_1 >> C_2$). 17.2. Point P_i globalement saillant, très stable $C_1 \approx C_2$.

successifs augmente et donc l'ordre k doit diminuer pour que la courbure soit significative. La courbure en P_i est ainsi définie par la relation suivante de la figure 16.

$$C_k = \text{angle}(P_{i-k}P_i, P_iP_{i+k})$$

Plus la courbure est forte, plus le point P_i correspond à un point anguleux, et plus il sera attractif aux yeux de l'observateur. Plus l'angle en P_i est saillant (et pas simplement localement saillant, voir figure 17), plus le valeur de la courbure reste stable avec des valeurs de k croissantes. L'extraction des points à forte courbure dépend donc de la largeur du voisinage choisi pour le point P_i considéré.

Finalement, nous pouvons tracer la fonction de courbure d'un objet mesurée en tout point discret de son contour. Les valeurs maximales de la courbe seront récupérées comme candidat potentiel au point de fixation suivant. Compte tendu du fait qu'un polygone régulier est constitué de N côtés et donc de N angles valant exactement $(N-2)\pi/N$, nous avons décidé de ne conserver parmi les points candidats que ceux dont la courbure était supérieure à la courbure d'un polygone régulier, soit : $C_k > \pi - (N-2)\pi/N$, soit $C_k > 2\pi/N$. En disposant de ce seuil de courbure minimale, nous limitons le nombre de points et ne conservons que ceux qui correspondent à des points visuellement saillants. L'exemple de la figure 18 illustre sur deux exemples la sélection de points de forte courbure à partir de l'approximation polygonale des contours.

La suppression des points redondants

Jusqu'ici, nous avons évoqué la définition des points de fixation successifs sans tenir réellement compte des points redondants (points de fixation choisis deux fois). Afin d'éviter la redondance d'information, on repère les secteurs du document qui ont été très peu visités (constitués des points dont la distance au centre de fixation courant est la plus élevée). Ces secteurs sont caractérisés par des concentrations de pixels à faible résolution. Ainsi, une zone dont les primitives visuelles sont nombreuses et qui *attire*

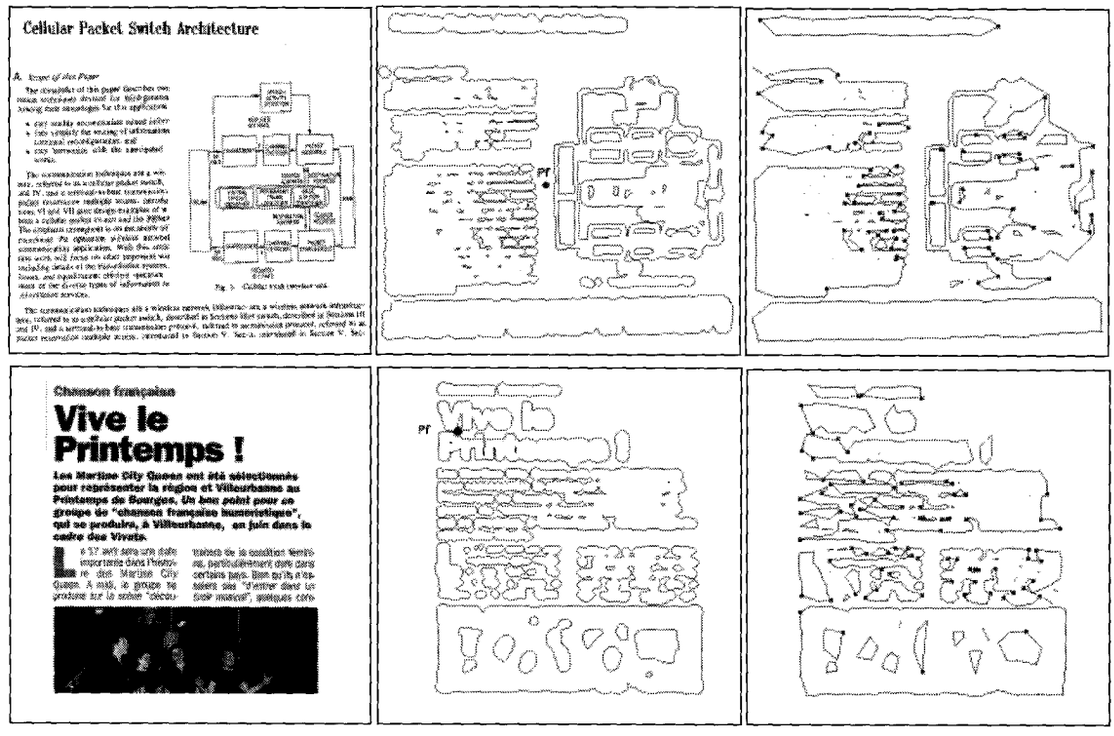


Figure 18. – Représentation des points de forte courbure à partir de l'approximation polygonale des contours des blocs.

de ce fait plus souvent l'*attention* induit un déséquilibre dans le parcours. Certaines zones entières ne sont jamais parcourues en fin de parcours. Aussi en l'absence de but immédiat de la part de l'observateur, nous avons choisi de rééquilibrer le parcours global de l'œil sur le document, en repositionnant à chaque fois qu'il est nécessaire les points de fixation dans les secteurs non visités. La segmentation finale s'en trouve améliorée. Pour cela, nous comparons dans les régions isolées les formes présentant des caractéristiques de surface, de symétrie de compacité ou de courbure et choisissons celle qui maximise les critères présentés au début de la section 4.2.

A partir de la présentation des zones d'intérêt sur le document, nous allons aborder la dernière étape de notre système : la reconstruction des formes et la convergence vers une représentation segmentée.

5. reconstruction des formes

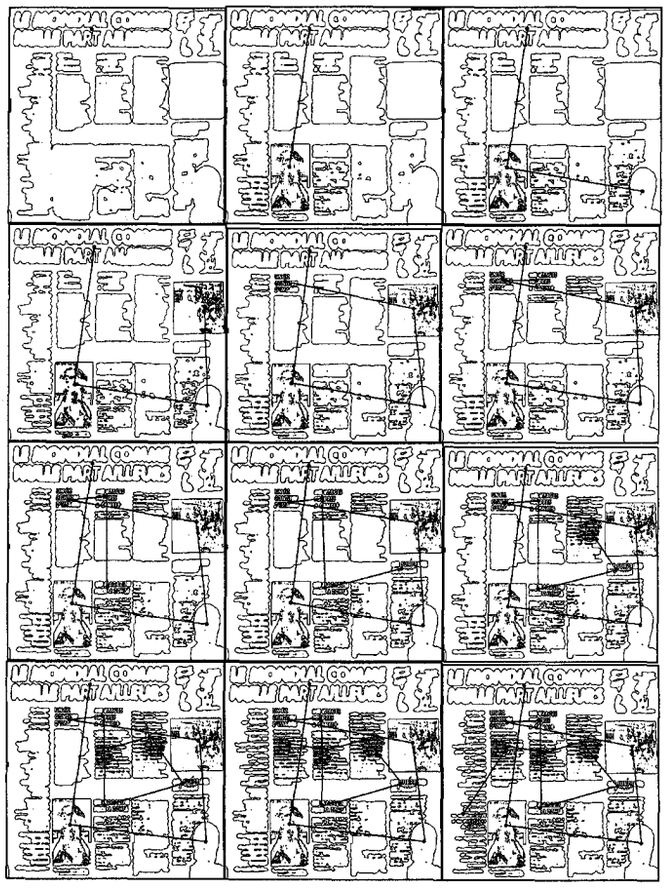
5.1. résultats sur deux exemples

Au fur et à mesure du déplacement du regard sur le document, le nombre de représentations du document augmente. Il est donc nécessaire, dans notre application, de disposer d'une représentation *synthétique* et globale de l'ensemble de l'information perçue par ces touches successives (points de fixations). Cette représentation est le résultat de l'intégration des mises à jour successives. Nous allons ainsi disposer de deux représentations complémentaires : une représentation issue de l'intégration des données à *haute résolution* et une représentation issue de l'intégration à *basse résolution*. A ce stade de l'analyse, nous n'avons utilisé comme données discriminantes, que les contours des objets. Ils présentent en particulier l'intérêt d'être plus facilement quantifiables (information binaire de contours) et manipulables (par des opérations géométriques). Les résultats de ces intégrations à haute et basse résolution sont illustrés sur un exemple, aux figures 19.2 et 19.3. Les résultats sont présentés au bout de 12 itérations, correspondant ainsi à 11 intégrations successives. Nous verrons dans la section suivante quel critère d'arrêt nous permettra de proposer une structuration satisfaisante et d'arrêter ainsi le positionnement de nouvelles fixations.

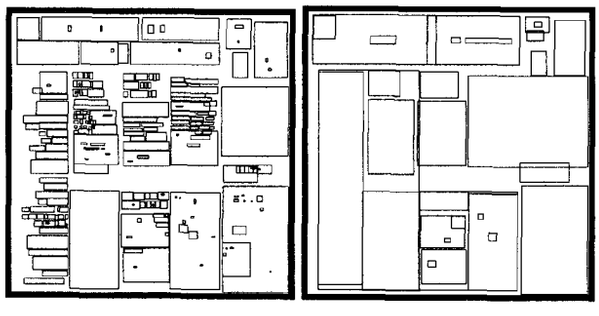
La représentation en blocs qui correspond au résultat de la structuration physique est obtenue à partir de la représentation des contours. Pour cela, on conserve pour chacune des formes la boîte englobante, ce qui ne demande qu'une recherche rapide de composantes connexes sur l'image des contours. Le résultat de cette procédure est illustré sur l'exemple de la figure 19. Parallèlement à cette structuration à haute résolution, on peut ne conserver que les contours de faible résolution et obtenir ainsi une description globale des contours à basse résolution. En ne

représentant que les boîtes englobantes à basse résolution, on obtient la structuration à basse résolution illustrée sur les figures 19.b et 19.c.

Avec une résolution spatiale irrégulière, nous parvenons ainsi à donner aux objets (blocs de textes, d'images) des formes différentes selon les points de vue. La convergence (ou en d'autres termes la reconstruction ou la segmentation de l'image d'origine) est obtenue lorsque le nombre de points de fixation est suffisant pour recouvrir à haute résolution l'ensemble des données considérées comme attractives pour l'œil (de la machine!). A ce stade, on dispose toujours de deux représentations : à haute et à basse résolution.



a)



b) c)

Figure 19 – a) Résultat de l'intégration à haute résolution des contours des formes au bout de 12 fixations. b) Résultat de la segmentation (par découpage en blocs rectangulaires) par intégration à haute résolution. c) Puis à basse résolution

5.2. principe de la reconstruction des formes

Principe de fusion

Le principe de la construction de l'image segmentée est donc le suivant : à chaque étape de mises à jour de l'image (voir figure 19, pour la génération des images intermédiaires), on compare la représentation courante (à partir du point de fixation courant), notée I_i, à la représentation suivante, centrée au point de fixation suivant; celle-ci est notée I_{i+1}. A partir de là, on a deux possibilités : soit on choisit de conserver une information finale à haute résolution (intégration à *haute résolution*), soit, on choisit de conserver une information à basse résolution qui dans ce cas, produira une représentation segmentée plus grossière où ne seront pris en compte que les contours périphériques à faible résolution. Dans le premier cas, on compare chaque pixel de l'image courante (I_i) au pixel de mêmes coordonnées polaires de l'image suivante (I_{i+1}). On ne conserve de ces deux valeurs, que celle qui correspond à la résolution la plus grande, ou encore à la distance la plus petite à un des deux points de fixation. La résolution utilisée est notée $R_p(x, y)$ en un point $M(x, y)$ distant du point de fixation courant $P(x_p, y_p)$; elle correspond à l'acuité visuelle. Cette fonction résolution s'exprime comme l'inverse de la distance de P à M. Plus la distance au point de fixation courant augmente, c'est-à-dire $|PM|$ augmente, plus la résolution $R_p(x, y)$ diminue.

$$R_p(x,y) = \frac{1}{1+\sqrt{((x-x_p)^2+(y-y_p)^2)}} \in [0,1]$$

Pratiquement, autour du centre de fixation, la valeur de $R_p(x, y)$ est maximale (vaut 1) et en périphérie la valeur décroît jusqu'à 0. Pour une segmentation à haute résolution, on retient alors pour une image de taille $n * n$:

$$\mathrm{Max}_{(x,y)(\{1,(n-1)*(n-1)\}}[I_i(R_p(x,y)), I_{i+1}(R_p(x,y))].$$

Pour la segmentation à faible résolution, on conserve pour chaque pixel le minimum des deux valeurs. A un instant donné de l'intégration, on dispose ainsi d'une représentation caractéristique de la *cinématique* du déplacement dans son rapport temps-espace. Finalement, l'image résultante (issue d'une série d'intégrations) est jugée satisfaisante lorsque le *taux de convergence* est atteint. Il est défini dans la section suivante.

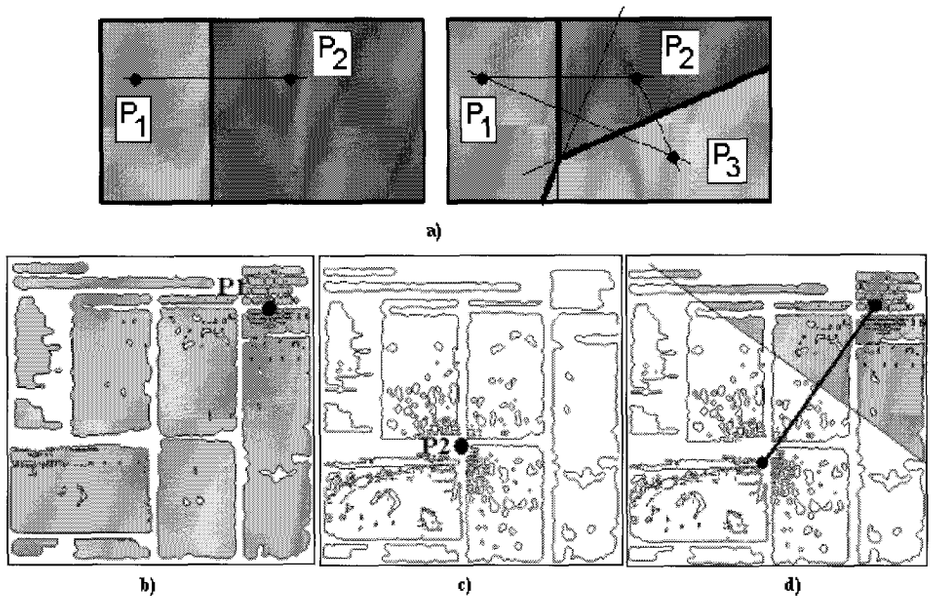


Figure 20. – a) Construction du pavage de Voronoï par intersection de demi-plans. b) Exemple de résultat sur l'image des contours avec une fixation en P1. c) Image des contours avec une fixation en P2. d) Résultat de la fusion par pavage de Voronoï.

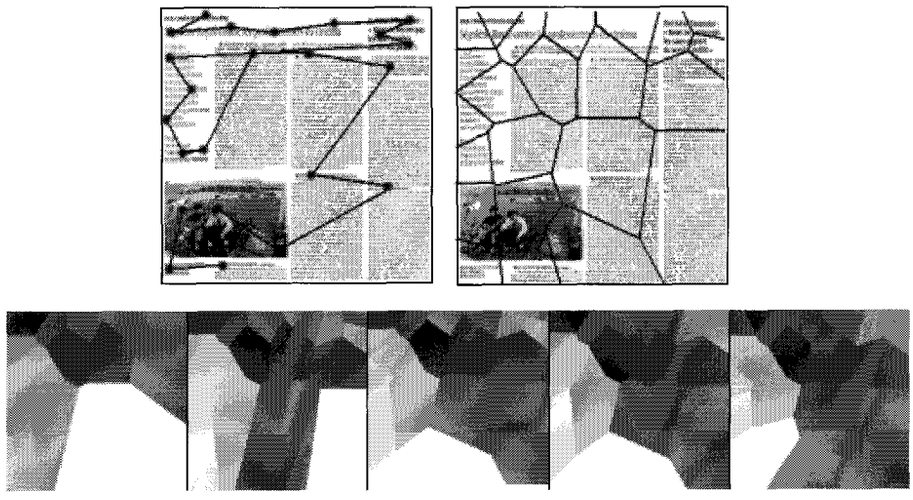


Figure 21. – Constitution du pavage de Voronoï sur un document.

Construction géométrique des régions d'intérêt

D'un point de vue purement géométrique, une autre façon de considérer l'opération de fusion des données consiste à remarquer que le simple tracé de la bissectrice du segment formé par deux points consécutifs P_i et P_{i+1} revient à partitionner le plan en deux régions R_i et R_{i+1}. La région R_i regroupe les pixels les plus proches de P_i, et R_{i+1} regroupe les pixels les plus proches de P_{i+1}. Chaque région correspond à la zone de résolution maximale par rapport au point de fixation à laquelle elle est associée. Au fur et à mesure du processus, on dispose d'un nombre de plus en plus grand de points de fixation, et donc d'un découpage en régions de plus en plus fin. La figure 20 applique la construction d'un pavage de Voronoï sur un document, pour deux puis trois points de fixation.

A chaque itération, le polygone de Voronoï du $i^{\text{ème}}$ point de fixation P_i (ou germe) est obtenu comme l'intersection des demi-plans contenant P_i et délimités par les médiatrices de l'ensemble des (P_iP_j). Le principe de construction du partitionnement en polygones de Voronoï est décrit dans [Chassery]. Le positionnement des *germes* successifs permet ainsi de mettre en évidence les zones fortement attractives. Relier ces germes, ou ces centres de fixations nous permet ainsi de visualiser le trajet effectué par le système dans sa recherche d'information. Ce trajet est représentatif de la dynamique du parcours. La figure 21 illustre un parcours

à partir de la numérotation des germes successifs. Les opérations d'intégration ont été simplifiées par l'utilisation des pavages de Voronoï : ils permettent de ne considérer à chaque itération qu'une portion du document entraînant de ce fait un gain de temps de traitement. Ces résultats mettent en évidence des zones fortement attractives à partir des germes.

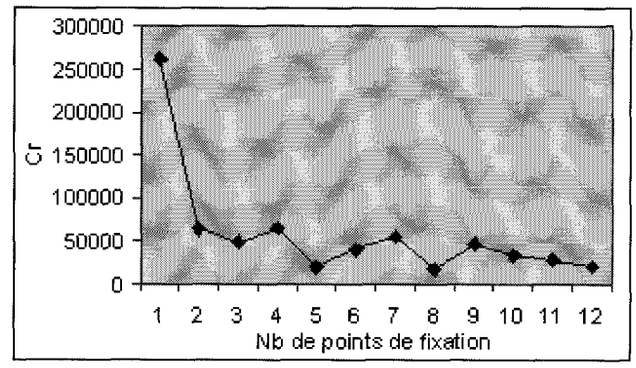
Convergence vers une représentation unifiée

Le principe de l'intégration consiste à définir un taux de convergence à partir de la succession des mises à jour. Ce taux est basé sur le calcul des cumuls des différences des valeurs de résolution pour chaque point de deux mises à jour successives I_i et I_{i+1}. Ainsi lorsque deux mises à jour sont très différentes (par exemple dans le cas où deux points de fixation ont été positionnés à deux endroits très éloignés l'un de l'autre), le cumul des différences est très important. C'est généralement le cas au tout début des itérations. En revanche, lorsque les points de fixation sont très nombreux et tendent à se regrouper dans certaines régions du document, les différences sont très faibles, induisant ainsi une valeur de cumul très faible et donc un taux de convergence (voir la définition ci-dessous) convergeant rapidement. Pour une image de taille $N * N$, on définit alors le taux de convergence de la manière suivante :

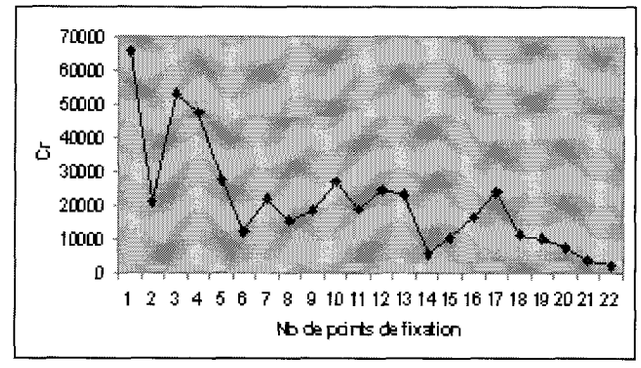
$$C_r = \lim_{i \in \{1,n\}} \left(\sum_{v \in \{1,(N-1)*(N-1)\}} |I_i(R_p(v)) - I_{i+1}(R_p(v))| \right)$$

La valeur de résolution en un point du champ visuel est évaluée comme l'inverse de la distance au centre du champ (point de fixation courant). La norme utilisée dans la définition de $R_p(v)$ correspond à la norme euclidienne classique (voir définition précédente).

En fonction des résultats obtenus par les différences successives, on décide d'arrêter la processus de recherche de points de fixation, lorsque la valeur C_r n'évolue plus de manière significative. C_r tend alors vers une valeur constante, minimale, voir figure 22. Dans la pratique, on arrête le processus lorsque le nombre de points modifiés correspond à un pourcentage du nombre de points total. Dans notre système, nous avons considérés une valeur de 5%. A partir de là, la représentation de l'image traitée à haute résolution (voir figure 19.2) maximise le rapport de l'information visuelle essentielle (celle que nous localisons à l'aide des fixations successives) sur l'information totale (toute l'information à haute résolution de l'image). Pour cette intégration, nous obtenons ainsi une représentation qui met en évidence les zones de focalisation de l'attention (haute résolution des contours), tout en conservant une information globale sur les éléments n'ayant pas été directement ciblés (basse résolution des contours). Dans le cas particulier de cet exemple, où seulement six points de fixation ont été positionnés, la convergence n'est pas complètement aboutie. Cependant, l'image étant de petite taille, il n'est pas nécessaire d'en positionner beaucoup plus. La figure 22 présente le résultat



a)



b)

Figure 22. – a) Représentation du graphe de convergence correspondant aux 12 intégrations successives de l'image de la figure 19.a. b) Convergence correspondant aux 22 intégrations de l'image de la figure 21.

de la convergence au bout de 12 itérations sur l'image test de la figure 19.1. Si le nombre de points de fixation n'est pas suffisant, ou si leur localisation n'est pas très significative (points placés au hasard), la représentation finale n'est plus représentative d'un taux de convergence informatif.

A ce niveau d'analyse, nous pouvons utiliser le découpage en pavés de Voronoï pour simplifier le calcul de la convergence. Plutôt que de recalculer à chaque itération une nouvelle valeur de cumul des différences, on constate à la figure 21 qu'il suffit de considérer à chaque nouvelle mise à jour une portion de plus en plus réduite du document sur laquelle s'effectueront les cumuls de différences. Compte tenu du principe d'intégration par pavage de Voronoï, seule une partie de l'image « nouvellement » intégrée a changé par rapport à la précédente intégration. C'est donc sur cette partie que figureront les différences entre une mise à jour I_i et la suivante I_{i+1}. Sur les portions laissées blanches, les différences sont nulles d'une mise à jour à la suivante. Le calcul du *taux de convergence* se simplifie ainsi au fur et à mesure des itérations. On pourrait donc proposer une définition du taux de convergence basé cette fois sur l'évaluation des surfaces (région blanche de la figure 21) sur lesquelles portent les différences d'une mise à jour à la suivante.

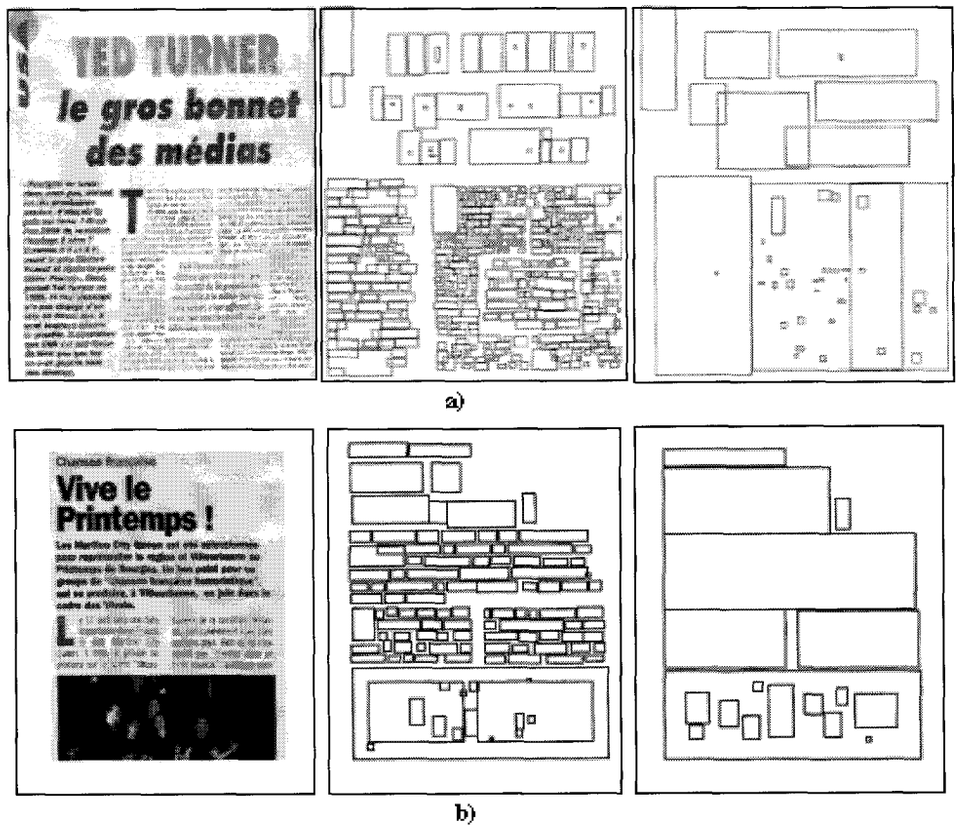


Figure 23. – a) et b) Résultats de découpage à haute et basse résolution.

6. analyse critique et validation

La base de documents qui a été testée est issue d'une sélection qui traduit une grande variété de mise en formes et de thématiques. Cependant, nous avons privilégié les documents contenant des données homogènes : images, textes et graphiques ne sont pas mélangés. Pour les données textuelles, nous avons écarté les inclinaisons ainsi que tous les textes non latins.

6.1. pertinence des découpages en blocs

La validation de ces travaux va porter sur deux points : la pertinence du découpage et la pertinence du trajet. Nous allons tout d'abord vérifier à partir des exemples testés que les différents blocs d'information du document ont bien été séparés. Nous allons attacher une attention particulière à vérifier la séparation entre les blocs d'images, de graphes et de texte mais également entre les blocs de textes de typographie différente : séparation entre les titres et les paragraphes, entre les paragraphes et les notes d'en-tête et de pied de page, et entre plusieurs paragraphes. La figure 23 présente quelques résultats de segmentation et met en évidence les cas de découpages difficiles, concernant notamment les cas de sous-segmentation et de sur-segmentation à basse résolution. La sous-segmentation peut avoir deux origines différentes. La première provient d'une particularité de la mise en forme du document qui ne respecte pas l'interlignage standard lié au type de typographie utilisé. Dans ce cas, deux zones de texte mitoyennes de polices différentes peuvent fusionner et ne former qu'un seul bloc (cas de l'exemple 23.a). La deuxième origine, la plus fréquemment rencontrée, est liée à notre principe de fusion à basse résolution de blocs. Lors des intégrations successives à basse résolution, la description des différentes régions converge vers un découpage de plus en plus global. Il arrive que ce découpage soit à l'origine du regroupement de deux blocs de natures différentes (cas typique d'un titre et d'un paragraphe, ou d'une lettrine et d'un paragraphe (cas des figures 23.a et 23.b).

La sur-segmentation se retrouve généralement dans les titres ou les régions comportant de grands caractères sur de petites surfaces. Les blocs au lieu de fusionner restent bien disjoints. Les opérations de filtrage périphérique ne permettent pas toujours de les regrouper.

Afin de limiter les effets de la sous-segmentation (et de la sur-segmentation) sur le label des blocs, nous avons choisi quelques pistes de recherche pour définir des heuristiques liées à une information de texture des blocs et sur lesquelles nous reviendrons dans la conclusion de cette étude. Nous envisageons également d'utiliser la complémentarité des descriptions à haute et basse résolution pour gérer les situations de conflit. Cette partie est également en cours d'étude.

6.2. analyse du trajet oculaire et vérification expérimentale

L'objectif de ces travaux n'est pas de simuler un parcours réel (car il en existe autant que d'observateurs, et tous très différents les uns des autres) : nous cherchons uniquement à localiser les régions présentant un intérêt visuel, ce qui nous conduit à construire un parcours qui apparaît comme un parcours possible. Nous allons donc proposer dans cette partie une manière de valider l'ordre de lecture que nous proposons. Le plus intéressant, pour ce faire, est de comparer cet ordre au trajet oculaire d'un lecteur humain sur le même document. Pour cela, nous avons exploité des mesures oculométriques sur des observateurs humains. Compte tenu des conditions expérimentales assez difficiles (mesures très sensibles au bougé de la tête, à la couleur de l'iris, aux conditions extérieures d'éclairage...) qui nous étaient imposées pour la prise des mesures, nous ne disposons que d'environ quatre cent résultats obtenus sur vingt documents. Pour les illustrer, nous présentons, sur la figure 24, des exemples de parcours réels, choisis pour leur ressemblance avec nos parcours simulés. La consigne donnée aux observateurs consistait (comme nous l'avons présenté dans l'introduction) à regarder sans lire les régions du document qui avaient un fort pouvoir d'attraction. Les images présentées étaient en niveau de gris (et pas en couleur). La durée totale du survol était variable d'un sujet à l'autre. Aucun temps limite n'était imposé; seule la lecture mot à mot était défendue.

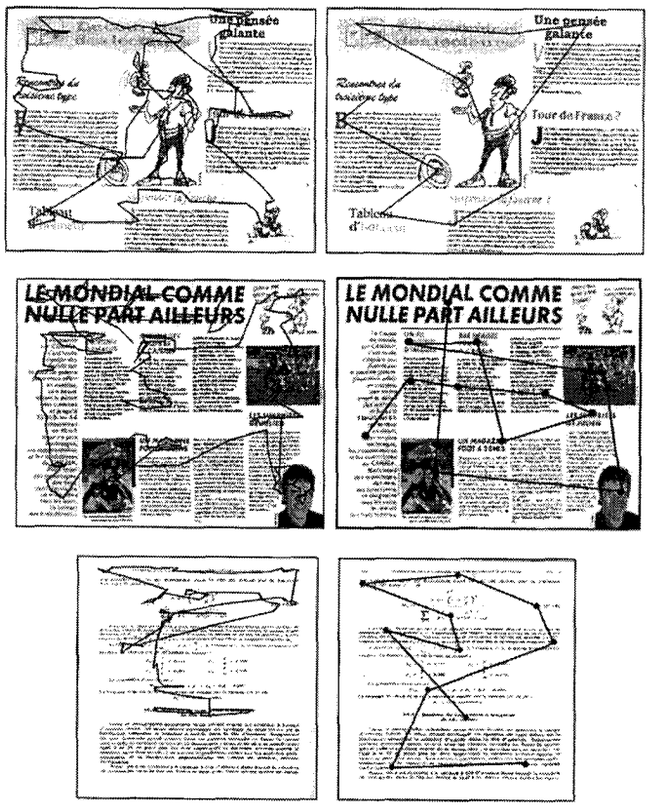


Figure 24. – Comparaison entre des trajectoires réelles d'un observateur humain choisies pour leur ressemblance avec des trajectoires simulées par notre système.

Le fait qu'à ce stade, nous ne disposons d'aucune information liée au contenu des blocs (lié à un étiquetage catégoriel des composantes homogènes du document) peut paraître pénalisant. Or, les résultats proposés à partir d'un paramétrage géométrique lié aux contours des formes sont satisfaisants : ils mettent notamment en évidence les régions riches et denses et privilégient les formes de grande taille. Notre étude peut donc être vu comme une première étape d'un processus complet faisant coopérer les informations des formes géométriques de bas niveau (que nous avons présentées ici) et des données de plus haut niveau liées à une analyse de texture des blocs. Cette coopération est actuellement en cours d'étude.

7. conclusion et perspectives

Conclusion

L'expérience commune a longtemps suggéré que l'homme percevait passivement son environnement, dont l'image s'imposait à lui sans qu'il puisse avoir sur elle aucune action. Les recherches en psychophysiologie ont ainsi pu émettre quelques doutes sur le bien fondé de cette réflexion, en proposant des mécanismes de capture d'informations intégrant directement l'activité perceptive et décisionnelle de l'observateur. On considère aujourd'hui la perception comme une conduite, où l'information perçue est une véritable *construction*, un ensemble d'informations sélectionnées et structurées en fonction des besoins, des intentions du sujet impliqué activement dans une situation perceptive donnée. Il en résulte notamment que la perception rétablit une sorte de continuité temporelle et spatiale nécessaire à la construction ou la « re-construction » de l'objet perçu.

Nous nous sommes donc très largement inspirés de ce constat pour mettre en place les bases d'une méthodologie de structuration des documents imprimés prenant le comportement exploratoire de

l'homme comme modèle de référence. Nous avons plus spécifiquement cherché à simuler ce comportement dans une situation élémentaire d'exploration correspondant au *survol* du document. L'intégration des données que nous proposons et qui est issue des fixations successives est à la base de la reconstruction des formes. Le choix des fixations est défini à partir d'une caractérisation géométrique de bas niveau des contours des objets. Il illustre autant que possible certaines propriétés de *bas niveau* de vision humaine liés aux éléments de contraste et à leur organisation (symétrie, courbure, compacité, proximité...). La *segmentation physique* que nous obtenons renseigne sur la localisation spatiale de l'information perçue (localisation des blocs circonscrits aux contours), sur les caractéristiques géométriques de ces blocs, ainsi que sur les relations de proximité qu'ils ont les uns avec les autres (alignement, chevauchement, adjacence...).

On a pu constater que la plupart des méthodes de traitement d'images perçoivent généralement « passivement » leur environnement en considérant l'ensemble des données à un niveau équivalent. Or, comme nous l'avons souligné dans cet article, un document possède une structure et des niveaux d'informations bien hiérarchisés. Qu'il s'agisse ainsi d'un titre ou d'une note de pied de page, l'information n'est pas perçue de la même manière. On peut ainsi affirmer que toute l'information du document n'a pas le même poids *sémantique*. Des résultats expérimentaux de suivi de regard nous ont notamment permis de mettre en évidence que les variations typographiques liées à la mise en forme matérielle du document sont de première importance. Elles traduisent en effet l'intention de l'auteur et son effort de mettre en relief les informations essentielles.

Nouveaux développements et perspectives

Il nous est donc paru assez naturel à la suite de ce travail de passer à une recherche de structuration *logique*, ou plutôt de structuration *fonctionnelle* mettant en évidence cette hiérarchie constitutionnelle et répondant aussi aux préoccupations de certains chercheurs tels Doerman qui dans [Doerman] parle de niveau *fonctionnel* d'analyse. A ce niveau, c'est non seulement la localisation spatiale du bloc (niveau physique) qui est informative mais également son identité catégorielle. Ces deux sources d'informations permettent une description fonctionnelle des éléments.

Les développements futurs que nous envisageons portent enfin sur la mise en place de nouvelles stratégies d'exploration pour une recherche d'informations particulières dans le document. Ce travail s'inscrit dans le contexte plus général de l'indexation des documents (par leur mise en forme ou leur contenu) impliquant davantage l'homme par sa volonté de comprendre et de trouver. Une première piste que nous commençons à exploiter consiste à « simuler » une stratégie de *lecture* par la recherche de mots-clés dans les zones d'intérêt du document. Des traitements de plus « haut » niveau liés à l'analyse sémantique des textes seraient ainsi indispensables pour amorcer la reconnaissance des mots.

8. remerciements

Nous tenons à remercier toute l'équipe du laboratoire CLIPS-IMA de Grenoble pour leurs multiples interventions et leur grande disponibilité dans l'élaboration de ce travail. Nos remerciements sont plus particulièrement destinés à Solange Hollard et Laurent Aublet-Cuvelier sans qui toute la partie expérimentale de ce travail n'aurait pu voir le jour. Leur gentillesse et leurs conseils pour l'ensemble des expérimentations nous ont été d'une aide très précieuse.

BIBLIOGRAPHIE

[Adelson] E.H. Adelson, J.R. Bergen, «Early Vision», *Computational models of visual processing*, Michael S.Landy, J.A.Movskon, 1991, p.36-45.

[Akindele] O.T. Akindele, A. Belaïd, «A labeling approach for mixed Document Blocks». *Proceedings of the Second Int. Conf. On Document Analysis and Recognition*, 1993, vol.4, pp.749-752.

[Amamoto] N. Amamoto, S. Torigoe, Y. Hirogaki, «Block segmentation and Text Area Extraction of vertically/ Horizontally written document», *ICDAR'93*, vol.4, 1993, pp.739-743.

[Baird90] H .S. Baird, S.E. Jones, S.J. Fortune, «Image segmentation by shape-directed covers», *International Conf. On Document Analysis and Recognition*, 1990, pp.820-825.

[Baird92] H.S. Baird, H. Bunke, K. Yamamoto, «Structured document analysis», Springer, 1992.

[Barbara] M.O. Barbara, M. Mojahid, J.Vivier, «Mise en forme matérielle des textes de consignes et repérage d'informations», *Colloque National sur l'Ecrit et le Document CNED'96*, 1996, pp.229-236.

[Bloomberg] D.S. Bloomberg, «Multiresolution morphological approach to document image analysis», *First Int. Conf. on Document analysis, ICDAR'91*, vol. 2, 1991, pp.963-971.

[Bonnet] C. Bonnet et B. Dresp. «Psychophysique de l'extraction des contours en vision humaine», *Reconnaissance de Formes et Intelligence Artificielle 3*, 1991, 102-109.

[Bruce] Bruce, V., Green, P.R. *La perception Visuelle : Physiologie, psychologie et écologie*. Grenoble : Presse universitaire de Grenoble, 1993, 431p.

[Chassery] J.M. Chassery, M. Mlekemi, «Segmentation d'images en diagramme de Voronoï. Application à la détection d'événements en imagerie multi-sources», 7ème *Congrès Reconnaissance de Formes et Intelligence Artificielle*, Paris, 1989, pp.781-790.

[Crettez] J.P. Crettez, «Modélisation des voies visuelles primaires, premières étapes de la perception des Formes «, *Thèse de Doctorat*, 1984, 242p.

[Crowley] J.L. Crowley, R.M Stern, «Fast Computation of the Difference of Low-Pass Transform». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, vol.6, pp.212-222.

[Déforges] O.Déforges, «Segmentation robuste d'images de documents par une approche Multirésolution», *Thèse de Doctorat*, 1995.

[Doermann] D. Doermann, A. Rosenfeld, E. Rivlin, «The function of documents», *Fourth Int. Conf. on Document analysis, ICDAR'97*, vol.2, Ulm, 1997, pp.1077-1081.

[Eglin97] V. Eglin, H. Emptoz. «Low-resolution boundaries for guiding eye-movement on a document», *In Proceedings of the fourth International Workshop on Visual Form*, Capri, Italy, 1997, pp.178-187.

[Eglin98] V. Eglin, S. Bres, H. Emptoz, . «Printed Text featuring using visual criteria of legibility and complexity». *Proceedings of the 14th International Conference on Pattern Recognition*, Brisbane (Australie), août 1998, pp.942-944.

[Fletcher] L.A. Fletcher, R. Kasturi, «A robust algorithm for text String Separation from mixed Text/Graphics Images», *IEEE Trans. On PAMI*, vol.10, N° 6, 1988, pp.910-918.

[Hacisalih] S.S. Hacisalihazade, L.W. Stark, J.S. Allen, «Visual Perception and Sequence of Eye Movements Fixations», *IEEE SMC*, vol.22, N° 3, 1992, pp.474-480.

[Ishitani] Y.Ishitani, *Document layout analysis based on emergent computation*, vol.1, pp.45-50, 1997.

[Kosslyn] S.M. Kosslyn, Image and brain : the resolution of the imagery debate. Cambridge, MA : MIT Press, 125p., 1994.

[Lecas] J.C. Lecas. *L'attention visuelle, de la conscience aux neurosciences : Problèmes fondamentaux et mécanismes de la perception visuelle*. Liège : Pierre Mardaga, 1992, 310p.

[Lévy] A. Lévy-Schoen, «Exploration et connaissance de l'espace visuel sans vision périphérique; quelques données sur le comportement oculomoteur de l'adulte normal», *Journal Psychologique*, 1976, vol.39, n° 1, pp.77-91.

[Likforman] L.Likforman-Sulem, C. Faure, «Une méthode de résolution de conflits d'alignements pour la segmentataion des documents manuscrits», *CNED 94, 3ème Colloque National Sur l'Ecrit et le Document*, 1994, pp.265-273.

[Manzanera] A. Manzanera, J.M. Jolion. «Pyramide irrégulière : une représentation pour la vision exploratoire». *Traitement du signal*, 1995, vol 12, n°2, pp.169-176.

[Marr] D. Marr, «Vision». New-york : W.H. Freeman and Co, 1982, 397p.

[Nagy] G. Nagy, M. Viswanathan, «Dual Representation of segmented Technical Documents», *int. Conf. On Pattern Recognition*, 1991, pp.141-151.

[Normand] N.Normand, C.Viard-Gaudin, «A background based adaptation page segmentation algorithm«, *Thirth Int. Conf. on Document analysis, ICDAR'95*, 1995, pp.138-141.

[Ogier] J.M. Ogier, R. Mullot, J. Labiche, Y. Lecourtier, «Interprétation de document par cycles«perceptifs» de construction d'objets cohérents. Application aux données cadastrales». *3ème Colloque National Sur l'Ecrit et le Document*, Rouen, 1994, pp.167-184.

[O'Gorman91] L. O'Gorman, «Subsampling Text images», *First Int. Conf. on Document analysis, ICDAR'91*, 1991, pp.219-227.

[O'Gorman93] L. O'Gorman, «The document spectrum for page layout analysis», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, 15(11) :1162-1173.

[Pavlidis] T. Pavlidis, J. Zhou, «Page Segmentation by white streams», *First Int. Conf. on Document analysis*, ICDAR'91, 1991, pp.945-953.

[Sankar] P.V. Sankar, E.V. Krishnamurthy, «On the compactness of subsets of digital pictures», *CGIP*, vol.8, 1978, pp.403-412.

[Shah] S. Shah, M. Levine, «Visual Information Processing in Primate Cone Pathway - Part I : A Model». *IEEE Transactions on PAMI*, 1996, vol. 26, n° 2, pp.259-273.

[Tang] Y.Y. Tang, C.Y. Suen, «Document structures : a survey». *Proceedings of Second ICDAR*, Montréal (Canada), 1993, vol.1, pp.99-102.

[Treisman] A. Treisman,«L'attention, les traits et la perception des objets», Folio Gallimard, 1992, pp.154-191.

[Tsotsos] J.K. Tsotsos, «The complexity of perceptual search tasks «. *In Eleventh International Joint Conference on Artificial Intelligence*, 1989, pp.135-160.

[Tsujimoto] S. Tsujimoto, H. Asada, «Major components of a complete text reading system», *in Proc. of the IEEE PAMI*, vol.80, n° 7, 1992, pp.1133-1149.

[Yamamoto] H. Yamamoto. «An active Foveated Vision System : Attentional Mechanisms ans Scan Path Covergence, Measures», *Computer Vision and Image Understanding*, vol. 63, N°1, 1996, 50-65.

[Watanabe] T. Watanabe, Q. Luo, N. Sugie, «Structure recognition methods for various types of documents». *Machine Vision and Applications*, 1993, 6(2-3) :163-176.

[Wertheimer] M. Wertheimer, *Untersuchungen zur Lehre von der Gestalt, II*. Psychologische Forshung, 4, 1923, 301-350.Traduit par «Laws of organisation in perceptual forms» in W.D. Ellis (1995). A source book og Gestalt psychology. London : Routledge and Kegan Paul.

[Wieser] J. Wieser, A. Pinz, «Layout analysis : finding text, titles and photos in digital images of newspaper pages». *Proceedings of the Second ICDAR*, 1993, vol.4, pp.774-77.

[Wilson] S.W. Wilson, «On the retino-cortical mapping«, *Int. J. Man-Machine Stud.*, 1983, vol.18, pp.361-389.

[Wong] K.Y. Wong, R.G. Casey, F.M. Wahl, «Document Analysis System «, *IBM Journal of Research and Development*, vol.25, N° 6, 1982, pp.647-656.

Manuscrit reçu le 21 juillet 1998.

LES AUTEURS

Véronique EGLIN

Véronique Eglin est ingénieur et docteur de l'Institut National des Sciences Appliquées de Lyon (1998). Depuis 1998, elle est attachée temporaire d'enseignement et de recherche (ATER) au laboratoire Reconnaissance de Formes et Vision de l'INSA. Ses recherches portent sur la segmentation et l'analyse des documents, plus particulièrement sur l'utilisation de la perception visuelle et de la multirésolution pour la recherche et la caractérisation d'informations.

Stéphane BRES

Stéphane Bres est ingénieur de l'Institut National des Télécommunications (INT) d'Evry (1988) et docteur de l'INSA de Lyon (1994). Il est maître de conférence du département informatique de l'INSA de Lyon depuis 1995 et enseigne le traitement du signal et l'analyse numérique. Ses recherches, au sein du laboratoire Reconnaissance de Forme et Vision, porte actuellement sur l'indexation d'images dans le cade de projets régionaux et européens.

Hubert EMPTOZ

Hubert Emptoz est professeur à l'INSA de Lyon où il dirige le laboratoire Reconnaissance de Formes et Vision. Il s'est intéressé à de nombreux aspect de la reconnaissance, notamment aux méthodes statistiques et à la théorie de l'information, à l'approche prétopologique (introduite dans sa thèse d'état, en 1983) et aux applications en neurophysiologie. L'essentiel de son activité actuelle est consacrée au domaine de l'écrit et du document.