

# Mémoire visuelle pour l'analyse de documents structurés

Joseph Chazalon, Bertrand Coüasnon, Aurélie Lemaitre

► **To cite this version:**

Joseph Chazalon, Bertrand Coüasnon, Aurélie Lemaitre. Mémoire visuelle pour l'analyse de documents structurés. Colloque International Francophone sur l'Écrit et le Document (CIFED2010), Mar 2010, Sousse, Tunisie. 2010, Actes du 11ème Colloque International Francophone sur l'Écrit et le Document. <inria-00545886>

**HAL Id: inria-00545886**

**<https://hal.inria.fr/inria-00545886>**

Submitted on 13 Dec 2010

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

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

---

# Mémoire visuelle pour l'analyse de documents structurés

**J. Chazalon\*** — **B. Couïasnon\*** — **A. Lemaitre\*\***

\* INSA de Rennes, Avenue des Buttes de Coësmes, F-35043 Rennes

\*\* Université de Rennes 2, Campus de Villejean, F-35043 Rennes

UMR IRISA, Campus de Beaulieu, F-35042 Rennes

Université Européenne de Bretagne, France

{joseph.chazalon,bertrand.couasnon,aurelie.lemaitre}@irisa.fr

---

## RÉSUMÉ.

Les méthodes d'analyse de documents actuelles proposent des systèmes qui sont adaptés pour reconnaître des lots de documents d'un même type (méthodes statistiques, analyses grammaticales), mais qui ne sont ensuite appliqués qu'isolément sur chacune des pages à traiter. Pourtant, dans un contexte de collection, il est important de profiter du résultat du traitement d'une ou plusieurs pages pour améliorer le traitement ultérieur d'autres pages. Nous proposons donc un cadre, la mémoire visuelle, pour intégrer aux systèmes de reconnaissance existants la prise en compte locale et en cours d'analyse d'informations contextuelles utiles pour chaque page. Nous détaillons ses spécifications qui permettent son implémentation dans n'importe quel système de reconnaissance. Nous présentons enfin une mise en œuvre dans un système grammatical et déclinons différents schémas de circulation d'information permis par la mémoire visuelle.

## ABSTRACT.

Current analysis methods propose systems which are adapted to recognize sets of documents of a same kind (statistical methods, grammatical analysis), but each page is then processed in isolation. However, in a collection context, it is important to make the most of the result of the processing of one or several pages in order to improve the later processing of other pages. Thus, we propose the concept of visual memory so as to enable existing recognition systems to use, locally and at run time, contextual information for each page. We detail its specifications which allow an implementation in any recognition system. Finally, we present our implementation using a grammatical method, and its application for several types of information flow.

**MOTS-CLÉS :** Analyse de documents, grammaire visuelle, mémoire, contexte, méthode générique

**KEYWORDS :** Document analysis, visual grammars, memory, context, generic method

---

## 1. Introduction

Dans le domaine de la reconnaissance d'images de documents, nous sommes fréquemment amenés à traiter des ensembles d'images ayant des caractéristiques communes : même type de contenu, structure similaire, contenu réparti sur plusieurs pages, etc. Actuellement, la plupart des systèmes de reconnaissance se contentent d'analyser les pages isolément, sans tirer profit des informations contextuelles propres à la notion de collection. Pourtant, l'utilisation du contexte d'une page au cours de son analyse peut permettre d'améliorer le résultat produit. On peut ainsi imaginer, par exemple, repérer les éléments structurels localement stables sur un ensemble réduit de pages pour ensuite mieux détecter la structure des pages endommagées.

Afin de permettre la production et l'exploitation d'informations contextuelles issues de la collection lors de l'analyse d'une image de document, nous proposons de dissocier deux niveaux dans un tel système : un niveau page qui produit ou exploite ces informations ; et un niveau global qui gère leur circulation entre les différents traitements de page. Cet article introduit alors la *mémoire visuelle*, concept générique et indépendant du système utilisé, destinée à être le *vecteur* de l'information contextuelle. L'aspect *mémoriel* permet la persistance d'informations variées entre les traitements de pages ; et l'aspect *visuel* permet de localiser les connaissances contextuelles, facilitant leur fusion avec les données issues des images.

En partie 2, nous présentons quelques approches existantes exploitant l'information contextuelle, et nous listons les besoins qu'un nouveau système doit satisfaire pour dépasser leurs limites. En partie 3, nous présentons le mécanisme générique de mémoire visuelle et l'ensemble des propriétés nécessaires à sa mise en œuvre. En partie 4, nous présentons un exemple d'implémentation basé sur le système de reconnaissance grammatical DMOS-P. Enfin, en partie 5, nous validons l'intérêt de l'utilisation du contexte sur une base de plus de 39 000 pages, et nous montrons plusieurs exemples de schémas de circulation de données contextuelles utilisant la mémoire visuelle.

## 2. État de l'art

Dans le cadre général de la reconnaissance d'images de documents structurés, quelques systèmes exploitent des informations contextuelles, c'est-à-dire produites précédemment par un autre traitement de page, pour améliorer le résultat du traitement de page courant. Nous en détaillons ici une sélection et dressons un bilan des besoins à satisfaire pour exploiter efficacement ce contexte.

### 2.1. Systèmes existants

Pour commencer, deux systèmes dédiés à un type précis de document permettent de montrer la variabilité des données contextuelles et de leurs schémas de circulation possibles selon les documents considérés. (Lin *et al.*, 1997), d'abord, ont proposé un système de reconnaissance de livres qui permet une vérification croisée des contenus présents à la fois dans la table des matières et dans les autres pages, en particulier les titres des sections. L'information extraite de la table des matières est ensuite distribuée

vers les traitements des autres pages. (He *et al.*, 2003) ont, quant à eux, proposé un outil pour la reconnaissance de cartes d'index de spécimens archivées au *UK Natural History Museum*. Les auteurs tirent profit de l'organisation hiérarchique de ces cartes (sous-famille, famille...) séparées par des cartes spéciales, pour propager d'une image de carte à sa suivante des informations de groupe, sauf si un séparateur est reconstruit.

Les deux systèmes précédents montrent l'intérêt du contexte lors de l'analyse d'une image de document, mais la nature de l'information qu'ils exploitent y est restreinte, et sa circulation se fait selon un seul schéma préétabli. L'adaptation de tels systèmes à de nouveaux types de documents est alors difficile, par manque de généralité. Une approche différente consiste à exploiter une base de données centrale, qui permet le stockage d'informations variées et plus de souplesse dans leur circulation. C'est le cas de smartFIX (Klein *et al.*, 2004), un système de traitement automatique de courriers entrants. Sa base centrale lui permet de partager immédiatement entre les traitements de pages des informations de nature variée : données reconnues, informations statistiques apprises, etc.

Malgré une certaine flexibilité de l'architecture à base centrale, elle impose à chacun des modules d'analyse de page d'être capable d'accéder à la base centrale pour publier ou récupérer l'information qui lui est utile. La connaissance relative à la structure de la collection, permettant le routage de l'information contextuelle, est alors *fragmentée* entre les descriptions des traitements de pages, complexifiant leur écriture et diminuant leur généralité. Nous avons récemment proposé (Chazalon *et al.*, 2010) de séparer clairement la description de la collection (nature des pages, structure de la séquence, et circulation de l'information) et les descriptions des traitements de pages, qui se concentrent alors sur l'extraction de contenus. Cette idée de centraliser la gestion de la circulation de l'information fait écho à d'autres travaux (Baird *et al.*, 1995) qui éliminent cette préoccupation des traitements travaillant sur des pages ou des fragments de pages. Notre approche utilise une grammaire attribuée pour décrire précisément la circulation de l'information entre les traitements de page, mais elle rend rapidement fastidieux l'ajout de nouvelles informations à échanger. Il apparaît donc nécessaire de dépasser ces limitations en permettant de systématiser la circulation de l'information.

## 2.2. Bilan des besoins

La séparation de la modélisation en deux niveaux, la page et la collection, nous a partiellement permis de centraliser au niveau de la collection l'expression de la connaissance sur la circulation d'information entre les pages, et de simplifier la description des traitements de pages en les rendant réutilisables au sein d'autres collections ou circulations. Nous souhaitons poursuivre ce travail pour mettre en œuvre un mécanisme systématique visant à exploiter pleinement, au sein du traitement *local* d'une page, l'information *globale* disponible dans la collection, provenant d'une autre page ou d'un autre traitement. Ce dernier doit donc permettre : (1) d'exploiter des informations de nature variée ; (2) de mettre en place, au niveau de la collection, des schémas de circulation variés et adaptables en cours d'analyse, entre différents traitements ou

différentes pages ; (3) de traiter chaque page sans devoir se soucier de la structure de la collection ; (4) de fusionner, au niveau de chaque page, les informations provenant du contexte avec celles provenant de l'image courante.

### 3. Mémoire visuelle

Afin de répondre aux besoins identifiés dans l'état de l'art, nous proposons dans cette section le mécanisme de *mémoire visuelle*. C'est un concept générique qui permet d'améliorer un système d'analyse de collection de documents, sans être tributaire d'une implémentation particulière. Il est donc adaptable à n'importe quel système existant d'analyse de documents. Cette mémoire visuelle est conçue autour de deux aspects essentiels : l'aspect mémoriel et l'aspect visuel.

L'aspect *mémoriel* fournit la capacité, au cours d'un traitement, de mémoriser et d'accéder à des informations variées, sans se soucier ni du traitement, ni de la page d'origine ou de destination de ces données. L'aspect *visuel* permet de manipuler, au cours de l'analyse d'une page, les données mémorisées comme si elles faisaient partie de l'image. Il fournit aux données une forme et une position au sein du même repère bidimensionnel que l'image, et facilite la fusion des deux types de données.

La mémoire visuelle sert de *support* à la circulation d'information entre deux niveaux : un niveau *collection*, gérant la circulation flexible d'information au sein du lot de document dont il décrit la structure et les stratégies d'analyse ; et un niveau *page*, où pour chaque type de page on peut exprimer un traitement centré sur le travail de reconnaissance.

Dans cette partie, nous détaillons les propriétés propres à la mémoire visuelle qui seront nécessaires pour sa mise en place : l'aspect mémoriel, l'aspect visuel, ainsi que les facultés de support de circulation de l'information.

#### 3.1. Aspect mémoriel

Afin de permettre la mise en place d'une méthode générique de persistance et de transmission de l'information entre les traitements de pages, il est nécessaire : (1) de connaître l'information qui va être manipulée ; (2) de déterminer quand et comment elle est manipulable.

##### 3.1.1. Nature variée des données

La mémoire visuelle doit être en mesure de stocker des éléments de natures très variées. Par exemple, on peut citer : des données provenant de la structure abstraite reconnue dans d'autres pages (paragraphe, article de journal, fragment d'un acte de vente) ; des dictionnaires recensant le vocabulaire autorisé pour la reconnaissance de certains éléments (qui pourront alors être étendus lors de la découverte de nouveaux éléments) ; des indices de position pour guider la recherche d'un élément particulier ayant une position assez stable au sein de la collection (titre, numéro de page) ; des seuils appris ; des statistiques sur la collection ; des demandes d'interaction avec un

opérateur humain ou ses réponses ; des éléments d'ordre supérieur, tels que des règles applicables pour la reconnaissance (si l'implémentation le supporte)...

Une telle variété peut être rendue possible par la définition d'un type d'éléments *mémorisables*, qui sont des conteneurs encapsulant toutes les informations nécessaires, et autorisant une extension aisée des éléments échangés.

### 3.1.2. *Omniprésence de la mémoire*

Il est nécessaire de pouvoir avoir accès à l'information mémorisée, c'est à dire provenant d'une autre page ou d'un autre traitement, lors de chaque traitement et à chaque moment de l'analyse. Cette information doit donc être omniprésente, et le système doit permettre, à n'importe quel moment de l'analyse d'une page : de *mémoriser* des informations nouvelles ; d'*utiliser* (se rappeler) des informations mémorisées ; d'*effacer* (oublier) des informations mémorisées. Cette mémoire doit donc être matérialisée, au sein du traitement d'une page, par une structure accessible à chaque étape de l'analyse. Bien entendu, la logique des actions réalisées au sein d'un traitement (mémorisation, oubli ou réminiscence) dépend du travail à réaliser sur la collection de documents, et doit être définie de façon cohérente par le concepteur.

## 3.2. *Aspect visuel*

Le caractère bidimensionnel des images traitées induit un aspect « visuel » des données manipulées, dans lequel les données mémorisées doivent pouvoir être fusionnées avec celles issues de l'image en cours. L'aspect visuel consiste donc à : **(1)** donner une *forme* et une *position* aux données mémorisées, dans le même repère bidimensionnel que celui de l'image ; **(2)** fournir un *accès homogène*, simultané, aux informations, qu'elles proviennent de la page ou de la collection.

### 3.2.1. *Représentation bidimensionnelle homogène*

Nous proposons de représenter les données mémorisées dans le même référentiel bidimensionnel que les informations issues de l'image. Cette *représentation homogène* des deux types de données facilite alors leur mise en correspondance. À chaque conteneur encapsulant l'information mémorisée (cf. sec. 3.1.1) doivent donc être associées : une *forme*, qui donne une portée spatiale à la donnée ; et une *position*, qui la localise dans le référentiel de l'image courante. Notons ici encore que la sémantique de cette forme et de cette position associées à l'information mémorisée dépend du traitement à réaliser, et doit donc être clairement définie par le concepteur.

### 3.2.2. *Accès homogène aux données*

Afin de mettre effectivement en correspondance les informations provenant soit de la collection (informations mémorisées), soit de la page (données issues de l'image), il faut être en mesure d'accéder de manière homogène et simultanée à ces deux types de données hétérogènes. Il faut également que les opérations primitives d'accès aux données mémorisées (alors omniprésentes) soient identiques à celles permettant de lire les données issues de l'image. Il faut donc assurer une *polyvalence* des opérateurs de consultation, de consommation, d'ajout et de transformation des données.

Il est également important de pouvoir continuer à distinguer l'origine de l'information : le concepteur pourra, par exemple, souhaiter définir des priorités entre les données mémorisées et les données issues de l'image, selon son problème. La définition de cette sémantique est à la charge du concepteur du système, car elle est, une fois encore, dépendante du type de travail à réaliser.

### **3.3. Support de circulation des données**

La mémoire permet de définir plusieurs niveaux d'abstraction au sein du système développé, en servant de *support* à la communication de l'information. Ceci autorise alors à : d'une part, mettre en place des échanges riches entre les pages ou les traitements ; et, d'autre part, simplifier la description du problème en structurant le système en deux niveaux : un niveau *global* et un niveau *page*.

#### **3.3.1. Rôles de chacun des niveaux**

Le *niveau global*, commun à la collection, centralise la gestion de la structure de la collection, des dépendances entre les traitements et de l'acheminement de l'information entre les pages ou entre les traitements. Il n'a pas connaissance des détails du fonctionnement de chaque traitement : il les compose pour former un tout cohérent.

Le *niveau page*, quant à lui, se concentre sur l'extraction du contenu de l'image. En exploitant la mémoire visuelle, il peut utiliser ou produire de l'information contextuelle sans savoir ni où, ni quand cette information a été produite ou sera utilisée.

#### **3.3.2. Circulation des données**

La mémoire visuelle sert de support pour la circulation d'informations mémorisées entre pages ou entre traitements. Pour permettre les échanges d'information, il faut définir une interface standardisée entre le niveau global et le niveau page. Celle-ci doit permettre, d'une part l'invocation à la demande, par le niveau global, d'un traitement de niveau page sur une image donnée, et d'autre part le passage ou la récupération d'informations mémorisées (contenu de la mémoire visuelle) entre ces niveaux. Le niveau global doit être capable de manipuler simplement les données mémorisées associées à une image, et chacun des deux niveaux doit disposer d'opérateurs d'envoi ou de réception de données mémorisées.

### **3.4. Bilan des propriétés requises**

Pour permettre la mise en place d'un mécanisme de mémoire visuelle et l'intégrer, il est donc nécessaire : **(1)** de définir un type de données mémorisables, autorisant une encapsulation facilement extensible ; **(2)** de fournir une structure de données mémorisées omniprésente, modifiable, accessible à tout moment lors de l'analyse d'une page ; **(3)** d'associer une forme et une position dans le repère de l'image pour les données mémorisables ; **(4)** de fournir des primitives de consultation, d'ajout, de modification et de suppression identiques entre les données issues de l'image et celles issues de la collection (en mémoire) ; **(5)** de fournir un mécanisme de communication entre le niveau page et le niveau global.

Ces propriétés répondent aux besoins exposés en conclusion de l'état de l'art (section 2.2). La propriété 1 permet de manipuler des données de types variés (besoin 1). La propriété 5 permet la persistance d'information contextuelle et sa circulation variée entre les traitements ou pages (besoin 2), tout en permettant une séparation claire de la modélisation. Les propriétés 2 et 5 permettent aux traitements de niveau page de ne plus se préoccuper de la structure de la collection (besoin 3), et donc de simplifier les descriptions au niveau de chaque page. La combinaison des propriétés 2, 3 et 4 permet la fusion efficace des informations locales, issues de l'image, et des informations contextuelles, issues de la collection (besoin 4).

#### **4. Exemple de mise en œuvre de la mémoire visuelle**

Les propriétés énoncées dans la partie précédente permettent d'introduire la mémoire visuelle dans tout système de reconnaissance. Nous en présentons maintenant un exemple de mise en œuvre dans une méthode grammaticale existante : DMOS-P.

##### **4.1. Méthode DMOS-P**

La méthode DMOS-P (Description et MOdification de la Segmentation) est une méthode de reconnaissance grammaticale d'images de documents structurés (Coüasnon, 2006). La description syntaxique et sémantique du type de document à reconnaître est écrite par un expert dans le formalisme EPF (*Enhanced Position Formalism*), puis compilée pour produire automatiquement l'analyseur associé. La généralité de cette méthode a été validée sur des types de documents variés (tableaux, partitions musicales, registres d'archives) et à grande échelle (plus de 500 000 pages traitées). De plus, cette méthode permet de prendre en compte plusieurs niveaux de perception pour chaque image à analyser (Lemaitre *et al.*, 2010). Par exemple, il est possible de confronter les visions d'une même image à différents niveaux de résolution.

C'est dans ce contexte que nous validons le concept de mémoire visuelle, en mettant en œuvre les différentes propriétés exposées dans la partie 3.4. Ce travail est réalisé de manière assez simple puisque la méthode DMOS-P offre déjà plusieurs des propriétés requises.

##### **4.2. Données mémorisables**

La mémoire visuelle nécessite de manipuler des données de types variés (propriété 1), encapsulées dans un type de données mémorisables. Dans notre implémentation, nous proposons donc des conteneurs capables d'encapsuler tous types d'objets. Ces conteneurs sont caractérisés par une position dans l'image et par une forme (rectangulaire ou rectiligne). Grâce à ce système, il est donc possible de manipuler des données de nature variée de manière homogène avec les terminaux de la grammaire en EPF (composantes connexes et segments).



### 4.3. *Données omniprésentes dans un repère homogène*

Nous avons ensuite montré (propriétés 2 et 3) qu'il est nécessaire de disposer de données mémorisées omniprésentes, dans un référentiel commun avec celui de l'image et d'uniformiser l'accès aux données, qu'elles proviennent de l'analyse de l'image courante ou de la mémoire.

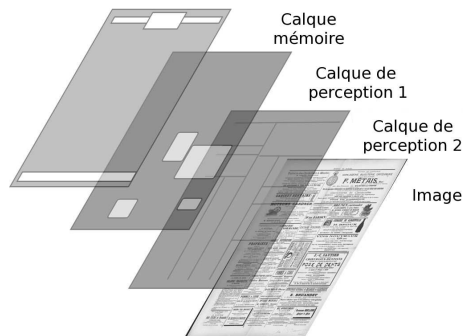
La méthode DMOS-P existante se base sur le concept des *calques perceptifs* (Lemaitre *et al.*, 2010) qui organisent les différents niveaux de perception de l'image en autant de structures accessibles à chaque moment de l'analyse. Chaque calque perceptif contient l'ensemble des terminaux de la grammaire qui ont été extraits dans l'image, à un niveau de perception donnée. Chaque terminal de l'image est caractérisé par une position dans l'image et une forme (rectangulaire ou longiligne). Lors de la description grammaticale du document, le concepteur de la grammaire peut donc définir, à chaque étape, le type de terminal à reconnaître, ainsi que le niveau de perception à considérer, c'est à dire le calque perceptif à étudier.

Nous avons proposé de créer un nouveau calque perceptif qui contient la mémoire visuelle : le calque de mémoire. Les éléments de la mémoire sont manipulés sous forme de terminaux de la grammaire, ayant une forme et une position dans l'espace de l'image. Par exemple, la figure 1 montre l'ensemble des terminaux disponibles pour l'analyse d'une page : ceux extraits aux deux résolutions étudiées, dans les deux calques de perceptions, et ceux présents dans le calque de mémoire. Cette utilisation des calques perceptifs permet donc bien d'assurer un référentiel commun, l'omniprésence des éléments contenus dans la mémoire ainsi qu'un accès homogène aux informations, qu'elles proviennent de la mémoire ou de l'image en cours de traitement

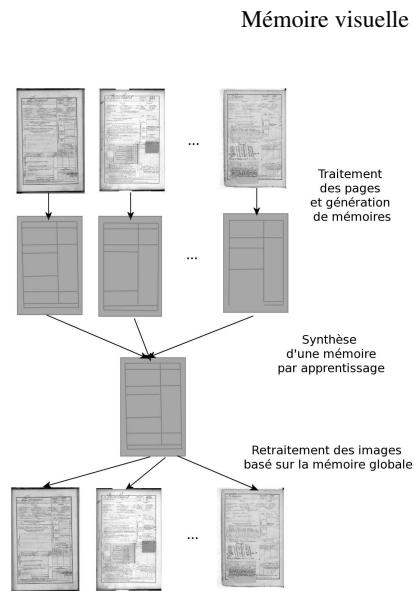
### 4.4. *Manipulation des données*

Nous avons montré (propriété 4) qu'il est nécessaire de prévoir trois types d'opérations sur la mémoire visuelle : (1) ajouter des informations en mémoire qui pourront être utilisées dans d'autres traitements d'images ; (2) consulter la mémoire pour y lire des informations stockées précédemment ; (3) effacer des informations mémorisées. Puisque les objets mémorisés sont des terminaux stockés dans un calque perceptif, ces actions reviennent dans le cas de notre implémentation à : *ajouter des terminaux* dans le calque perceptif de mémoire ; *analyser des terminaux* présents dans le calque perceptif de mémoire, *avec ou sans consommation* (effacement).

L'ajout d'informations dans le calque mémoire est réalisé grâce au nouvel opérateur EPF `STORE_IN_LAYER Term` dont le rôle est de stocker le terminal `Term` dans la mémoire. Ce dernier permet de spécifier, au fur et à mesure de l'analyse de n'importe quel document, quelles sont les informations à stocker dans la mémoire pour être réutiliser lors de l'analyse future d'autres pages. Pour *consulter* le contenu de la mémoire, *avec ou sans consommation*, il est possible d'utiliser les opérateurs classiques du langage EPF servant à la reconnaissance des terminaux puisque les éléments mémorisés sont considérés comme des terminaux de la grammaire.



**Figure 1.** Avec les calques, utilisation conjointes de perceptions à deux résolutions et d'informations contextuelles contenues dans la mémoire



**Figure 2.** Utilisation de la mémoire visuelle pour intégrer de l'apprentissage de dimension dans la reconnaissance de pages de registres matricules

L'originalité de ce travail est donc d'intégrer complètement le remplissage et la consultation de la mémoire dans la description grammaticale, ce qui autorise le concepteur à choisir, en fonction du type de document étudié, les endroits et les cas où il est pertinent d'utiliser des informations provenant du contexte global.

#### 4.5. Circulation des données

Le dernier élément de notre implémentation (propriété 5) est lié à la circulation des données d'un traitement de page à l'autre. Nous avons créé deux opérateurs dans le langage EPF : `LOAD_MEMORY FichierMem` et `SAVE_MEMORY FichierMem` dont le rôle est d'importer (charger dans le calque mémoire) et d'exporter (sauvegarder le calque mémoire) des informations contextuelles mémorisées.

### 5. Applications

Nous détaillons ici trois exemples, qui montrent les possibilités que peut offrir la mémoire visuelle dans n'importe quel système de reconnaissance de documents. Nous nous intéressons à montrer que la gestion des échanges entre les outils de traitement des images peut bénéficier de cette mémoire visuelle, car elle simplifie l'expression et la mise en œuvre de comportement globaux sophistiqués. La façon dont les données mémorisées sont exploitées au sein des traitements de page n'est pas présentée ici.

### 5.1. *Apprentissage de dimensions*

Nous présentons ici des documents d'archives (fin XIX<sup>e</sup> et début XX<sup>e</sup>) à structure tabulaire : les registres matricules. Cette collection de formulaires de recrutement militaire contient à la fois des documents pour lesquels la structure tabulaire apparaît clairement, mais aussi beaucoup d'autres où la structure tabulaire est masquée, mal imprimée ou très abîmée. L'utilisation de la mémoire visuelle (illustrée figure 2) permet dans ce cas de mettre en place un apprentissage des dimensions des cases de la structure tabulaire pour aider à reconnaître les pages plus difficiles.

Dans un premier temps, on essaie de reconnaître au mieux la structure de chaque page, et on mémorise le résultat dans un calque mémoire. Ensuite, un mécanisme d'apprentissage non supervisé parcourt l'ensemble des mémoires produites et génère une nouvelle mémoire visuelle contenant les dimensions supposées de la structure tabulaire à reconnaître. Cette mémoire peut alors servir de support pour le nouveau traitement des images qui pourront être mieux reconnues. Ce mécanisme d'apprentissage des dimensions des cases peut être mis en place car, si d'un registre à l'autre (env. 500 pages) la structure physique peut varier fortement (pré-imprimé différent), au sein d'un même registre les dimensions sont stables puisque c'est le même pré-imprimé.

Nous avons effectué une évaluation de cette utilisation du contexte apporté par la mémoire visuelle en comparant les taux de reconnaissance (tableau 1) de la structure tabulaire sur 84 registres de 1910 à 1914, soit 39 653 pages. Les documents étant fortement dégradés, une reconnaissance isolée de chaque page ne permet de détecter correctement que 62,67% des structures, alors que la détection en contexte, à l'aide de la mémoire visuelle, permet de passer à 98,22% de structures correctement détectées.

	Sans contexte		Avec contexte	
	Nb pages	Taux	Nb pages	Taux
Structure bien détectée	24 854	62,67%	38 948	98,22%
Structure non détectée	14 799	37,32%	705	1,78%

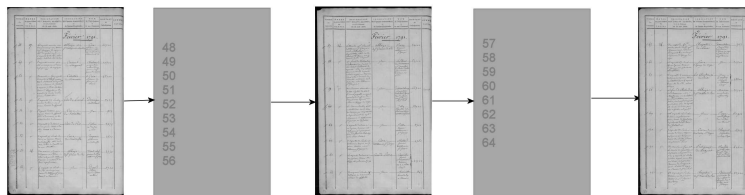
**Tableau 1.** *Reconnaissance de structures de 39 653 pages de registres matricules de 1910 à 1914, sans puis avec le contexte apporté par la mémoire visuelle*

### 5.2. *Optimisation sur un lot de page*

Nous nous intéressons ici à un autre type de document d'archive à structure tabulaire : des registres de ventes datant de la Révolution Française. Ces registres se présentent sous forme de pages de tableaux où chaque ligne décrit une vente. Les colonnes indiquent alors : numéro de vente, nom de l'acheteur, désignation du bien, etc. Nous souhaitons localiser et reconnaître les numéros de chacune des ventes, et nous savons que ces numéros sont en ordre croissant, respectant un incrément de 1 entre entrée successive d'un registre. L'exploitation de la mémoire visuelle autorise alors la mise en place d'une circulation des numéros reconnus entre pages successives, per-

mettant de guider le processus de reconnaissance des numéros au sein d'une page, en se basant sur ceux reconnus dans la page précédente.

La figure 3 illustre cette stratégie : lors de l'analyse d'une page, le traitement associé localise les nombres à reconnaître, et les traite en s'aidant des nombres mémorisés provenant de la page précédente. Ensuite, il stocke à son tour dans la mémoire ces nombres pour guider l'analyse de la page suivante. La mémoire visuelle permet ici la collecte et l'optimisation d'une information répartie sur un ensemble de pages.



**Figure 3.** Utilisation de la mémoire visuelle pour utiliser le principe de croissance des actes en transmettant les numéros trouvés

### 5.3. Indice de position en interactivité

Le troisième exemple d'application proposé concerne la reconnaissance de pages de presses anciennes. Nous cherchons à localiser la date de parution du journal, dont la position approximative dans la page peut rester stable pendant plusieurs années pour une édition donnée. En revanche, comme les pages peuvent être mal imprimées, déchirées, abimées, la date n'est pas toujours facilement positionnable sur le document.

En utilisant la mémoire visuelle, nous pouvons introduire la connaissance d'un expert qui donne un indice de position de la date pour la collection en cours d'étude. L'information localisée est introduite dans la mémoire visuelle prise en entrée lors de l'analyse de chacune des pages. Cet exemple montre comment la mémoire visuelle peut servir de support pour une interaction ponctuelle avec un expert.

Les exemples présentés ici illustrent la variété des mécanismes de circulation permis par la mémoire visuelle, qui pourraient être combinés pour gérer des documents plus complexes, sans que cela ne complique la description de chaque page.

## 6. Conclusion

Nous avons proposé dans cet article un mécanisme original et générique, la *mémoire visuelle*, qui vise à améliorer la reconnaissance de pages de documents similaires, sans être dépendant d'une implémentation particulière. La mémoire visuelle permet d'exploiter facilement, lors du traitement d'une page, les informations pertinentes qui pourraient avoir été extraites précédemment dans d'autres pages. Ceci permet de modifier localement et en cours d'analyse le comportement d'un analyseur de page.

Grâce à sa capacité de stockage d'informations variées, la mémoire permet la mise en place de plusieurs flots d'information riches entre les différents traitements. L'omniprésence de l'information mémorisée au sein d'un traitement, l'homogénéité de sa représentation et de sa manipulation avec les données issues de l'image, permettent une intégration immédiate de l'information contextuelle dans les traitements de pages. La séparation claire entre les traitements au niveau des pages, qui se concentrent sur l'extraction d'information, et le processus global, chargé de router l'information entre les pages, permet une clarification des descriptions, et la mise en place de plusieurs stratégies de circulation de l'information pour les mêmes traitements au niveau page.

Afin de valider cette mémoire visuelle, nous l'avons implémentée de manière simple dans une méthode de reconnaissance existante, DMOS-P. Ce mécanisme permet de répondre à plusieurs contraintes posées par le cadre d'une analyse grammaticale. Il permet de manipuler des informations de types variés (dimensions, indices de positionnement, règles du second ordre, etc.) de manière homogène. De plus, la manipulation des données est réalisée de manière transparente puisque les informations provenant du contexte sont considérées de la même manière que les terminaux de la grammaire qui sont extraits dans l'image en cours d'analyse.

Les exemples présentés montrent la grande souplesse de la mémoire visuelle qui a pour vocation d'être intégrée dans les systèmes de reconnaissance actuels pour permettre de nouveaux mécanismes de coopération contextuelle. Enfin, nous avons validé l'intérêt d'utiliser l'apport du contexte, puisque nous avons montré sur plus de 39 000 pages de documents d'archives abimés que l'utilisation du contexte permet d'augmenter le taux de reconnaissance de la structure de 62% à 98%.

## Remerciements

Ce travail a été réalisé en coopération avec les *Archives départementales des Yvelines* en France, avec le support du *Conseil Général des Yvelines*.

## 7. Bibliographie

- Baird H. S., Ittner D. J., « Data Structures for Page Readers », in , A. L. Spitz, , A. Dengel (eds), *Document Analysis Systems*, World Scientific, Singapore, 1995.
- Chazalon J., Coüasnon B., « Using definite clause grammars to build a global system for analyzing collections of documents », *Proc. SPIE*, vol. 7534, San Jose, CA, 2010.
- Coüasnon B., « DMOS, a Generic Document Recognition Method : Application to Table Structure Analysis in a General and in a Specific Way », *IJDAR*, vol. 8(2), p. 111-122, 2006.
- He J., Downton A. C., « User-Assisted Archive Document Image Analysis for Digital Library Construction », *Proc. ICDAR*, vol. 1, p. 498, 2003.
- Klein B., Dengel A., Fordan A., *smartFIX : An Adaptive System for Document Analysis and Understanding*, vol. 2956 of *LNCS*, Springer, p. 166-186, 2004.
- Lemaitre A., Camillerapp J., Coüasnon B., « Interest of perceptive vision for document structure analysis », *Proc. SPIE*, vol. 7527, San Jose, CA, 2010.
- Lin C. C., Niwa Y., Narita S., « Logical structure analysis of book document images using contents information », *Proc. ICDAR*, vol. 2, p. 1048-1054, Aug, 1997.