

Interprétation de documents techniques par « cycles perceptifs » à partir d'une perception globale du document

Interpretation of Technical Drawings Using Perceptive Cycles Starting with its Global Perception

par Jean-Yves RAMEL*, Nicole VINCENT**, Hubert EMPTOZ*

* Reconnaissance de Formes & Vision
Bât. 403 - INSA de Lyon
20, av. A. Einstein 69621 Villeurbanne Cedex (France)
[ramel,emptoz]@rfv.insa-lyon.fr

** LI / E3i
Université François Rabelais
64, avenue Jean Portalis 37200 Tours (France)
vincent@univ-tours.fr

résumé et mots clés

Nous présentons, dans cet article, un système de lecture automatique de documents techniques. Les dessins techniques, au sens où nous l'entendons par la suite, sont des documents essentiellement graphiques (constitués de traits, de formes pleines, de texte, ...) produits par un dessinateur, à la main ou avec l'aide d'un système d'édition. L'architecture que nous avons choisie et développée permet d'extraire l'information graphique et alphanumérique du document pour reconstituer la sémantique associée aux différents objets du dessin et à leurs relations. En nous inspirant des connaissances acquises sur les mécanismes de la vision humaine, nous avons élaboré une méthode comportant deux phases :

- Durant une première phase de perception globale, une représentation initiale du document est obtenue à l'aide d'un nombre limité de primitives : le Pixel, le Vecteur, le Quadrilatère. Le choix de ces primitives permet à cette représentation de s'adapter à la description de documents techniques de types variés (plans, cadastres, partitions musicales, ...). Cette représentation, bien que fondée sur des primitives élémentaires, fournit une description précise de toutes les formes du document initial (image binaire). Outre leur adéquation aux formes à décrire, ces primitives constituent la base de la structure de données qui nous permettra de définir une représentation efficace des connaissances acquises sur le document durant l'analyse.

- Ensuite, l'utilisation duale d'un système multi-spécialistes et de la représentation préalablement construite permet, au cours d'une phase de lecture et d'interprétation, d'utiliser les données fournies par la vision globale pour réaliser les analyses locales. Un mécanisme de "cycles perceptifs" permet aux différents spécialistes de coopérer pour réduire au maximum les erreurs d'analyse et aboutir à une interprétation plus robuste. Chaque spécialiste réalise une tâche précise telle que, par exemple, la détection du texte, des courbes, des éléments mécaniques ou la reconstruction 3D du mécanisme. Tous utilisent la représentation construite pour focaliser leur attention sur des zones précises de l'image qu'ils étudient en détail. En fin de chaque cycle, la complétude de la compréhension du document est vérifiée et un nouveau cycle d'activation des spécialistes approfondit, si nécessaire, l'interprétation. Cette phase de lecture progressive a été mise en place pour le cas particulier des plans cinématiques mécaniques. Le système obtenu comporte alors six spécialistes et s'accorde parfaitement au traitement de ce type de document. Les exécutions continues et entrelacées des différents spécialistes permettent de réaliser des analyses contextuelles par segmentations successives augmentant ainsi l'interaction et l'efficacité du système.

Grâce à cette démarche, nos travaux se distinguent de ceux employant la chaîne habituelle de traitements : pré-traitements - segmentation - analyse - reconnaissance. Notre système s'efforce, dans la vision globale, d'extraire des connaissances puis, à partir des informations possédées sur le document, d'extraire des formes de manière hiérarchique et réfléchie, et il les transcrit dans une représentation évolutive décrivant à chaque instant l'état du contexte d'interprétation.

Interprétation d'images, documents techniques, perception, extraction de primitives, représentation des documents.

In this paper, we propose a new method for automatic understanding of technical documents. In the case we are concerned with, technical documents are essentially constituted of graphics (lines, filled shapes, text, ...) and have been drawn either by hand or with the help of an editor. The chosen architecture allows to extract, from the document, the graphic and alphanumeric information in order to rebuild the semantic associated to the different objects in the drawing and to their relationships. To realize this work, we have tried to automatise natural techniques normally associated with human perception.

- A first stage of global perception of the document is realized, using a limited number of elementary primitives: the Pixel, the Vector and the Quadrilateral, to obtain a first representation of the document. This first representation allows to depict correctly several kinds of technical documents (diagrams, cadastral maps, musical scores, ...). This representation will provide a precise description of all the shapes from the original document (binary images) along the different levels of the analysis.

- Next, the dual use of this representation of the document and of a multi-specialist system allows to improve the local analysis during all the reading and understanding steps. A mechanism of «perceptive cycles» allows the different specialists to cooperate. Each specialist possesses some elementary, precise and sufficient information in some particular domains: extraction of text, curves, mechanical elements and it uses the global vision to focus its attention on some places of the image. The hypotheses that the specialists interjected are verified and induce the evolution of the representation of the document. This step of progressive reading has been tested in the particular case of kinetic diagrams. The entire system, using six different specialists, seems to be perfectly suited to this kind of document.

This method is quite different from those using classical procedure: pre-processing - segmentation - analysis - recognition. Every step of the process brings some new information that is used either to confirm a previous hypothesis or states a new one until the total understanding of the document is achieved.

Image understanding, perception, technical documents, feature extraction, representation of drawings.

1. introduction

Le dessin reste toujours le moyen préféré de communication entre concepteurs, réalisateurs et utilisateurs. Mécaniciens, électroniciens et architectes souhaiteraient disposer d'une méthode simple et fiable de lecture et d'interprétation des plans qu'ils préfèrent toujours réaliser d'abord sur papier laissant libre cours à leur imagination. L'objectif de l'interprétation automatique de documents est alors de faciliter l'exploitation de ces dessins manuscrits en s'appuyant sur un ensemble d'étapes permettant la transformation du schéma papier en un document numérique qui puisse être interprété automatiquement.

Beaucoup de travaux concernent l'analyse de documents imprimés tels que les sommaires de périodiques, les pages de journaux, les partitions musicales [Martin90], ... mais très peu traitent de l'interprétation des documents techniques [Ts95]. Les difficultés qui proviennent de la diversité des types d'informations (texte, traits, formes pleines, ...) constituant ces documents, de leur manque de normalisation (grande liberté du dessinateur) ou de la nécessité d'effectuer finalement la reconstruction 3D du mécanisme représenté, expliquent probablement l'intérêt limité porté à ces documents.

C'est dans ce contexte que nous avons conçu et mis en œuvre une méthode qui, s'inspirant des mécanismes de la perception humaine, conduit à une interprétation automatique des plans, qu'ils soient imprimés ou manuscrits. Les expérimentations auxquelles nous nous sommes livrés en utilisant des méthodes classiques, ont confirmé l'hypothèse que les traitements de bas niveau entraînent le plus souvent des pertes d'information significatives et restent donc à améliorer [Kasturi90]. De plus, les systèmes utilisant un

contexte global pour l'interprétation, dès les premières étapes du traitement, sont encore très rares [Dori95]. Cette étude constituera la première partie de cet article.

Ensuite, nous présenterons le système de lecture automatique mis en place pour le cas particulier des schémas cinématiques. Ce système opère selon 2 phases distinctes : l'une dite de **perception globale** permettant l'acquisition d'une représentation du document, l'autre dite de **lecture et de compréhension**. Après avoir présenté chacune de ces deux phases, nous décrirons les résultats obtenus par la mise en œuvre des « cycles perceptifs ». Grâce aux résultats présentés tout au long de l'étude, nous montrerons les apports fournis par le système mais aussi son aptitude à s'adapter à l'analyse de toutes sortes de dessins techniques.

2. des démarches humaines à un système de compréhension automatique

2.1. les mécanismes de la perception humaine

Bien que le fonctionnement du système visuel humain soit encore loin d'être complètement compris, il est possible d'en tirer

des enseignements intéressants. Nous sommes tous capables de reconnaître le type d'un document technique (plan, projection orthogonale, organigramme,...) sans en comprendre le contenu, de discriminer le texte du graphique, le manuscrit de l'imprimé, de nous placer très rapidement au niveau de détail approprié à la compréhension et ceci pratiquement sans effort; ces remarques montrent que l'interprétation de ce qui nous entoure ne se réduit pas seulement à une simple mise en correspondance d'images c'est-à-dire à du « Pattern Matching ». Les performances du système visuel humain peuvent être expliquées, d'une part, par sa faculté à utiliser le contexte environnant et les connaissances préalablement acquises, et d'autre part par son mode de fonctionnement; selon B. Julesz [Lecas92], la vision humaine opère selon 2 modes distincts :

- la vision pré-attentive (pré-consciente) qui permet l'acquisition instantanée de l'image, sans scrutation, et en couvrant un large champ visuel (tampon visuel=image entière).
- la vision attentive, limitée à de petits domaines (fenêtre d'attention=zone de focalisation), qui correspond à une scrutation pas à pas avec une attention bien marquée. Pendant cette seconde période, l'activité perceptive peut être représentée par un mécanisme d'oscillation entre 2 phases : l'acquisition d'une image, puis sa confrontation avec l'appris. Ce cycle perdure jusqu'à ce que la compréhension totale de la scène soit obtenue. Kosslyn résume ce fonctionnement par le schéma suivant :

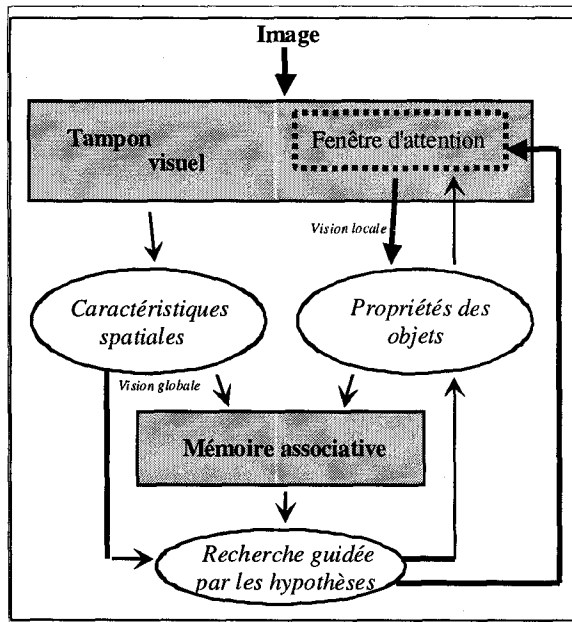


Figure 1. – Organisation fonctionnelle de la perception visuelle selon Kosslyn [Simon93].

2.2. le mécanisme de la lecture

Les études du mécanisme de reconnaissance de la mise en forme matérielle (structure physique des documents), pour la perception humaine, n'en sont qu'à leur début [Clavier97] [Eglin97]. Pour ce qui concerne la lecture de mots ou de phrases, et même si on

ne parvient pas encore à modéliser l'acte de lire pour le transcrire sous forme de règles, on sait maintenant que le lecteur enchaîne sans cesse des activités complexes telles que la prise d'indices (graphiques, lexicaux, syntaxiques, sémantiques), l'élaboration d'hypothèses (permettant d'anticiper) et la réalisation de corrections éventuelles.

Durant la lecture, des échanges s'effectuent constamment entre l'information visuelle qui se limite à l'image brute du document et l'information non visuelle qui utilise un arrière-plan conceptuel qui rassemble ce que l'on sait déjà du document [Bruce93]. L'interface entre les deux est la mémoire associative (ou immédiate), de faible capacité, sorte de file d'attente des prises d'indices dans laquelle la dernière entrée chasse impitoyablement la première de la file.

Ainsi, lorsqu'il interprète un phénomène, l'Homme effectue des focalisations successives, chacune ayant un objectif précis et contribuant à faire évoluer, aussi bien la représentation mentale de la scène en cours que ses connaissances, et à les faire converger vers un tout cohérent.

C'est donc par une approche plus locale de la scène, et en utilisant des connaissances pour cibler les recherches, que sont trouvées les informations précises et pertinentes qui conforteront le raisonnement et contribueront à l'interprétation. Guidé par la vision globale, le système visuel humain explore des points particuliers de la scène. Il est à la recherche d'informations précises, dont la présence est supposée à des endroits donnés et sur lesquelles il possède aussi des connaissances. Durant cette phase d'exploration visuelle, toutes les connaissances acquises auparavant (pendant le cycle et encore avant) par l'individu sont utilisées [Fortin92].

2.3. le schéma classique des méthodes de reconnaissance

2.3.1. analyse du document technique sur la base de sa structure physique

Si l'on excepte le cas des documents composites, on constate que, traditionnellement, en reconnaissance de formes, une distinction est faite entre deux phases successives : la phase de segmentation et la phase de reconnaissance. Cela s'explique par le fait que les techniques de segmentation se veulent souvent universelles donc applicables à différents types de documents tandis que les méthodes de reconnaissance sont plus spécifiques puisqu'elles utilisent les données de leur domaine d'étude.

Les premières tentatives de création de systèmes d'analyse de documents techniques se voulaient générales; elles n'introduisaient que très tard, dans la chaîne des traitements, les connaissances a priori sur la nature du document traité. Elles mettaient en évidence une organisation en couches pour chacun de ces documents [Belaïd92], ce qui constitue la structure physique.

Dans le cas des documents techniques, celle-ci se compose :

- d'une couche graphique qui contient la majeure partie de l'information. On peut parfois la diviser en plusieurs sous-couches selon des critères purement structurels ou géométriques (épaisseur du trait, style du trait, hachures, ...).

- d'une couche texte qui se présente généralement sous la forme de groupes de composantes connexes de petites tailles constituant, soit un mot, soit un nombre, soit une annotation. La position du texte par rapport au graphique indique à quelle partie du dessin se rapporte la cote ou la légende; il est donc important de conserver les positions relatives de ces composantes.

- d'autres couches éventuelles composées, par exemple, des petits symboles (têtes de flèches, disques, ...)

Ainsi, le problème de l'interprétation automatique de documents semblait résider dans la phase de séparation puis de reconnaissance de chacune de ces couches, et dans le choix de l'ordre d'extraction le plus approprié. Cela ne va pas sans difficultés car toutes les couches sont constituées d'objets géométriques, eux-mêmes formés à partir de primitives d'un même type (arcs et segments). Ainsi, ces méthodes ont vite révélé leurs limites; elles étaient, en effet, incapables de présenter, en sortie des modules d'interprétation, un niveau d'abstraction suffisamment élevé.

2.3.2. interprétation se basant sur la structure logique

Les systèmes à base de connaissances ont alors été introduits. Ils étaient adaptés au problème puisqu'ils permettaient d'incorporer des connaissances de natures très différentes; celles-ci pouvaient être de bas niveau (caractéristiques structurelles ou géométriques des objets de l'image) ou de très haut niveau (contraintes de cohérence sémantique qu'un objet doit respecter) [Tombre92]. Les techniques d'interprétation de documents techniques ont ainsi évolué et la décomposition en couches n'a plus été réalisée en fonction de critères géométriques (couches des traits forts, couches des traits fins, ...), c'est-à-dire des traitements à mettre en place, mais plutôt sur des *critères sémantiques*, c'est-à-dire en fonction de la signification des informations fournies par les regroupements judicieux des formes de l'image, en utilisant sa structure logique.

Il est à noter qu'avec ce principe, seules les méthodes d'extraction des primitives de base et les modèles de description des objets constituant les dessins peuvent être communs à tous les systèmes d'interprétation. L'ordre d'extraction des entités dépend, lui, du domaine d'application et non plus de la structuration logique des données en couches superposées.

Les différents documents obéissent le plus souvent à des règles strictes, définies par différentes normes; elles précisent les entités et les relations entre entités qui sont susceptibles d'apparaître. On ne parle plus alors de structure mais plutôt de syntaxe propre au schéma. Ces règles varient selon le type de documents et d'applications auxquels on s'intéresse : les banques de données d'un système de CAO mécanique sont totalement différentes de celles d'un système de simulation de circuits électriques ou même de celles d'un système de CAO d'un autre type. La connaissance de ces différentes règles permet de vérifier la cohérence de la représentation obtenue. C'est en tenant compte de ce type d'information qu'il faut définir, par exemple, l'ordre d'extraction des différentes entités constituant un dessin, ou essayer de lever certaines ambiguïtés concernant l'appartenance d'une primitive à une entité plutôt qu'à une autre. On limite ainsi les incohérences et leur propagation dans les niveaux supérieurs.

2.4. notre approche

Nous nous sommes donc efforcés, d'une part, de ne pas nous contenter de la stratégie usuelle : « segmentation en couches » puis « reconnaissance », et d'autre part, nous avons tenté de conserver au maximum l'aspect évolutif de la perception humaine (qui nous apparaît fondamental). La succession : phase de segmentation puis phase de reconnaissance est de plus en plus remise en cause; par exemple en reconnaissance de texte, on fait souvent référence à K.M. Sayre [Sayre73] qui faisait déjà la remarque suivante : « il est paradoxal de vouloir segmenter le mot en lettres avant de l'avoir reconnu ».

Dans le cas des schémas cinématiques, nous avons, avec l'aide d'experts mécaniciens, remarqué qu'il était possible de déduire, du dessin lui-même, une partie des connaissances nécessaires à la constitution d'un contexte de départ. Remarque qui est d'ailleurs valable dans la plupart des types de documents. C'est la raison pour laquelle, nous avons choisi d'inclure dans notre système une étape préliminaire d'extraction de caractéristiques globales qui nous permet d'acquérir les éléments indispensables à la réalisation d'une analyse locale plus adaptée et donc plus précise. Pour cela, nous avons dû, d'abord, définir une représentation initiale performante, capable de décrire et de gérer toutes les formes présentes et toutes les connaissances acquises dans le document initial.

Nous avons choisi, ensuite, de tirer parti des stratégies à base d'agents permettant, notamment, une meilleure distribution des connaissances pour réaliser notre système. Plus précisément, après l'intégration des systèmes experts dans le domaine de l'interprétation automatique, nous pensons que l'utilisation de systèmes multi-spécialistes peut permettre de nombreuses avancées.

L'utilisation conjointe d'une représentation évolutive du document pour le partage de l'information et de plusieurs spécialistes réalisant chacun une tâche précise, permet, au cours de la reconnaissance et de la reconstruction des objets, d'utiliser de façon duale une vision globale et une vision locale du document (figure 2).

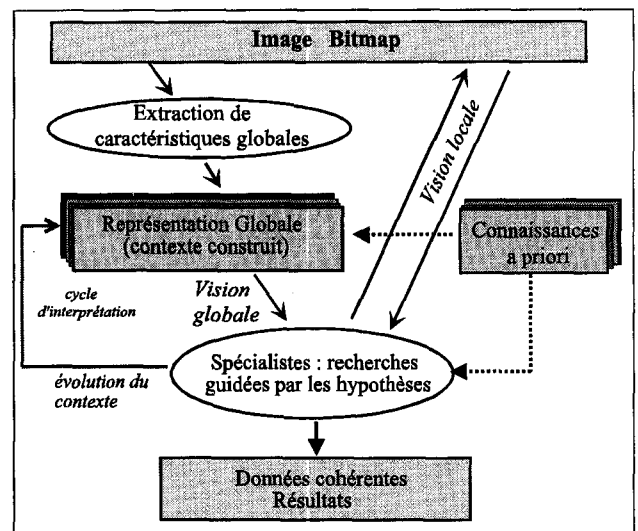


Figure 2. – Architecture du système proposé.

Une description détaillée des deux phases : extraction des caractéristiques globales et interprétation du document que comporte cette approche, est effectuée dans les deux parties suivantes.

3. l'approche globale du document

Comme nous l'avons déjà rappelé, notre système visuel permet d'avoir d'abord une appréhension globale de l'environnement (vision globale pré-attentive). Durant cette phase, notre système visuel ne tient pas compte des détails mais semble simplement essayer de situer le problème nouveau par rapport aux situations voisines déjà vécues; une extraction rapide de données structurantes et sûres dans la scène suffit généralement. Ces constatations avaient d'ailleurs conduit D. Marr à parler d'élaboration d'une esquisse primaire [Marr82], par extraction d'indices visuels, permettant d'obtenir une vue d'ensemble; cette vision globale étant, par la suite, utilisée comme contexte.

Après une description sommaire des documents concernés par l'étude, nous allons voir comment il est possible de mettre en place une approche similaire dans le cas de l'analyse des documents.

3.1. les documents techniques concernés par l'étude

Parmi les documents, les dessins techniques forment une catégorie particulière; ils sont constitués de lignes, de régions pleines, de régions hachurées, de texte, ... Ils contiennent toujours une quantité importante d'informations, peuvent être d'une grande complexité, et leur domaine d'application est très étendu. On s'accorde à distinguer parmi eux quatre grands types qui font l'objet de travaux de recherche sensiblement comparables [Tombre92] :

- les projections orthogonales
- les schémas de principe
- les plans
- les organigrammes et autres dessins

Les projections orthogonales sont des vues planes d'objets réels représentés depuis différents points d'observation. L'image comprend la plupart du temps un ensemble de vues 2D, avec des symboles et une nomenclature (liste complète des pièces ou parties de l'objet représenté). Une partie des lignes représente le mécanisme, l'autre partie fait référence à la cotation. La cotation fournit les informations complémentaires, indispensables à la compréhension totale du dessin. Toutes ces conventions ou règles sont définies dans différentes normes : DIN, BSN, NDS, ... et témoignent de la rigueur avec laquelle ces dessins sont exécutés.

Le but de l'analyse est, dans ce cas, la conversion du dessin (souvent celui d'une pièce mécanique) en une description utilisable par un système de CAO. Les informations à extraire de l'image sont donc les entités qui figurent dans la base de données de tels

systèmes ou bien des informations d'un niveau supérieur (groupement d'objets élémentaires) représentant une structure élaborée (boîte de vitesse, ...), entité que l'on retrouve, cette fois, dans les bibliothèques graphiques du système.

Les plans sont des représentations schématiques de vues d'ensemble, plus ou moins détaillées (vues aériennes ou satellitaires, plans cadastraux, ...) et comportant différents types de renseignements (selon l'application visée). De tels documents sont généralement constitués d'une superposition d'informations de caractère géographique, géologique, météorologique, ... Un symbolisme ou des codes de couleurs, d'épaisseurs des traits, de hachures, ... sont utilisés pour différencier chaque type de données (réseaux routiers, rivières, bâtiments, canalisations, courbes de niveau, ...).

Généralement, le but recherché est l'extraction d'une partie des informations contenues dans ces images (par exemple la détection des immeubles), c'est-à-dire la détection de certaines zones ou formes (souvent construites à partir de segments de droites ou de courbes), noyées parmi toutes les autres données que l'on peut alors considérer à ce stade comme du bruit.

La catégorie **des organigrammes et autres dessins** englobe les schémas représentant une organisation (description d'algorithmes, de fonctionnement, ...) et les documents de type tableau. Ils sont caractérisés par la présence de nombreuses zones de texte. Les formes à reconnaître sont généralement des polygones, lignes, cercles, et autres objets géométriques élémentaires.

Les schémas de principe nous intéressent plus particulièrement. Ils comprennent les plans de câblages électriques, électroniques, ou téléphoniques, ... ou encore des schémas cinématiques. L'information véhiculée évoque assez rarement l'aspect externe d'un objet réel; il s'agit plutôt de la description du principe de fonctionnement d'un dispositif à partir de formes logiques (les symboles), de liens, de connexions entre les symboles, et de texte.

C'est à l'aide de dessins de cette dernière classe que nous décrirons, en la justifiant, l'approche que nous proposons; il s'agit de schémas de fonctionnement de processus mécaniques ou de schémas cinématiques. Deux exemples sont présentés figure 3.

Même si ces derniers répondent complètement à la définition donnée ci-dessus, il est à noter qu'ils reflètent, en outre, la structure externe de l'objet au moyen d'une schématisation du mécanisme parfois dans l'espace en 3 dimensions. Les formes à extraire des schémas de principe sont, généralement :

- un ensemble de symboles « normalisés » (dépendant de l'application et constitués d'une variété de segments, de boucles, ou de régions pleines, ...),
- des liens de connexion entre ces symboles,
- du texte.

Le nombre des symboles est souvent très important et leurs formes comportent de nombreuses similitudes. Dans le cas des schémas cinématiques, l'analyse porte à la fois sur la position relative des objets, sur les liens physiques qui existent entre eux et sur leur reconnaissance.

De même que pour les dessins de projections orthogonales, le but de l'analyse est la conversion du dessin afin de le rendre utilisable par un système de CAO qui permettra l'archivage, les modifications, ou même les simulations de fonctionnement du

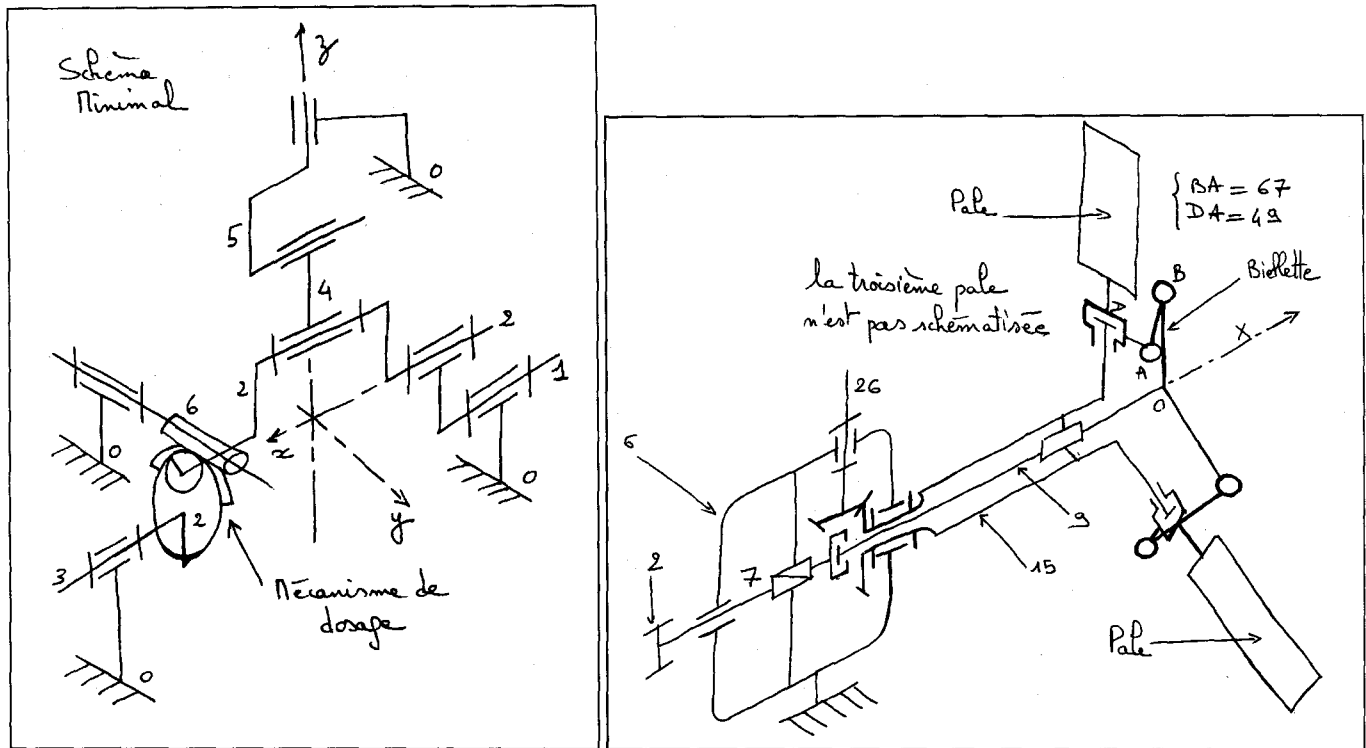


Figure 3. - Images initiales (schémas cinématiques).

processus. Ce dernier type d'application est caractérisé par la présence, dans ces schémas, d'informations sémantiques. Ces informations sont éventuellement utilisables durant les différentes phases du traitement.

Il est à noter que tous ces documents techniques peuvent se présenter sous différents aspects. Ils peuvent avoir été dessinés à main levée ou, au contraire, à l'aide d'instruments graphiques (règles, normograpes, ...) et peuvent avoir ou non subi des dégradations (photocopies successives, ...). La qualité d'une image se définit alors sur des critères d'orientation, de bruits (taches noires parasites ou blancs induisant des discontinuités), et de distorsion des formes. Face à ces défauts, dus aux techniques d'acquisition, mais aussi à la mauvaise qualité du document original lui-même, les méthodes choisies devront être plus ou moins robustes et il conviendra donc de bien définir le type et la nature des documents traités par chaque système.

Dans notre cas, le point de départ de l'interprétation est l'image binaire du document numérisé à 300 points par pouce (d'autres résolutions peuvent néanmoins être utilisées). Il s'agit pour la plupart des documents traités, de photocopies de dessins réalisés, soit à la main, soit à l'aide d'outils (règles, systèmes CAO), sur une feuille

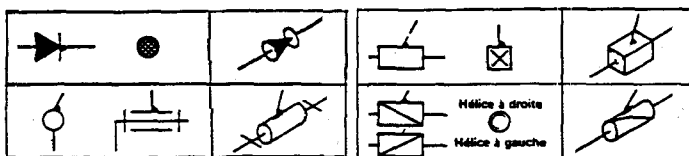


Figure 4. - Exemples de symboles « normalisés ».

de papier blanc de taille A4 par des concepteurs en mécanique. Les tests réalisés pour l'instant portent sur une base de 30 images binaires dont 20 sont des schémas cinématiques mécaniques. Les résultats obtenus ont été jugés très corrects, malheureusement aucune méthode ne permet de quantifier numériquement les résultats obtenus. La possibilité d'utiliser ces résultats, stockés dans un fichier, par un logiciel d'étude des systèmes, a été testée sur les schémas cinématiques. Les résultats présentés tout au long de cet article permettront à chacun de juger de la valeur de la méthode.

3.2. l'analyse d'ensemble des documents

Un examen des différents documents, la détermination des formes contenues, de leur type, de leur fréquence d'apparition et de leur disposition suffit pour en réaliser une première caractérisation. Il est alors possible de choisir quels primitives ou paramètres « informatiques » pourront traduire ces caractéristiques afin d'être plus facilement utilisés durant toute la phase d'interprétation.

L'analyse des formes susceptibles de se trouver dans ces documents montre que les objets à décrire sont principalement constitués par des traits et par des formes pleines. Les traits ont globalement une forme allongée, à côtés parallèles; les formes pleines ont des aspects plus variés. Les dessins comportent presque toujours de petits symboles et des zones de texte qui constituent les annotations. Ces éléments se reconnaissent souvent à leur allure particulière. Les dessins techniques, et plus particulièrement les schémas cinématiques, possèdent en outre, pour les experts

mécaniciens, deux caractéristiques intéressantes qu'il est nécessaire d'extraire en priorité; ce sont, d'une part, les directions privilégiées ou du moins la valeur moyenne des directions de traits (calcul de la rose des directions) dans l'image et d'autre part, les épaisseurs des traits.

3.3. le choix des primitives

Les primitives de description choisies devront être à même de décrire le document durant toute la suite de l'analyse. Le choix de ces primitives de description, outre leur adéquation aux formes à décrire, se fonde sur leurs aptitudes à représenter les connaissances acquises au fur et à mesure de l'analyse.

3.3.1. les vecteurs

Une forme binaire est décrite de manière équivalente par son contour ou par les pixels de la région noire, qui la constituent. Une information sur sa direction peut être obtenue par une étude locale de ses contours. Quant à l'épaisseur, qui correspond à la distance entre les deux frontières du trait, elle est surtout intéressante pour l'étude des traits. Ces raisons nous ont conduits à choisir la description des formes par les contours au moyen de la primitive Vecteur (figure 5). Ainsi, une approximation polygonale des contours des formes constituant l'image initiale fournira une suite ordonnée (par le suivi des contours) de vecteurs. La suite est notée (SV).

Nous avons choisi de reprendre la méthode d'approximation polygonale proposée par Wall et Danielsson [Wall84]. Comme le montre la figure 6, elle se base sur un calcul d'aire où le seuil concerne l'erreur cumulée autorisée (en gris sur la figure)

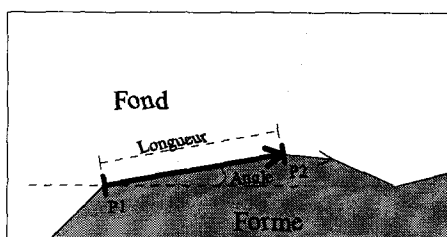
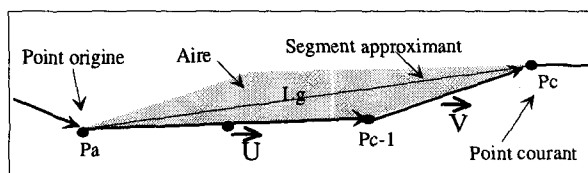


Figure 5. - Objet Vecteur.



$$Erreur = Erreur + \|\vec{U} \wedge \vec{V}\|$$

$$Erreur = Erreur + |(x_{c-1} - x_a)(y_c - y_{c-1}) - (y_{c-1} - y_a)(x_c - x_{c-1})|$$

Condition d'arrêt : $Erreur > Lg \times Epsilon$

soit : $Erreur > Epsilon \times \sqrt{(x_c - x_a)^2 + (y_c - y_a)^2}$

Figure 6. - Critère de l'approximation polygonale.

par rapport à la longueur du segment approximant ($Aire / Lg < Epsilon$). On détermine ainsi les sommets de la ligne polygonale : les points de contrôle.

Les segments obtenus sont stockés dans une liste chaînée en respectant l'ordre fourni par le suivi des contours. Ensuite, pour améliorer les performances de sélection des points de contrôle et diminuer la variabilité suivant le seuil choisi, on étudie les frontières à différents niveaux de détails. Pour cela, la phase d'approximation polygonale est opérée de manière itérative (en utilisant toujours le même algorithme d'approximation) sur les points de contrôle nouvellement obtenus.

Le nombre de points de contrôle est ainsi réduit par fusion de certains des segments obtenus lors des étapes successives. Ce procédé est répété jusqu'à stabilisation, c'est-à-dire jusqu'à ce que plus aucune fusion ne soit possible. Le nombre d'itérations avant stabilisation varie généralement entre 2 et 5 suivant les caractéristiques du contour (courbure, manuscrit ou imprimé, ...).

La représentation de l'image sous forme de chaînes de Vecteurs correspondant aux contours des formes permet de limiter la perte d'information et donne une approximation très fidèle du dessin (figure 7), mais elle n'est pas facilement exploitable et ne fournit pas suffisamment d'information sur la structure du document. On obtiendrait alors seulement une vectorisation des contours. La structuration supplémentaire décrite ci-dessous permet de faire évoluer cette représentation de manière hiérarchique et d'acquérir des renseignements intéressants.

3.3.2. les Quadrilatères

Pour obtenir et gérer aisément des données supplémentaires, il a fallu définir un outil de description plus évolué. La nature des formes à extraire nous a incités à choisir le *Quadrilatère*.

Une étape d'appariement des *Vecteurs* de la suite SV permet la mise en place des primitives *Quadrilatères*. Chaque *Quadrilatère* est défini par une paire de Vecteurs (chacun appartenant à l'une des deux frontières opposées d'une forme fine). Cette description peut être comparée à celle utilisée par Shih et Kasturi basée sur des trapèzes [Shih89]. Le Quadrilatère étant une primitive plus souple, il n'est pas nécessaire, dans notre cas, d'effectuer, au préalable, l'extraction du texte contenu dans l'image. Afin d'obtenir une description plus robuste des formes fines, la construction des Quadrilatères s'effectue en plusieurs étapes :

- Appariement des *Vecteurs* (éléments de SV) pour construire les *Quadrilatères* et mise à jour de SV (ensemble des vecteurs non appariés).

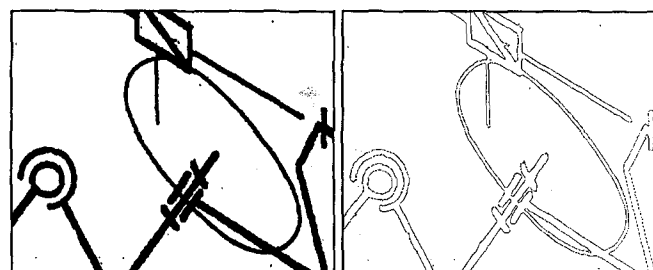


Figure 7. - Approximation des contours par chaînes de Vecteurs.

Interprétation de documents techniques par «cycles perceptifs»

- Tri des *Quadrilatères* suivant leur proximité de manière à reconstruire une « chronologie » du tracé
- Fusion de certains *Quadrilatères* voisins

Pour augmenter la robustesse de l'algorithme d'appariement, on commence par sélectionner le vecteur V1 de longueur maximum dans la liste des vecteurs encore non appariés, puis on recherche le vecteur le plus proche (suivant la distance euclidienne entre points) de chacune des extrémités de V1 et qui vérifie certains critères. Pour que la mise en correspondance de 2 vecteurs ait lieu, des conditions (critères d'appariement), correspondant chacune à une propriété physique des traits, doivent être remplies. Ces critères ont été choisis « empiriquement » mais en respectant la définition intuitive d'un trait :

- un vecteur ne peut être apparié qu'avec le vecteur le plus proche de lui (la distance entre les extrémités des 2 vecteurs doit être inférieure à 60 pixels)
- la différence entre les directions (pentes) des 2 vecteurs doit être inférieure à $\pi/8$
- présence de pixels noirs entre les extrémités (P1Q2 et P2'Q1 dans la figure 8)
- les vecteurs doivent être de sens opposé
- la longueur du 1^{er} vecteur doit être supérieure aux épaisseurs (e1 et e2 calculées à partir des extrémités)

Pour améliorer l'appariement, certains vecteurs peuvent être décomposés en 2 vecteurs fils, le point de coupure est obtenu par projection d'une extrémité d'un vecteur sur le support de l'autre (figure 8). Les longueurs des deux vecteurs constituant le *Quadrilatère* sont, de cette manière, plus semblables. Le second vecteur fils reste dans la liste des vecteurs à étudier. Le processus d'appariement s'arrête lorsque plus aucune mise en correspondance n'est réalisable (selon les critères d'appariement imposés).

Du fait de la localisation des points de contrôle (provenant de l'approximation polygonale) et de la méthode d'appariement, une phase de fusion des *Quadrilatères* est nécessaire pour obtenir des objets se rapprochant au mieux des traits effectifs sur l'image (*Quadrilatères* de dimensions optimales et en nombre réduit).

Cette fusion sera effectuée après une étape de tri des *Quadrilatères* selon leur proximité. Pour cela, on définit une distance δ (figure 9) :

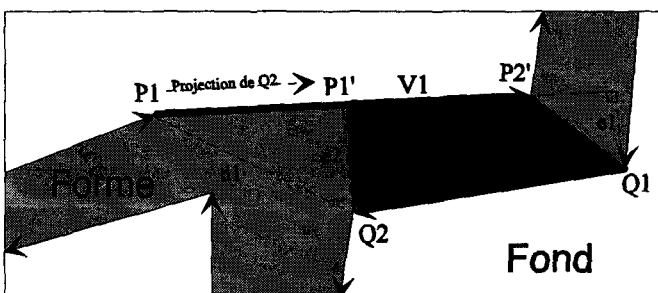


Figure 8. – Appariement des Vecteurs.

Q et Q' désignant des quadrilatères (voir notations sur le schéma) :

Si $|\text{angle } Q - \text{angle } Q'| < \pi/8$ ET Noir. entre (Q, Q') alors

$$\delta(Q, Q') = \text{Min}(P_1P'_2 + Q_2Q'_1, P_1Q'_2 + Q_2P'_1, P_2P'_1 + Q_1Q'_2, P'_2Q_1 + Q'_1P_2)$$

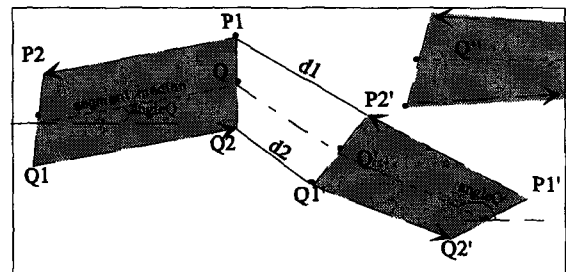
Sinon

$$\delta(Q, Q') = +\infty$$

L'amélioration apportée est visible sur la figure 10.

Pour rendre le résultat de la fusion le plus stable possible, les *Quadrilatères* obtenus sont rangés de manière à reproduire au mieux un ordre chronologique plausible du tracé. Le principal problème qui se pose est le choix du point de départ. Une fois ce choix réalisé, une recherche de son plus proche voisin permet de définir de proche en proche le nouveau *Quadrilatère* courant et ceci jusqu'à la rencontre d'un *Quadrilatère* sans voisin (suivant marquant la fin d'un tracé.

Pour le choix du *Quadrilatère* à l'origine d'un tracé, nous sélectionnons un *Quadrilatère* ayant un voisin le plus éloigné de lui.



Ici, $\delta(Q, Q') < \delta(Q, Q'')$

Figure 9. – Distance entre Quadrilatères.

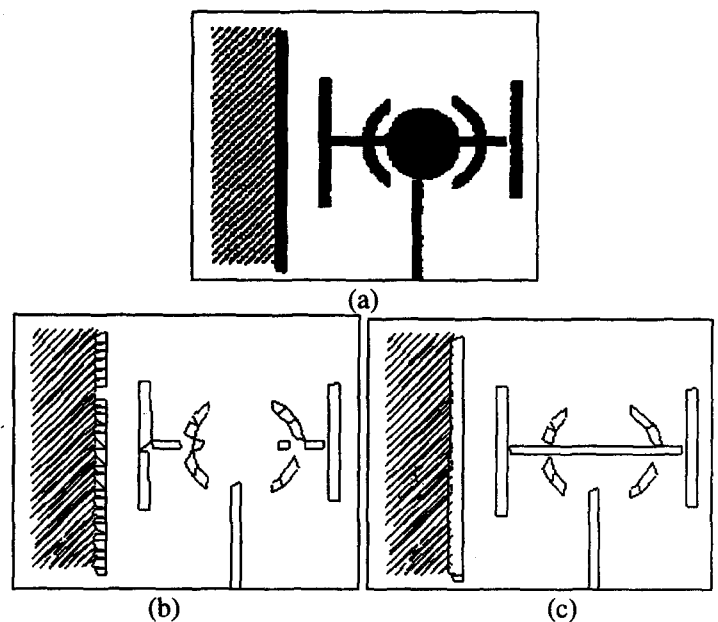


Figure 10. – Utilité de la phase de fusion des *Quadrilatères* (a) Image initiale, (b) avant fusion, (c) après fusion.

La probabilité que ce Quadrilatère corresponde à l'extrémité d'un tracé est ainsi accrue (figure 11). Ici la distance utilisée repose sur la distance euclidienne entre les extrémités des segments médians des Quadrilatères.

Ce tri permet de revenir à la manière dont a été dessinée l'image de départ. Une chronologie plausible du tracé du dessin est ainsi reconstituée et stockée dans la suite SQ (stockage des *Quadrilatères* construits). En se basant toujours sur le principe d'analyse de l'image à différents niveaux de détails, la phase de fusion des *Quadrilatères* est réalisée suivant le même algorithme d'approximation polygonale que celui utilisé lors de la construction des vecteurs, cette fois-ci, appliqué aux extrémités des segments médians des *Quadrilatères* telles qu'elles ont été précédemment ordonnées (figure 12). L'algorithme tient cependant compte des attributs de chaque *Quadrilatère* (notamment de l'épaisseur) pour autoriser ou refuser la fusion.

Les vecteurs ne pouvant être mis en correspondance avec aucun autre, correspondent aux contours des formes pleines. Les résultats obtenus lors de l'étape d'appariement des *Quadrilatères* sur les schémas donnés figure 3 sont présentés sur la figure 13 (suite SQ).

L'ensemble des objets *Quadrilatères* finalement obtenu est représentatif des formes fines présentes dans le document initial; il ne faut cependant pas voir cette étape comme une véritable étape de vectorisation du document mais plutôt comme la méthode que nous avons choisie pour obtenir une description initiale des formes fines d'un document à l'aide d'une primitive très simple à manipuler.

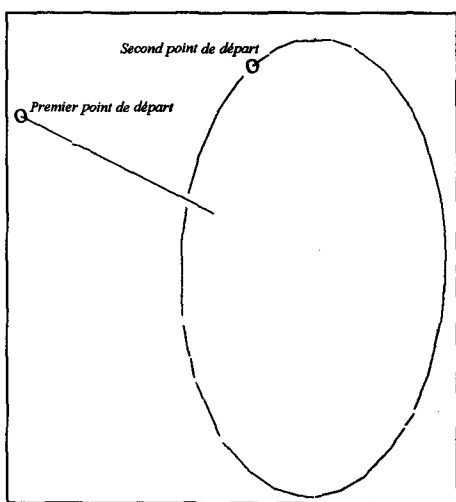


Figure 11. - Choix d'un point de départ, aisé pour une droite, indifférent pour un cercle.

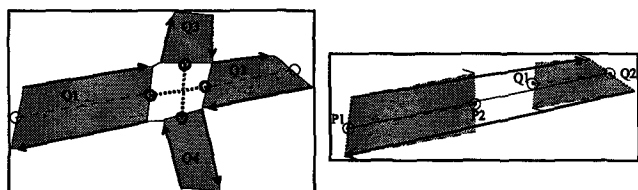


Figure 12. - Fusion des Quadrilatères.

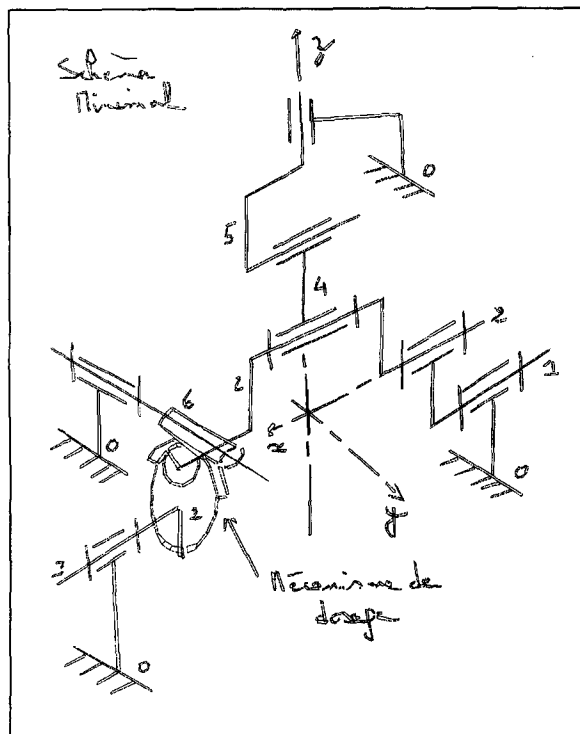
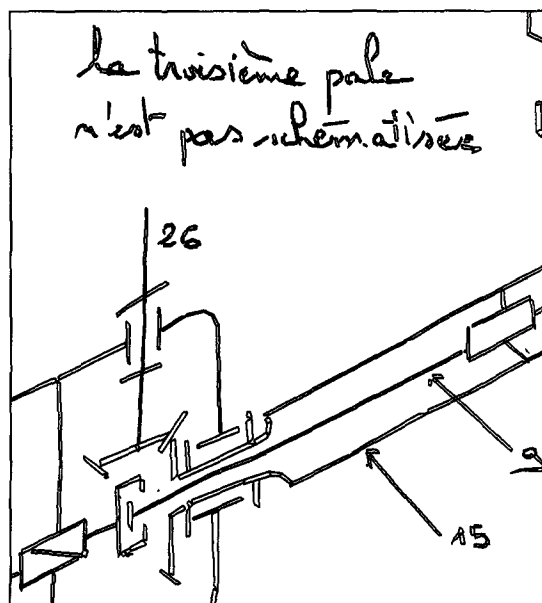


Figure 13. - Image des Quadrilatères.

Le premier inconvénient de cette représentation est qu'elle ne traduit pas directement les intersections entre traits du document original, comme le montre la figure 14a. Ensuite, la phase de fusion des *Quadrilatères* (sensible aux changements de direction) engendre parfois l'apparition de bruits au niveau de certaines intersections (figure 14b). Enfin, la gestion de l'épaisseur des traits peut provoquer le découpage d'un trait en plusieurs *Quadrilatères* (figure 14c). Par contre, la plupart du temps, il y a bien reconstitution double au niveau d'une intersection de traits, ce qui correspond bien à la façon dont le dessin a été obtenu.

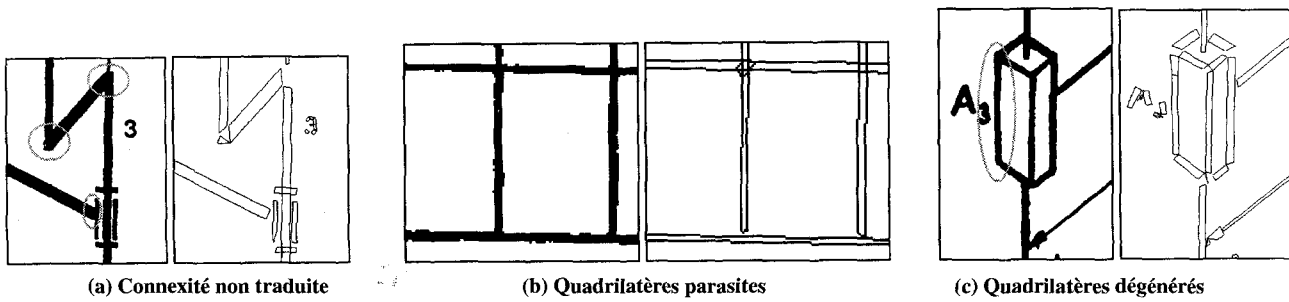


Figure 14. – Quelques difficultés

3.4. structure hiérarchique de la représentation

Pour mettre en place cette représentation initiale du document, nous utilisons une organisation hiérarchique des primitives de description (figure 15). A chaque primitive de représentation correspond une classe d'objets; cela facilitera la gestion des connaissances, la vérification de la cohérence des résultats et la coexistence des formes et permettra, à tout niveau de la hiérarchie, d'utiliser toutes les informations exprimées dans les niveaux inférieurs. Au bas de la hiérarchie se trouvent les objets *Pixels* qui constituent l'information de base.

Pour gérer les connaissances qui lui sont associées, chaque classe d'objets possède des attributs. Tout au long du traitement qui fait évoluer la représentation, nous conservons intact un exemplaire de l'image initiale; elle permet, à tout moment, de procéder à des vérifications avant de prendre une décision.

3.5. apport de la représentation

L'utilisation de cette représentation semble avoir permis de résoudre certains problèmes concernant les traitements de bas niveau notamment :

- Les formes pleines et les jonctions ou extrémités sont moins altérées que lors d'une véritable squelettisation (préalable fréquent à la vectorisation). La figure 16 présente la représentation obtenue (Vecteurs et Quadrilatères) pour une partie d'image contenant des formes pleines.
- La représentation des traits obtenue paraît de meilleure qualité qu'avec certains algorithmes d'approximation polygonale

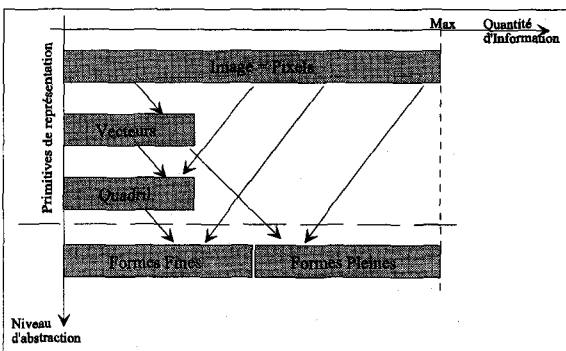


Figure 15. – Hiérarchie des classes d'objets.

[Kadonaga95] fournissant des points de contrôle pas toujours représentatifs de la forme initiale (barbules, jonctions, ...) comme on peut le voir sur la figure 17. Cette amélioration provient essentiellement de l'étape de fusion des quadrilatères (approche multi-échelles).

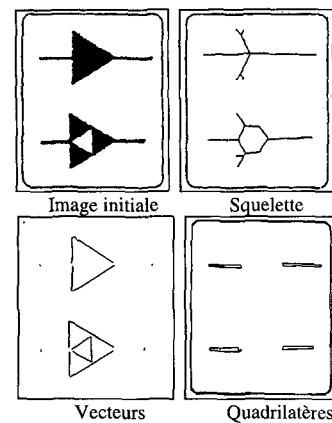


Figure 16. – Réponse aux problèmes de la squelettisation.

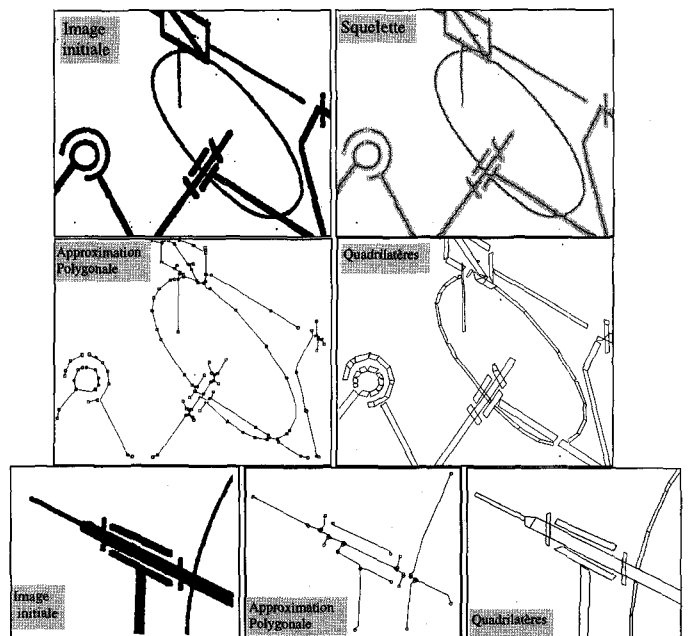


Figure 17. – Réponse aux problèmes de l'approximation polygonale.

De plus, notre représentation permet de distinguer les formes pleines (limitées par des Vecteurs), des formes fines (Quadrilatères), les directions et épaisseurs des formes fines, et leur disposition dans l'image. Sa construction constitue la phase d'approche globale du document, les primitives utilisées seront donc une source très importante d'informations utilisées tout au long de l'analyse.

Cette représentation initiale du document prend appui sur l'image initiale, son mode de construction se fonde sur une analyse préalable de la structure des documents à interpréter (connaissances a priori sur le type du document : dessin technique, partition musicale, ...). Elle peut donc être considérée comme une **représentation** (structure de données) pouvant s'adapter pour décrire différents types de documents comme le montrent les figures 18, 19 et 20.

Une fois les primitives de base extraites, les systèmes classiques d'interprétation de documents complexes se heurtent à d'autres difficultés. Il devient très difficile, notamment après la séparation des différentes couches que comporte un schéma, d'effectuer la reconstruction et la reconnaissance d'entités de haut niveau

puisque leurs composants peuvent être dispersés dans plusieurs des couches. Les représentations engendrées sont très différentes les unes des autres, la comparaison et la mise en correspondance

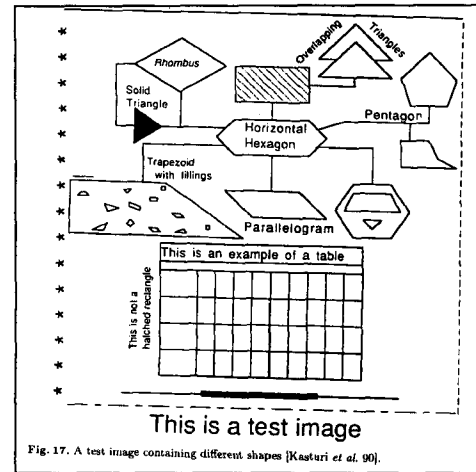


Fig. 17. A test image containing different shapes [Kasturi et al. 90].

(a) Image initiale

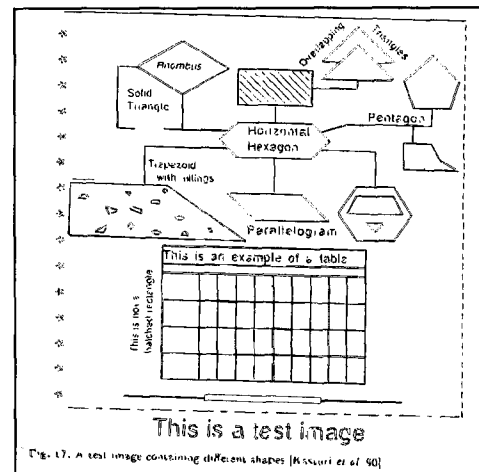
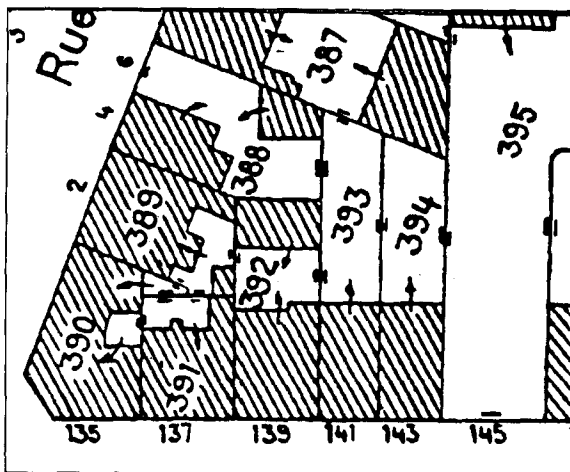
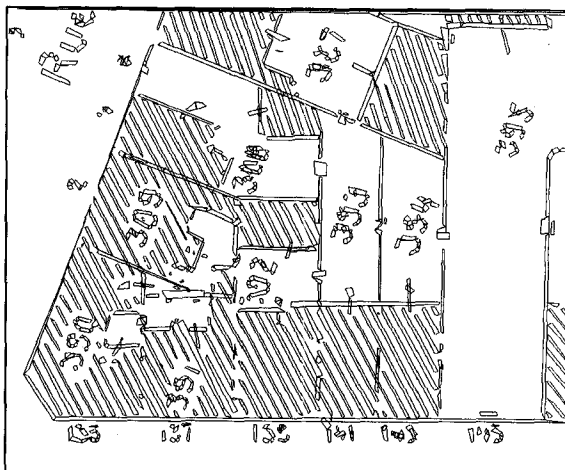


Fig. 17. A test image containing different shapes [Kasturi et al. 90].

(b) Ensemble des formes fines

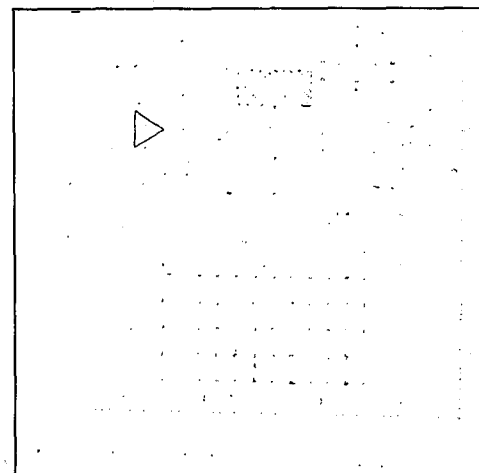


(a) Image initiale (extrait de [Antoine92])



(b) Ensemble des formes fines

Figure 18. – Représentation de planches cadastrales (ensemble SQ).



(c) Vecteurs non appariés

Figure 19. – Représentation d'organigrammes (ensemble SQ et SV).



(a) Image initiale



(b) Ensemble des formes fines

Figure 20. – Représentation des partitions musicales (ensemble SQ).

des résultats obtenus pour chaque couche deviennent quasiment impossibles. La plupart des réalisations ne fournissent pas de solution satisfaisante; les recherches s'interrompent donc à ce

niveau en raison d'une interprétation isolée de chacune des couches. De plus, il semble que les retours en arrière soient pratiquement irréalisables sans remise en cause de la totalité de l'analyse.

Dans notre cas, nous allons montrer, dans l'étape suivante, comment il est possible de faire évoluer progressivement la représentation initiale obtenue ici jusqu'à ce qu'elle traduise la compréhension du contenu sémantique du document. On peut remarquer que l'analyse de bas niveau utilise peu de connaissances a priori afin de pouvoir être utilisée sur différents types de documents techniques. Cependant, il ne s'agit que de la phase de mise en place d'une représentation initiale du document qui évoluera par la suite grâce aux connaissances acquises par les spécialistes.

4. interprétation du document par «cycles perceptifs»

L'interprétation du document est réalisée par différents **spécialistes**. Tous utilisent la vision globale pour focaliser leur attention sur des zones précises de l'image (génération d'hypothèses) qu'ils étudient en détail (vérification) pour faire avancer l'analyse. Pour effectuer ce travail d'exploration - confrontation, les spécialistes possèdent des compétences élémentaires, précises et suffisantes dans des domaines particuliers comme l'extraction du texte, des courbes, des éléments mécaniques, ... Le mécanisme, génération d'hypothèses - validation, souvent employé en reconnaissance de formes, peut ainsi être reproduit. Ici, la vision globale est à la base des hypothèses générées, tandis que les analyses locales permettent la validation ou la réfutation de ces hypothèses.

4.1. les différents spécialistes

Du fait des types de tâches à effectuer et des connaissances utiles pour réaliser correctement chacune d'elles, nous avons, dans le cas des dessins cinématiques, eu recours aux six spécialistes cités ci-dessous. Une description détaillée du fonctionnement de ces différents spécialistes peut être trouvée dans [Ramel96]. Nous allons cependant en donner un aperçu afin de pouvoir, par la suite, illustrer le fonctionnement des cycles perceptifs.

- Le premier spécialiste est chargé d'extraire les zones de **texte**. La séparation des caractères du graphique pose des difficultés lorsque la taille des composantes connexes n'est plus significative [Kasturi92], c'est-à-dire quand un caractère est connecté à une autre partie du dessin ou quand il s'agit de manuscrit (figure 21).

Dans notre représentation, la taille, la forme et la disposition des quadrilatères sont représentatives de la texture de l'image initiale (figure 22). Les zones de texte sont caractérisées par des regroupements de quadrilatères courts. Ils sont détectés par leur proximité physique et sans tenir compte de leur orientation. Bien que d'autres formes (arcs, lignes pointillées, ...) produisent des

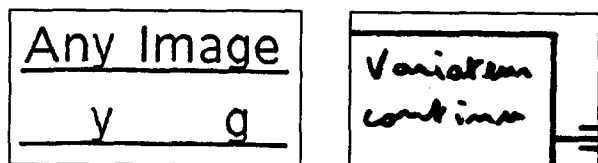


Figure 21. – Problèmes de localisation du texte.

quadrilatères ayant des caractéristiques semblables, une analyse locale de ces groupes de quadrilatères et de leur voisinage permet la construction de zones de focalisation (susceptibles de contenir du texte manuscrit ou imprimé), ceci sans avoir recours aux composantes connexes. A partir de chaque amas de quadrilatères, il est possible de générer une **composante de texte virtuelle** (zone de focalisation) constituée par le rectangle englobant l'ensemble des quadrilatères regroupés pour former l'amas; la figure 22 montre quelques exemples de zones de focalisation.

En couplant cette information avec celle fournie par l'étude des composantes connexes présentes dans le document initial, le spécialiste prend une décision et classe chaque zone dans l'une des catégories suivantes : texte, non texte, indécision.

Dans les zones d'indécision, en étudiant la taille et la position des composantes de texte virtuelles par rapport aux zones de texte déjà localisées, on procède à une confrontation avec le contexte, et certaines ambiguïtés peuvent alors être levées (figure 23). Enfin dans les cas où le doute subsiste, on peut faire appel à un module

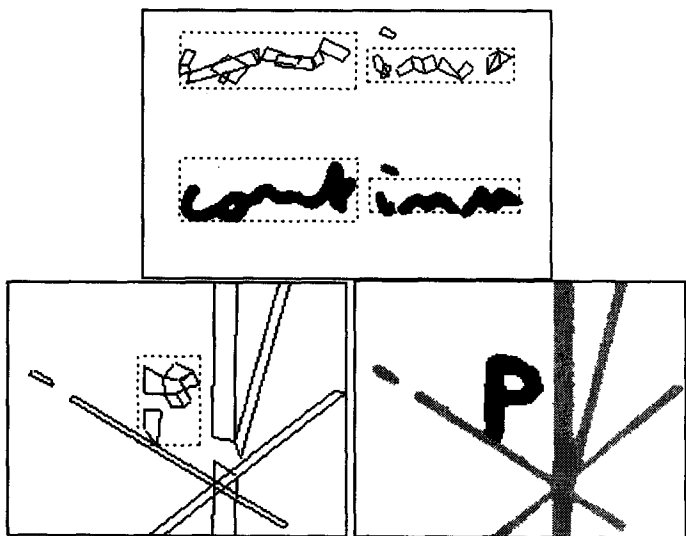


Figure 22. – Composantes de texte virtuelles.

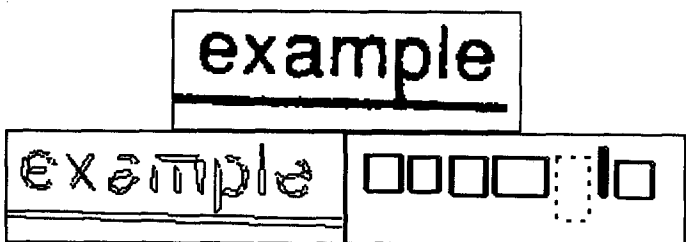


Figure 23. – Localisation de texte grâce au contexte.

d'OCR ou, dans un système interactif, interroger l'utilisateur pour prendre d'autres décisions.

- Un second spécialiste recherche les **formes pleines**. Dès que les quadrilatères ont été construits, la suite (SV) ne contient plus que les vecteurs pour lesquels la mise en correspondance a échoué; la plupart de ces vecteurs (de longueurs significatives) appartient alors au contour d'une forme pleine. Ainsi, les formes pleines peuvent être localisées par recherche de chaînes de contour quasi-fermées dans la suite (SV) et par vérification dans l'image initiale (figure 24). La position de chaque forme est sauvegardée.

- Le spécialiste « Pointillés » est chargé de définir le **type de chaque ligne composant** le dessin. Ce type est attribué après une étude de la taille et de la position relative des quadrilatères consécutifs. Les paramètres calculés permettent de classer chaque ligne parmi trois catégories : pointillés réguliers, pointillés mixtes, ligne continue.

- Les **courbes** sont, elles aussi, extraites par un spécialiste à partir de la représentation initiale. Chaque succession de quadrilatères consécutifs et de faibles longueurs est susceptible de correspondre à une courbe; le spécialiste va tenter d'approximer cette succession à l'aide d'une courbe de Bézier; si le résultat de l'approximation est satisfaisant, la suite de quadrilatères est remplacée par sa représentation sous forme du polynôme de Bézier.

Avant d'être approximés, les ensembles de points susceptibles de constituer des courbes doivent être localisés dans l'image. Les courbes sont des formes fines; l'ensemble des primitives quadrilatères du modèle de description de l'image traduit donc leur présence. De même que pour les caractères, une courbe dans le document, de par sa nature, correspond à une succession de quadrilatères de faible longueur dans le modèle de représentation.

Pour pouvoir être approximée par une courbe de Bézier, chaque courbe est représentée par une suite de points $M_k(x_k, y_k)$ suivant l'abscisse curviligne croissante; à ces points, on tentera de faire correspondre les points $H_k(P_x(t_k), P_y(t_k))$ qui sont les points de la courbe de Bézier approximante. Les points M_k sélectionnés sont les extrémités des axes médians des quadrilatères susceptibles de constituer une courbe. Les quadrilatères sont ordonnés (selon la proximité entre leurs extrémités) de manière à reconstruire une chronologie du tracé de la courbe. Grâce à cette technique, il est possible d'associer un t_k à chaque extrémité des axes médians.

La courbe est considérée comme un signal continu observable dans le temps à différents instants correspondant aux points M_k précédents. Les temps t_k d'observation associés aux différents points sont calculés en fonction de la longueur à parcourir pour

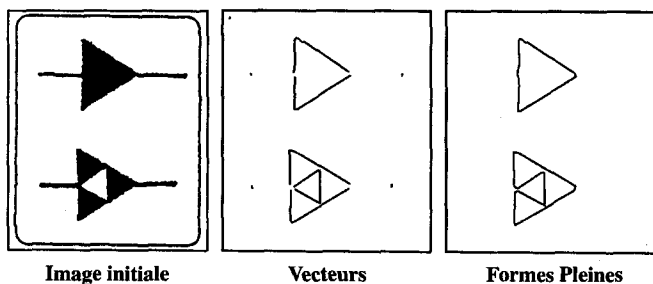


Figure 24. – Détection des contours des formes pleines.

passer d'un point à l'autre, dans un mouvement uniforme :

$$t_0 = 0 \text{ et } t_k = t_{k-1} + d(M_k, M_{k-1})/L \text{ avec } L = \sum_{j=1}^m d(M_j, M_{j-1})$$

L'erreur commise en considérant la courbe de Bézier trouvée, est estimée en évaluant la distance entre les 2 courbes, relative à la norme de la convergence uniforme. Elle est évaluée par un calcul ponctuel à différents instants t_i , et seule l'erreur maximum est prise en compte. Les points de la courbe de Bézier pour chaque t_i sont calculés et sont projetés orthogonalement sur la courbe à approximer qui est une suite de segments. La distance séparant le point et sa projection définit l'erreur à $t = t_i$ (figure 25).

Le degré et les pôles correspondant à la meilleure approximation obtenue constituent les paramètres finals de l'approximation.

Le dernier spécialiste est, en revanche, beaucoup plus lié à l'application qui nous intéresse.

• Pour reconnaître les éléments mécaniques, il nous a semblé nécessaire d'organiser les informations extraites de manière à mieux utiliser le contexte lors de l'interprétation; une représentation sous la forme d'un graphe est la plus adaptée pour lier les différents objets qui constituent le dessin en fonction de leur voisinage, c'est aussi la plus apte à traduire les relations structurelles entre primitives.

Les objets extraits (quadrilatères, symboles pleins) constituent les noeuds du graphe; les arcs reliant ces noeuds traduisent les relations qui existent entre les primitives; le graphe représente la structure globale du document. Pour le construire, on calcule pour chaque primitive une zone d'influence. Des arcs décriront la nature des liens existant entre la primitive en cours d'étude et celles appartenant à sa zone d'influence.

Nous construisons, dans un premier temps, le graphe correspondant uniquement aux formes fines (suite SQ); le tableau ci-dessous (figure 26) répertorie les différents types de relations qui servent à la construction du graphe.

La figure 27 décrit le processus de construction de la partie du graphe correspondant au Quadrilatère 0 (élément de SQ). Une fois cette partie construite, on recherche l'existence d'un sous-graphe correspondant au graphe d'une des entités mécaniques normalisées stockées dans un dictionnaire. Le graphe global est ainsi construit et les entités mécaniques non ambiguës sont reconnues et étiquetées durant cette étude.

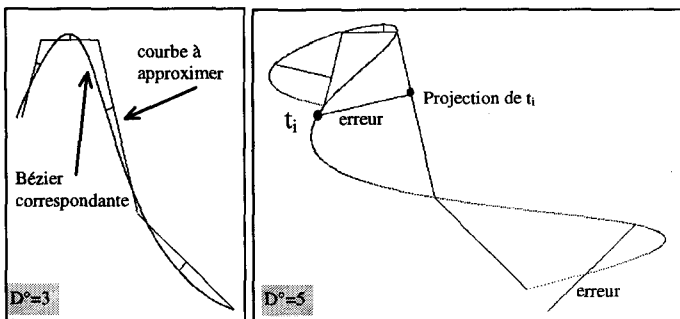


Figure 25. - Calcul de l'erreur d'approximation.

Type de liaison entre deux primitives	Exemple correspondant
T (Liaison orientée en T)	
X (Intersection)	
P (Parallèles)	
L (Liaison en L)	
S (Successifs)	

Figure 26. - Types d'interaction.

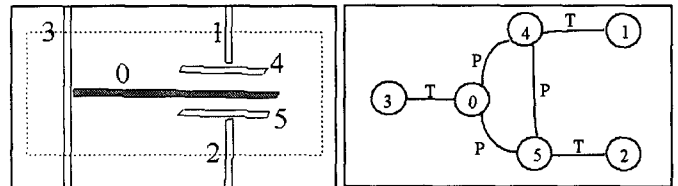


Figure 27. - Zone d'influence du Quadrilatère 0 et graphe correspondant.

• L'étude des attributs de chaque primitive, associée à un mécanisme de propagation d'étiquettes dans le graphe, permet de résoudre certaines ambiguïtés, de vérifier la cohérence des données extraites et permet aussi de positionner les différentes entités les unes par rapport aux autres. Ce travail est réalisé par le spécialiste chargé d'effectuer la « reconstruction 3D » du mécanisme.

4.2. les « cycles perceptifs »

Dans l'architecture choisie, un mécanisme de « cycles perceptifs » permet de faire coopérer les différents spécialistes. Les spécialistes sont activés de manière séquentielle et contribuent ainsi à simplifier et à enrichir la représentation de manière progressive (figure 28). La mise en place informatique de ce mécanisme

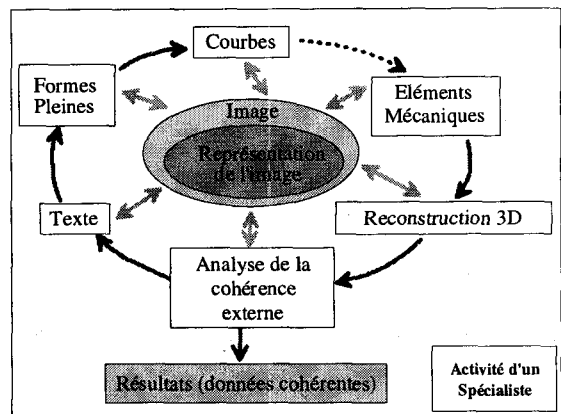


Figure 28. - Cycle perceptif (pour les schémas cinématiques).

implique la séquentialisation. Les cycles permettent de relier les différentes perceptions, de combiner les opérations pour atteindre un but commun. Des données ambiguës à un instant donné, pour un spécialiste donné, peuvent constituer des évidences pour un autre spécialiste. De sorte que, lors d'une prochaine activation d'un spécialiste la représentation aura été simplifiée et le nombre d'ambiguïtés réduit par rapport à l'état courant d'interprétation. Les problèmes, parfois complexes, de choix de la répartition des connaissances a priori et du raisonnement dans le système sont ainsi résolus par l'utilisation duale de la représentation du document et du mécanisme d'activation cyclique de spécialistes.

Un des problèmes de cette méthode réside dans le choix de l'ordre d'activation des différents spécialistes. Ce choix est basé, pour l'instant, sur les connaissances a priori tirées de l'étude de la structure des documents. Telle sera la stratégie exploratoire de capture de l'information.

En outre, si l'homme possède une habileté très forte à extraire les évidences en faisant appel aux vastes connaissances qu'il a acquises, le mécanisme de cycles perceptifs est le moyen que nous avons trouvé pour permettre aux spécialistes de coopérer pour extraire seulement des données certaines et fiables : **les évidences**, et ainsi de limiter les erreurs d'analyse. En fin de cycle, la complétude de la compréhension du document est vérifiée et si nécessaire, un nouveau cycle d'activation des spécialistes permet une nouvelle avancée dans l'interprétation. Dans notre cas, chaque spécialiste, à son niveau, utilise les connaissances dont il dispose pour renvoyer des données cohérentes. Il analyse la cohérence de la représentation de son point de vue par confrontation avec ses propres connaissances. Suivant le travail qu'il a à réaliser, il gère des entités de plus ou moins haut niveau sémantique. Un travail intéressant sur la gestion des incohérences dans le cas du cadastre français est proposé par J.M. Ogier [Ogier95] et peut être adapté à notre problème : chaque spécialiste vérifie alors la cohérence interne des objets qu'il construit tandis que la cohérence externe est vérifiée en fin de chaque cycle par un spécialiste particulier.

Le paragraphe suivant met en lumière l'utilité du mécanisme de cycles perceptifs, par la description détaillée d'un exemple.

4.3. illustrations

4.3.1. le premier cycle perceptif

La figure 13 montre la représentation de l'image, obtenue après la phase de vision globale (c'est-à-dire la mise en place de la représentation); le dessin choisi ne comportant pas de formes pleines, l'information se trouve contenue dans l'ensemble des quadrilatères.

Le premier cycle perceptif débute par l'activation du spécialiste chargé d'extraire les zones de texte; on peut remarquer, sur la figure 29, qu'à cet instant, le nombre des zones de focalisation (composantes texte virtuelles) auxquelles est confronté le spécialiste est assez important. Néanmoins les caractères isolés et les caractères non attachés sont détectés sans problème (figure 30). On peut aussi remarquer que les têtes de flèches des axes x et y sont considérées comme des caractères. Il n'en a pas été de même pour le manuscrit qui ne sera appréhendé ici que durant le second cycle perceptif.

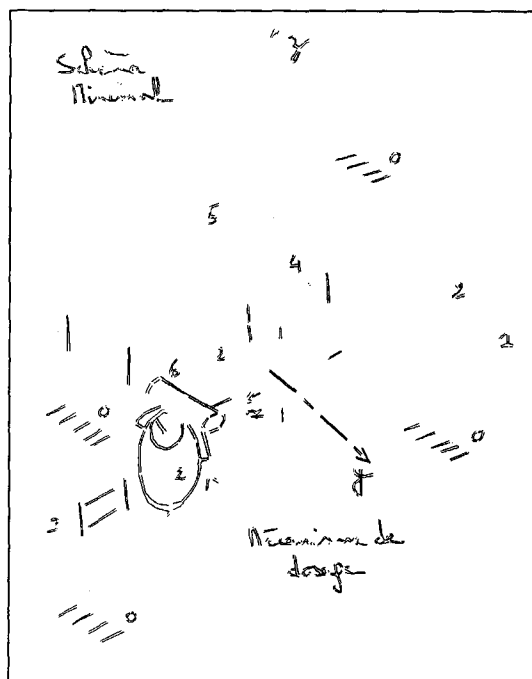


Figure 29. - Zones de focalisation (1er cycle).

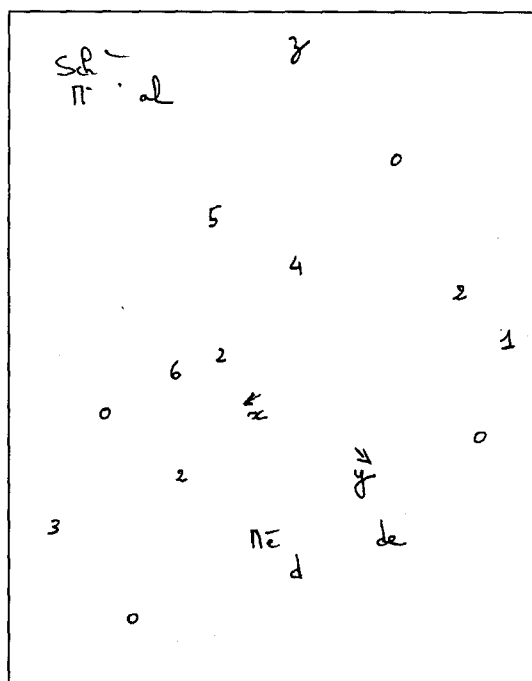


Figure 30. - Zones de texte localisées.

La figure 31 décrit les résultats obtenus par le processus chargé de localiser les courbes présentes dans l'image. Celles-ci sont bien localisées, néanmoins une succession de quadrilatères traduisant un trait non rectiligne a été prise comme courbe et approximée par une courbe de Bézier; elle est néanmoins devenue une entité unique.

Les entités mécaniques sont localisées correctement (figure 32); seule une liaison pivot fixe a été confondue avec une liaison pivot glissant (encadré par des pointillés). L'erreur est due, comme précédemment, à la mauvaise représentation d'un trait : un trait mal dessiné (non rectiligne) s'est traduit par la génération de deux quadrilatères.

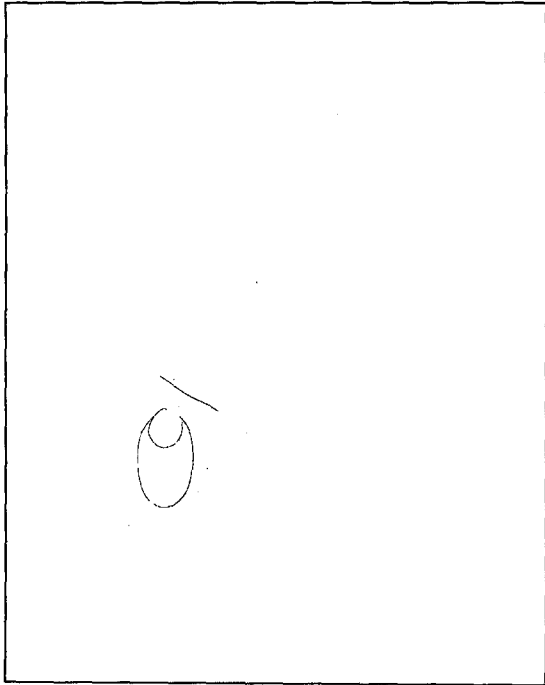


Figure 31. – Courbes localisées.

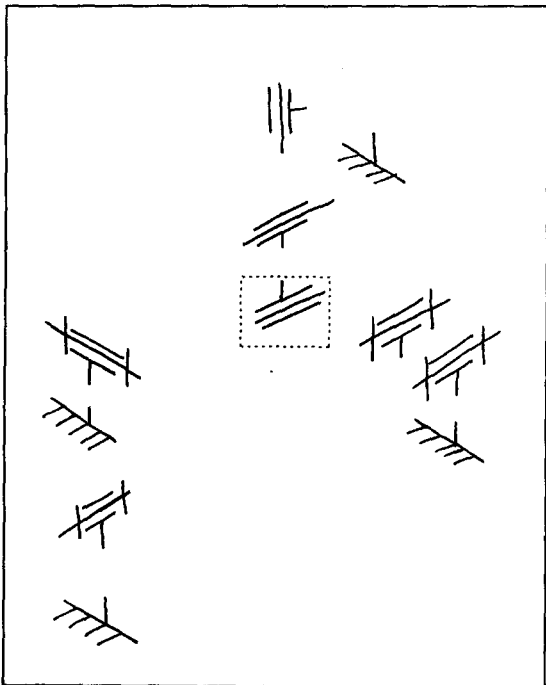


Figure 32. – Entités mécaniques localisées.

L'interprétation du schéma en trois dimensions permet de positionner les différentes liaisons les unes par rapport aux autres et de poursuivre la reconstruction des pièces constituant le mécanisme représenté par étiquetage des primitives encore non interprétées (figure 33). Ici, le positionnement des liaisons les unes par rapport aux autres n'est pas réalisé complètement car certaines liaisons n'ont pas pu être mises en correspondance avec d'autres, non pas à cause d'une mauvaise interprétation mais parce que le dessin initial est ambigu, même pour l'homme. Notre système est capable de signaler à l'utilisateur cette ambiguïté.

A première vue, le premier cycle perceptif peut sembler différent des autres puisque, à un instant donné, il n'existe encore aucun objet du type recherché par le spécialiste activé. En fait, la méthode choisie associant cycle perceptif et minimisation de la prise de risque, permet que le changement de l'ordre d'activation des processus influe peu sur les résultats finalement obtenus mais seulement sur la rapidité du processus d'interprétation. Certains objets sont plus facilement localisables que d'autres ou constituent un préalable à la localisation d'entités de plus haut niveau mais tous sont détectés de la même façon à plus ou moins longue échéance.

4.3.2. les autres cycles perceptifs

Comme le montre la figure 34, le nombre de zones de focalisation devant être analysées par le spécialiste Texte durant le second cycle a diminué depuis sa première activation (figure 29). Certaines zones problématiques (pendant le premier cycle) pour ce spécialiste ont constitué des évidences pour certains de ses confrères (courbes, éléments mécaniques, ...). La résolution (élimination)

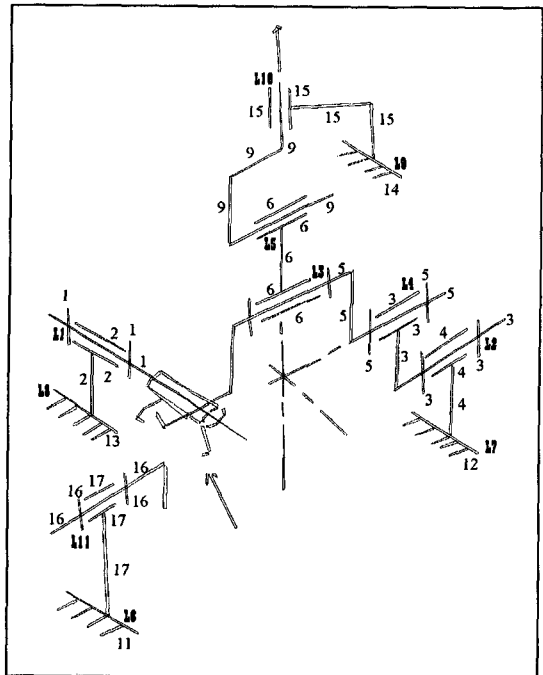


Figure 33. – Reconstruction 3D.

de ces ambiguïtés ainsi que le texte déjà localisé facilitent la localisation de nouvelles zones de texte durant les cycles suivants (figure 35).

L'interprétation est terminée lorsque un cycle perceptif entier n'apporte plus aucune modification au modèle de représentation.

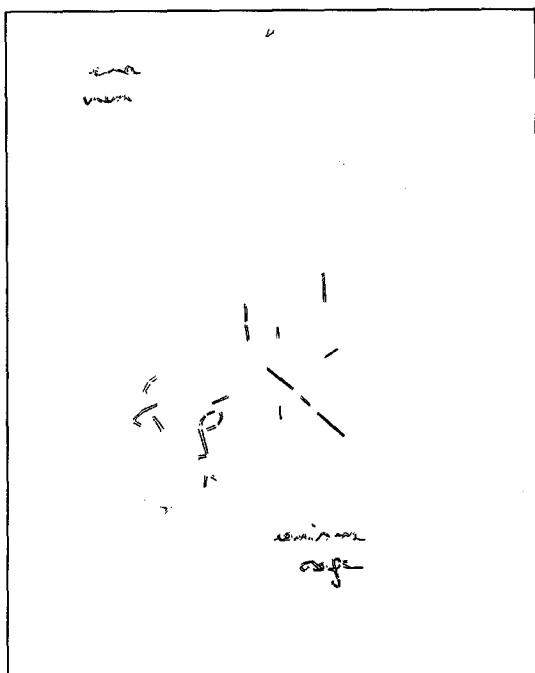


Figure 34. – Zones de focalisation (2e cycle).

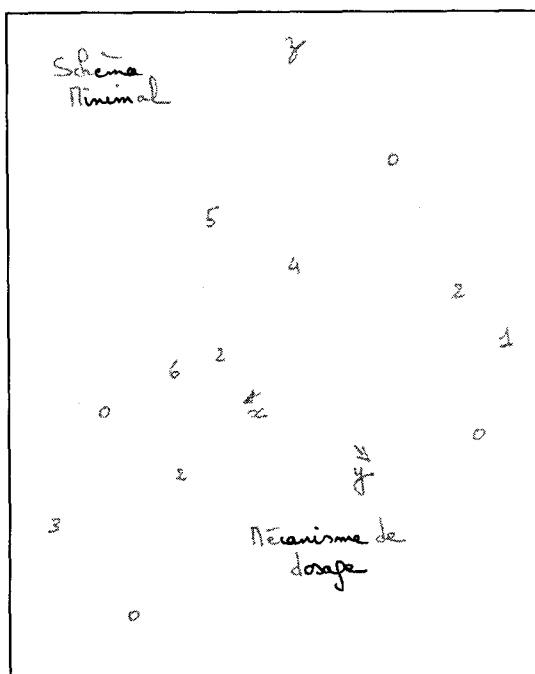


Figure 35. – Zones de texte.

La compréhension est complète si les données fournies par le modèle de représentation ne présentent aucune zone d'indécision ou d'incohérence.

Dans le cas contraire, le système interrogera l'utilisateur qui devra l'aider à comprendre et à résoudre les problèmes insolubles automatiquement.

4.3.3. d'autres exemples

En modifiant le plan régissant l'enchaînement des processus faisant évoluer la représentation du document, le système peut aussi s'adapter et analyser différents types de documents. En effet, la liste des spécialistes peut être modifiée dans son ordre et dans son contenu ou complétée en fonction du problème à résoudre sans que cela n'influence l'architecture globale du système. Les figures 36 et 37 montrent des exemples de résultats d'extractions réalisées

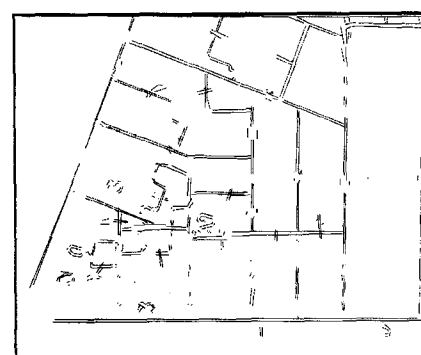
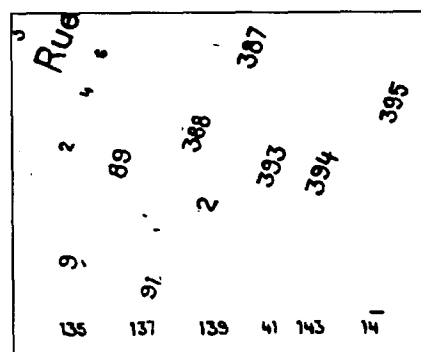
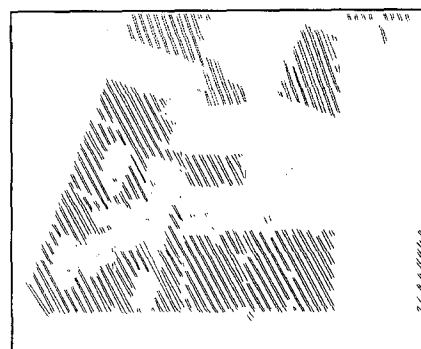


Figure 36. – Spécialistes hachures, texte, parcelles appliqués à l'image de la figure 18.

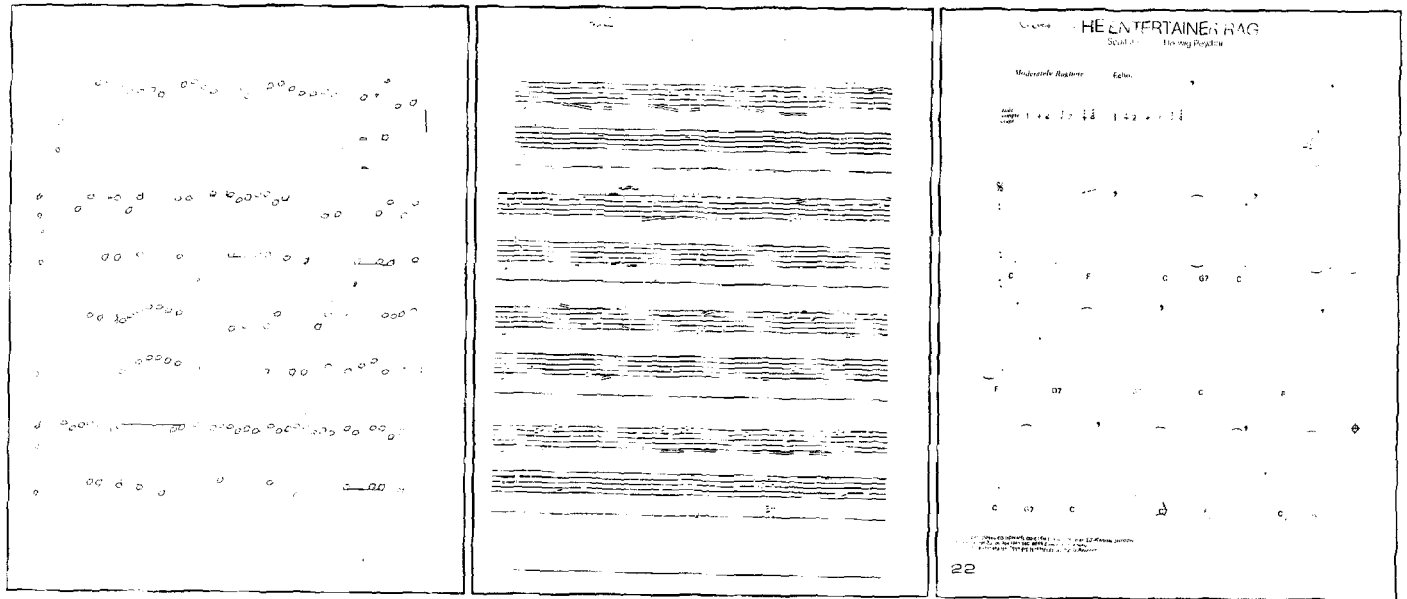


Figure 37. - Spécialistes têtes de notes noires, lignes de portées, texte, appliqués à l'image de la figure 20.

par des spécialistes adaptés à l'étude des planches cadastrales ou à l'étude des partitions musicales.

Même si les tests portent, pour l'instant, sur un nombre très limité d'images, les résultats sont encourageants et laissent entrevoir la possibilité de construction, sur cette base, d'une architecture logicielle générique permettant l'analyse de différents types de documents techniques.

5. conclusion

5.1. apports de la méthode

La validité de la méthode de construction de la représentation initiale a déjà été discutée dans le paragraphe 3.5, nous ne la reprendrons pas ici. Les spécialistes utilisent les primitives de la représentation pour communiquer. Ils se les approprient en modifiant les attributs, ils les associent pour constituer des objets plus complexes (courbes, éléments mécaniques, ...) et font évoluer la représentation. Cependant, dans le système actuel, les résultats fournis par la phase de perception globale (la représentation initiale) ne sont jamais remis en cause et un spécialiste ne peut pas utiliser une entité de la représentation déjà interprétée par un autre spécialiste. Chaque décision prise par un spécialiste constitue donc un résultat définitif. Une gestion des conflits entre spécialistes est envisageable sans grande modification du système et de récents travaux réalisés en analyse d'images [Lefèvre96] [Bellet96] montrent que cette voie est prometteuse. Ils mentionnent aussi de nombreux problèmes sur lesquels nous réfléchissons actuellement comme :

- la difficulté de définir exhaustivement les conditions d'utilisation des traitements (spécialistes) en fonction du contexte (appels d'offre),
- la difficulté d'évaluer les résultats des traitements,
- la difficulté d'élaborer des stratégies fusionnant des résultats fournis par des traitements différents.

Après chaque cycle, une analyse des informations contenues dans cette représentation permet de vérifier si l'analyse a été complète et, si nécessaire, de relancer un nouveau cycle perceptif.

La gestion du contexte à l'aide d'une représentation évolutive permet aux spécialistes de tenir compte des données extraites pour générer des hypothèses. Le mécanisme de cycles perceptifs permet une meilleure utilisation des connaissances obtenues et augmente les possibilités de coopération entre les processus d'analyse. L'extraction de données sûres durant les cycles perceptifs semble une solution efficace pour limiter les erreurs d'interprétation ainsi que les étapes de résolution des incohérences et leur effet boule de neige.

Il arrive que, lors de la détection ou de la reconstruction d'entités de haut niveau sémantique, une autre difficulté survienne au niveau de l'amorçage du processus lorsque le contexte sur lequel on se base est faible.

Le mécanisme d'évolution de la représentation du document, et de son mode de construction nous permet de résoudre le problème des éléments contenus simultanément dans deux entités. Bien souvent, la soustraction des entités reconnues (hachures, textes, ...) de l'image initiale pose beaucoup de problèmes, d'autre part leur maintien dans l'image rend plus difficile la détection des autres données. La fusion des quadrilatères, le suivi de tracé ont permis de déceler la bi-appartenance d'une intersection de traits à deux entités distinctes.

5.2. vers une architecture logicielle

Les systèmes de lecture automatique de documents techniques utilisant un contexte global (l'environnement) pour l'interprétation, dès les premières étapes du traitement, sont encore très rares. Très peu d'architectures permettent, aujourd'hui, de gérer et d'utiliser facilement les connaissances possédées sur une image à un instant donné. Lors de l'élaboration de notre système de lecture de plans cinématiques nous nous sommes efforcés de combler cette lacune. Pour s'adapter aux conditions changeantes de l'environnement et construire de manière progressive une solution en cohérence avec le contexte effectif, notre système comporte une première phase de perception globale du document suivie, ensuite, d'une phase de lecture et de compréhension.

La phase de perception globale consiste en la construction d'une représentation du document à partir d'un nombre limité de primitives élémentaires pour fournir une description précise de toutes les formes composant l'image initiale. Ces primitives possèdent de grandes capacités à représenter les connaissances recherchées et/ou celles que l'on a pu acquérir à chaque niveau d'abstraction. Nous avons montré que cette représentation initiale permet la description de documents techniques de types variés (dessins, partitions musicales, cadastres, ...) lorsque ceux-ci se trouvent, au départ, sous forme d'images binaires.

Ensuite, des actions effectuées par différents spécialistes peuvent être coordonnées afin d'arriver à lire et à appréhender le contenu d'un document. Dans le système que nous proposons les spécialistes s'efforcent, durant chacun des cycles perceptifs, d'interpréter les schémas cinématiques, de manière hiérarchique et réfléchie pour faire évoluer la représentation des documents afin d'obtenir une analyse plus robuste de leur contenu.

L'architecture choisie s'articule autour de spécialistes chargés de réaliser des tâches précises contribue, là aussi, à augmenter la généralité de la méthode puisqu'il est possible de créer facilement de nouveaux spécialistes et de modifier l'ordre de leur activation afin de traiter d'autres types de documents.

Dans le système proposé, les connaissances *a priori* sur la structuration des documents à traiter se trouvent autant dans la manière dont on choisit de faire évoluer la représentation, que dans la structure de la représentation elle-même. La représentation initiale permet la description fidèle de différents types de documents techniques ce qui ouvre d'intéressantes perspectives aussi bien dans la parallélisation des tâches des différents spécialistes que dans le domaine de la prise de décision.

BIBLIOGRAPHIE

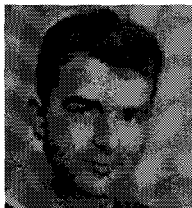
- [Antoine92] Antoine, D., Collin, S. and Tombre, K. Extraction d'indices : la méthode de REDRAW. In : Structured Document Analysis. Baird, H.S., Bunke, H. and Yamamoto, K. Eds. Berlin (Allemagne) : Springer Verlag, 1992. p. 385-402.
- [Belaïd92] Belaïd, A. et Tombre, K. Analyse de documents : de l'image à la sémantique. *Bigre*, 1992, N 80, p. 3-28.
- [Bellet96] Bellet, F. et Garbay, C. Des processus adaptatifs et coopératifs pour la vision bas niveau. In : Actes du 10e congrès Reconnaissance de Formes et Intelligence Artificielle, Rennes (France), 16- 18 janvier, 1996. p. 516-525.
- [Bruce93] Bruce, V. et Green, P.R. La perception visuelle : Physiologie, psychologie et écologie. Grenoble : Presses universitaires de Grenoble, 1993. 511 p.
- [Clavier97] Clavier, E. et Clavier, S. Tri et catégorisation de documents numérisés. 2e Journée jeunes chercheurs « Ecrit et Document ». Paris. 13 juin 1997.
- [Dori95] Dori, D. and Tombre, K. From engineering drawings to 3D CAD models : are we ready now? *Computer Aided Design*, 1995, Vol. 29, N 4, p. 243-254.
- [Eglin97] Eglin, V. and Emptoz, H. Low-resolution boundaries for guiding eye-movement on a document. Third International Workshop on Visual Form, Capri (Italie), 28-30 mai 1997.
- [Fortin92] Fortin, C. et Rousseau, R. Psychologie cognitive : une approche du traitement de l'information. Québec : Presses de l'université du Québec, 1992. 434 p.
- [Kadonaga95] Kadonaga, T. and Abe, K. Comparison of methods for detecting corner points from digital curves. *Lecture Notes in Computer Science 1072. Graphics Recognition, Methods and Applications*. R. Kasturi and K. Tombre Eds. 1995. p. 23-34.
- [Kasturi90] Kasturi, R. A system for interpretation of line drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, Vol. 12, N 10, p. 978-991.
- [Kasturi92] Kasturi, R. An Overview of techniques for graphics recognition. In : *Structured Document Analysis*. Baird, H.S., Bunke, H. and Yamamoto, K. Eds. Berlin (Allemagne) : Springer Verlag, 1992. p. 285-324.
- [Lecas92] Lecas, J.C. L'attention visuelle : de la conscience aux neurosciences. Bruxelles (Belgique) : Mardaga Eds, 1992. 300 p.
- [Lefèvre96] Lefèvre, V., Pollet, Y., Philipp, S. et Brunesseaux, S. Un système multi-agents pour la fusion de données en analyse d'images. *Traitement du Signal*, 1996, Vol. 13, N 1, p. 99-111.
- [Marr82] Marr, D. *Vision : A computational investigation into human representation and processing of visual information*. New-York (USA) : W.H. Freeman and company, 1982. 397 p.
- [Martin90] Martin, P. et Bellissant, C. Segmentation et classification par réseau pour la reconnaissance de partitions. In : *Colloque de Reconnaissance Automatique de l'Ecrit*. 1990. p. 102-111.
- [Nagy95] Nagy, G. Document image analysis : what is missing? In : *Proceedings of the 8th International Conference on Image Analysis and Processing*, San-Remo (Italy), september, 1995. p. 577-587.
- [Ogier95] Ogier, J.M., Mullot, R., Labiche, J. et Lecourtier, Y. Interprétation de documents par cycles perceptifs de construction d'objets cohérents. Application aux données cadastrales. *Traitement du Signal*, 1995, Vol. 12, N 6, p. 627-638.
- [Ramel96] Ramel, J.Y. *Lecture automatique de dessins, méthodes d'analyse et de reconnaissance, application aux plans cinématiques*. Thèse de doctorat (INSA de Lyon). 1996. 156 pages.
- [Sayre73] Sayre, K.M. Machine recognition of Handwritten words : A project report. *Pattern Recognition*, 1973, Vol. 5, p. 213-228.
- [Shih89] Shih, C. and Kasturi, R. Extraction of graphic primitives from images of paper based line drawings. *Machine Vision and Applications*, 1989, Vol. 2, p. 103-113.
- [Simon93] Simon, P. Contribution de l'analyse des mouvements oculaires à l'évaluation de la charge de travail mental. Thèse de doctorat (Université de Valenciennes). 1993. 183 pages.
- [Tombre92] Tombre, K. Technical drawing Recognition and understanding : From Pixel to semantics. In : *Proceedings of the IAPR Workshop on Machine Vision and Application*. Tokyo (Japon), 7-9 December, 1992. p. 393-401.
- [Ts95] Numéro Spécial Ecrit et Documents. *Traitement du Signal*, Vol. 12, N 6, 1995.
- [Wall84] Wall, K. and Danielsson, P. A fast sequential method for polygonal approximation of digitized curves. *Computer Vision, Graphics and Image Processing*, 1984, Vol. 28, p. 220-227.

Manuscrit reçu le 11 mars 1997.

Interprétation de documents techniques par «cycles perceptifs»

LES AUTEURS

Jean-Yves RAMEL



Jean-Yves Ramel travaille actuellement au sein du laboratoire Reconnaissance de Formes et Vision de l'INSA de Lyon. Il a soutenu sa thèse de doctorat concernant la lecture automatique de plans mécaniques en novembre 1996. Il travaille aujourd'hui, toujours dans le domaine de l'interprétation automatique de documents, plus particulièrement sur les techniques de coopération et d'intelligence artificielle distribuée pour l'analyse d'images.

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 aspects 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 consacré au domaine de l'écrit et du document.

Nicole VINCENT



Ancienne élève de l'ENS de Fontenay et agrégée de Mathématiques, Nicole VINCENT s'est tournée vers l'Informatique au cours des années quatre vingt et a soutenu en 1988 une thèse de doctorat sur la reconnaissance multipolice des caractères dactylographiés et imprimés off line. Maître de conférences à l'INSA en 1989, puis professeur à l'Université de Tours, elle a orienté tous ses efforts vers la reconnaissance de l'écrit. Elle a attiré l'attention sur l'intérêt des fractales dans l'analyse des écritures et mis à l'épreuve différentes techniques de recherche qu'elle s'efforce de faire coopérer, persuadée que la solution ne peut résulter de l'utilisation d'une méthode unique.