
Adaptation des caractéristiques pseudo-Haar pour le word spotting dans les documents manuscrits

Adam Ghorbel*** — Jean-Marc Ogier** — Nicole Vincent *

* *Laboratoire SIP-LIPADE, Université Sorbonne Paris Descartes*
12, rue de l'école de médecine, 75006 Paris
adamghorbel@hotmail.com
nicole.vincent@mi-parisdescartes.fr

** *Laboratoire L3I, Université La Rochelle*
Pôle Science et Technologie, avenue Michel Crépeau, 17042, La Rochelle
Jean-marc.ogier@univ-lr.fr

RÉSUMÉ. On présente dans ce papier une approche analytique multi-échelle pour le word spotting dans les documents manuscrits. Le modèle proposé comporte deux niveaux. D'abord, un module de filtrage global permettant de définir des zones candidates répondant à la requête dans le document testé. Ensuite, l'échelle de l'observation est modifiée à un niveau plus fin afin d'affiner les résultats et de sélectionner uniquement ceux qui sont vraiment pertinents. Le présent article se focalise sur la première étape de la méthode. Les contributions de cet article sont l'utilisation et l'adaptation des caractéristiques pseudo-Haar dans le word spotting. De plus, une technique permettant de modéliser les requêtes sélectionnées par l'utilisateur à partir d'un clavier est proposée. Notre méthode ne nécessite pas la segmentation préalable du document en ligne et mots. L'approche est évaluée sur la base de données publique George Washington.

ABSTRACT. This paper addresses the problem of word spotting in handwritten documents. We propose a coarse-to-fine segmentation free approach. This approach is based on two filtering phases, which are a global filtering followed by a local filtering after changing the observation scale. The contribution of this work is the use and the adaptation of the Haar-like-features in word spotting task for each tested document and the introduction of a new technique permits modelling queries typed by the user. The approach is evaluated using the George Washington manuscripts database.

MOTS-CLÉS: Word spotting, caractéristiques pseudo-Haar, analyse de documents

KEYWORDS: Word spotting, Haar-like features, document analysis

1. Introduction

Dans les documents, l'extraction de l'information est habituellement effectuée par des méthodes d'analyse de texte. La performance des systèmes optiques de reconnaissance des caractères désignés aussi par l'abréviation (OCR) est encore trop faible pour convertir certaines images de documents en texte, en particulier pour la reconnaissance des manuscrits. L'OCR ne présente donc pas une solution adaptée au problème de la recherche d'information dans des bases d'images de documents en raison de ses limites dans le traitement des écritures manuscrites et dans le traitement des collections dégradées anciennes. Plus précisément, les techniques d'OCR ne sont pas assez robustes pour les écritures manuscrites dans un contexte de vocabulaire ouvert. Pour cela, le word spotting est considéré comme une alternative à l'OCR traditionnel pour différentes applications comme par exemple l'indexation de documents et l'extraction d'information dans des collections de documents numériques.

Le word spotting a pour objectif de trouver dans les images de documents les multiples occurrences d'une requête, un mot désigné par l'utilisateur. Le word spotting, sans s'attacher à retrouver les lettres des mots, utilise des méthodes d'appariement, des méthodes de mesures de similarité plus globales entre images de mots. Elles permettent ainsi de créer des indexes partiels pour le document manipulé. Donc, le word spotting facilite l'indexation et la récupération de l'information suggérée comme une requête dans des documents historiques ou modernes quand ils sont relativement complexes et dégradés.

Dans la littérature, les approches de Word Spotting ont été développées, concernant divers scripts comme des scripts latins, arabes ou grecs, etc. Ces scripts diffèrent les uns des autres par la nature des alphabets, le nombre de caractères, la direction de l'écriture, la forme et la cursivité entre les lettres. Ils peuvent être soit manuscrits soit imprimés.

Les approches de Word Spotting ont été divisées en différentes catégories suivant des critères différents par les chercheurs en analyse de documents. Par exemple, elles peuvent être divisées en deux catégories principales basées sur des techniques de mise en correspondance des images, certaines sont des techniques d'appariement basées images et d'autres sont basées caractéristiques (Rothfeder *et al.*, 2003). Les premières techniques contiennent des méthodes qui calculent directement les distances entre les mots à partir des pixels de l'image tels que l'appariement par templates en utilisant la corrélation. Par contre, les dernières méthodes calculent certaines caractéristiques sur les images de mots et ensuite ce sont les objets portant ces caractéristiques qui vont être appariés.

On trouve aussi d'autres classifications dans la littérature (Lladós *et al.*, 2012). Deux principales approches de word spotting existent selon la forme de la requête. Ces deux approches sont basées soit sur des requêtes construites par une chaîne de caractères (QBS) soit sur des requêtes basées images (QBE). Les méthodes QBS

(Cao *et al.*, 2007) utilisent comme entrée des séquences de caractères. Elles exigent généralement une grande quantité de matériels d'apprentissage car les caractères sont appris *a priori* et le modèle de la requête est construit à l'exécution à partir des modèles des caractères constitutifs. Dans les méthodes basées QBE (Manmatha *et al.*, 1996), la requête d'entrée est une ou plusieurs images exemples du mot requête. Le problème est alors traité comme un problème de recherche d'images par contenu. Par conséquent, il ne nécessite pas l'apprentissage mais la collecte d'un ou plusieurs exemples du mot requête. Une autre technique populaire de catégorisation divise les méthodes de Word Spotting soit dans une catégorie basée sur la segmentation, soit dans une catégorie de méthodes sans segmentation (Gatos *et al.*, 2009).

L'article est organisé comme suit. La section 2 introduit une classification des approches de word spotting. Ensuite, la section 3 est consacrée à la description de l'approche proposée en détaillant ses différentes étapes. Une étude expérimentale de l'approche est proposée dans la section 4. Finalement, les conclusions et les perspectives sont indiquées dans la section 5.

2. Classification des approches de word spotting

Dans cette section, nous proposons une classification des approches de word spotting en deux grandes catégories:

- Les techniques d'analyse globale ou holistique : techniques sans segmentation, elles traitent une image de mot comme une unité globale.
- Les techniques de reconnaissance analytiques : techniques basées sur la segmentation d'un mot ; une image de document ou une image de mot est segmentée en unités plus petites qui peuvent être reconnues indépendamment ou après regroupement partiels.

En outre, nous classifions les méthodes au sein de chaque principale catégorie selon deux sous classes:

- Les techniques basées QBE.
- Les techniques basées QBS.

Nous allons maintenant décrire brièvement les méthodes relatives aux différentes classes et leurs sous classes.

2.1. Les techniques d'analyse holistique

Les techniques holistiques considèrent chaque image de mot comme une unité. Elles s'appuient principalement sur un processus de segmentation en mots qui doit être réalisée préalablement sur les documents dans lesquels se fait la recherche. En fait, les résultats dépendent de la qualité des mots segmentés dans les images de

documents. Nous divisons les techniques holistiques selon la façon dont la requête est formulée : QBE ou QBS.

En fait, Dans la plupart des techniques holistiques, plus précisément dans la sous-catégorie des techniques basées QBE, chaque mot est représenté selon un des trois types de modèles classés en modèles statistiques, modèles pseudo-structurels et modèles structurels. En premier lieu, les modèles statistiques (Shekhar *et al.*, 2012) (Yalniz *et al.*, 2012) représentent l'image comme un vecteur de caractéristiques à n dimensions. Ils peuvent être définis à partir de caractéristiques globales ou locales. Pour les descripteurs globaux, les caractéristiques scalaires qui sont calculées à partir de la totalité de l'image telles que la hauteur, la largeur, le rapport d'aspect peuvent être utilisées. Les descripteurs locaux sont calculés à partir de régions locales de l'image ou même à partir de primitives extraites de ces régions. Par exemple, ces caractéristiques peuvent être des points, des croisements, des points d'intérêt ou relatives à des régions, position/nombre de trous, etc. En deuxième lieu, certaines approches holistiques accumulent des informations pseudo-structurelles comme les informations de Loci dans des descripteurs afin de représenter les mots images (Fernandez *et al.*, 2011). Enfin, certaines approches décrivent l'image comme une séquence de primitives géométriques et topologiques et génèrent les relations entre elles (Wang *et al.*, 2014).

De plus, la plupart des travaux utilisant une approche holistique basée QBS propose à l'utilisateur d'intervenir après une phase automatique, on les qualifie d'approches guidées (Kesidis *et al.*, 2011) et d'approches non guidées (Aldavert *et al.*, 2013). En fait, plusieurs propositions d'occurrence du mot recherché sont présentées à l'utilisateur qui indique si la proposition convient ou non, et en fonction de ce retour, le système peut améliorer les résultats de spotting.

2.2. Les techniques d'analyse analytique

Les techniques analytiques consistent à segmenter l'image de mot ou même une image de document en unités plus petites qui seront reconnues lorsqu'elles sont isolées ou regroupées. En effet, nous dégageons deux catégories de techniques analytiques. Certaines approches basées sur la segmentation exigent que chaque mot soit segmenté en caractères. Ceci pour donner de meilleurs résultats de reconnaissance. D'autres approches utilisent une segmentation explicite des mots en unités plus petites qui sont censées être des portions de caractères qui seront reconnues après. Comme pour les approches holistiques, nous divisons ces approches en deux sous-catégories en fonction de la formulation de la requête.

Ainsi, Nous classifions les approches analytiques basées QBE en deux sous-classes qui sont les techniques basées sur la segmentation en caractères (Cao *et al.*, 2007) et les techniques basées sur le non segmentation en caractères (Almazán *et al.*, 2012). Dans la catégorie basée sur la non segmentation en caractères, la plupart des techniques utilisent des fenêtres glissantes afin d'en extraire différentes caractéristiques, par exemple dans (T. Mondal, N. Ragot, J.Y. Ramel, U. Pal 2015)

les auteurs utilisent les techniques de matching de type DTW. La plupart des travaux qui s'appuient sur la segmentation des caractères manipulent des documents en chinois ou japonais car la segmentation de l'image de document en caractères est plus facile que dans les autres alphabets.

Par ailleurs, les approches QBS utilisent les techniques de matching DTW, les modèles de HMM et les kNNs pour résoudre le problème de spotting. Ainsi, nous pouvons identifier deux sous-classes qui sont les techniques basées sur un apprentissage avec un modèle (Fischer *et al.*, 2012) et les techniques basées sur un apprentissage sans modèle (Liang *et al.*, 2012).

Finalement, à partir de cette étude, nous remarquons que les approches analytiques sont très robustes et donnent de meilleurs résultats que les approches holistiques. De plus, fort de l'analyse des travaux précédents, le premier choix que nous avons fait est d'exprimer la requête par la suite de caractères du mot, le plus simple étant une saisie au clavier. En effet, cela permet d'utiliser le système dans toutes les circonstances, même si le mot requête n'est pas présent dans le document. Dans ce cas, on propose une nouvelle technique qui permet de générer une séquence de formes rectangulaires pour chaque requête entrées par l'utilisateur. Cette technique permet, en première approximation, la manipulation des formes rectangulaires au lieu des codes ASCII des requêtes dans le processus de word spotting. Par contre, sans la connaissance des mots du texte, les caractéristiques de l'écriture dans le document ne sont pas connues, parmi ces caractéristiques notons la hauteur du corps des mots. De ce fait, on propose une méthode qui permet de calculer automatiquement la hauteur approximative des caractères dans chaque document manipulé.

De plus notre méthode ne repose pas sur un ensemble de caractéristiques fixé *a priori* et qui indexe les documents dans lesquels se fait la recherche mais sur une famille de caractéristiques qui vont s'adapter simultanément au mot recherché et au document dans lequel se fait la recherche.

Un autre aspect que nous avons voulu éviter est la présence d'une phase d'apprentissage sur une base de documents, ce qui spécialise le système en un système dédié aux documents anciens, aux documents imprimés ou aux documents en anglais par exemple. Il faut pour cela concevoir un système capable d'adaptation et d'auto apprentissage.

3. Une approche multi-échelle pour le word spotting

Dans notre travail, on se concentre sur les documents manuscrits, mais il s'appliquerait d'autant mieux sur des documents imprimés. Ces documents peuvent représenter une grande variabilité à différents points de vue. Les documents anciens se caractérisent par des variabilités de styles non utilisés de nos jours. Ils se caractérisent aussi par des présentations et des écritures très variées. Les difficultés

principales de ce type de documents qu'ils soient anciens ou modernes sont la fragmentation des caractères due souvent à la non homogénéité de l'encre, la variabilité du style de l'écriture et surtout le chevauchement de composantes comme par exemple des composantes appartenant à plusieurs lignes de texte du fait de la présence de hampes et jambages.

A cause de cette variabilité, les systèmes de reconnaissance ne sont pas encore opérationnels. Le word spotting permet sans déchiffrer un texte de retrouver un mot cherché. Dans la méthode que nous proposons, les documents sont traités d'une façon globale, c'est-à-dire sans avoir recours à une phase de segmentation, des lignes ou des mots, phase qui est souvent nécessaire dans les méthodes actuelles de word spotting. La manipulation de l'intégralité du document nécessite souvent un temps de calcul très grand. En tenant compte de toutes ces contraintes, nous proposons une approche comportant 3 grandes phases:

- La représentation de la requête par un ensemble de rectangles.
- Le filtrage global : permettant de réduire très fortement l'espace de recherche en sélectionnant des zones de mots candidats.
- L'affinage : permettant de ne retenir que les mots identiques au mot recherché.

Ces deux dernières grandes phases s'appuient sur des filtrages conçus à des échelles différentes. La figure 1 montre le diagramme de notre approche.

Comme nous avons opté pour une approche de word spotting dont la requête est constituée d'une suite de caractères saisis au clavier par l'utilisateur, contrairement aux méthodes générant une image synthétisée de la requête, nous allons maintenant décrire l'étape permettant une représentation automatisée de chaque requête.

3.1. Codage rectangulaire pour la représentation des requêtes

Le principe de la perception humaine nous a aidé à déduire une technique automatique de codage des chaînes de caractères. En effet, en regardant un mot par exemple, on peut le caractériser par une forme englobante qui est principalement construite par une suite de formes rectangulaires. Ces rectangles dépendent des caractéristiques de chaque caractère. La figure 2 montre un exemple de représentation d'une chaîne de caractères par des rectangles adjacents. On considère ici une vision grossière du mot qui a pour but d'éliminer un grand nombre de zones de document ne pouvant pas correspondre au mot recherché.

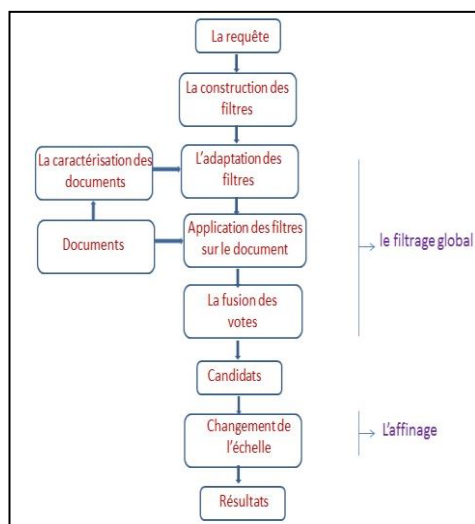


Figure 1. Le diagramme de l'approche proposée



Figure 2. Modélisation d'une chaîne de caractères par des formes rectangulaires.
 (a) La requête (b) La suite de rectangles modélisant la requête.

Cette technique est alors capable de modéliser grossièrement chaque requête entrée par l'utilisateur par une suite de rectangles dont les tailles dépendent de la taille de chaque élément de la chaîne de caractères. Cette technique de codage n'est pas réversible, c'est-à-dire, qu'à une chaîne de caractères on peut associer la forme englobante, mais la connaissance d'une suite de rectangles ne permet pas de spécifier un unique mot. Par exemple, la suite de rectangles de la figure 2(b) correspond aussi aux autres mots comme Paring, etc.

De ce fait, l'idée est de construire un tableau d'index, contenant d'une part les symboles (lettres de l'alphabet) et d'autre part 2 informations. L'une correspond au rectangle dont la taille est liée à chaque classe de lettres et un nombre index qui représente la proportion entre la hauteur du rectangle souhaité et la hauteur d'une lettre minuscule sans hampe et ni jambage se trouvant dans le document manipulé. Pour la création de ce tableau, on va classifier les symboles écrits en différentes

classes selon leurs caractéristiques (lettre avec hampe, lettre avec jambage, nombre, ponctuation, etc.). Ce tableau a 5 classes et un rectangle associé pour chaque classe. Chaque classe contient soit : les lettres majuscules, les lettres minuscules, les nombres, les symboles de ponctuation et l'espace blanc. En outre, la classe contenant les lettres minuscules peut être aussi divisée en 4 sous-classes qui contiennent les lettres avec hampe, les lettres avec jambage, les lettres avec hampes et jambages et le reste de l'alphabet. En conclusion, ce tableau est composé par :

- Classe 1 : lettres minuscules
 - Sous classe 1 : lettres avec hampe : 2
 - Sous classe 2 : lettres avec jambage : -2
 - Sous classe 3 : lettres avec hampe et jambage : 4
 - Sous classe 4 : le reste de l'alphabet : 1
- Classe 2 : lettres majuscules : 2
- Classe 3 : les nombres : 2
- Classe 4 : les symboles : 1
- Classe 5 : l'espace blanc : 0

La figure 3 montre un exemple de cette technique de codage.

3.2. Le filtrage global

La technique de filtrage global proposée représente la phase majeure de notre méthode. Elle permet d'avoir le moins possible de candidats à l'issue de la première phase de word spotting. A partir d'une requête constituée de caractères tapés au clavier, le filtrage global permet de réduire le nombre de zones possibles qui correspondent plus ou moins à la requête. Le principe de cette partie s'appuie en premier lieu sur un ensemble de points de vue. Sur chaque page de document testée, et en fonction du mot recherché, nous définissons un certain nombre de filtres caractérisant la forme du mot. Pour adapter ces points de vue au document testé, nous tenons compte des caractéristiques de ce dernier. Les caractéristiques du document sont la taille (largeur et hauteur) des caractères. Elle joue un rôle important dans la construction des points de vue et aussi dans l'indexation du tableau de codage. A ce niveau, nous allons maintenant justifier le choix des points de vue et décrire une technique permettant l'estimation automatique de la hauteur moyenne des différents caractères se trouvant dans le document manipulé.

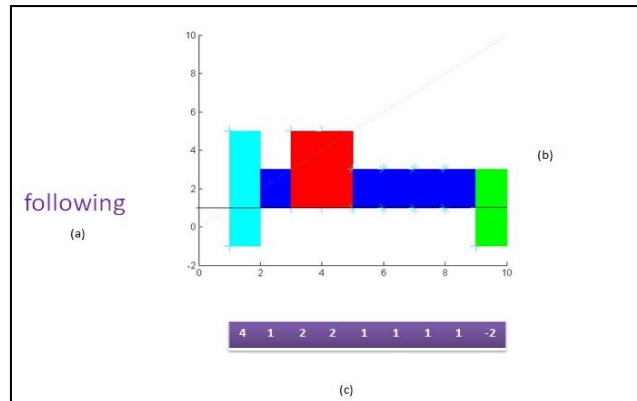


Figure 3. La technique de codage (a) La requête (b) La suite de rectangles qui modélise la forme de la requête. Les rectangles en bleu ciel représentent les lettres avec hampe et jambage, les rectangles bleus représentent les lettres sans hampe ni jambage, les rectangles rouges représentent des lettres avec hampes et les rectangles verts représentent les lettre avec jambage (c) le tableau contenant la hauteur relative de chaque rectangle.

3.2.1. Le choix des points de vue

Notre approche est une approche globale qui s'applique à chaque pixel de l'image de document. Donc, le temps de calcul représente une contrainte majeure. Ainsi, les deux étapes de construction et l'application des points de vue ne doivent pas prendre un grand temps d'exécution trop long. Les filtres choisis sont des filtres de Haar calculés en complexité constante à partir d'une image intégrale (Viola *et al.*, 2001). En effet, les filtres de Haar sont des filtres rectangulaires asymétriques qui capturent les changements d'intensité dans des zones dont nous choisissons les localisations relatives. Ainsi, on obtient des caractéristiques appelées les caractéristiques pseudo-Haar. Nous combinons ensuite les réponses aux filtres en considérant un accumulateur de votes. Nous obtenons une nouvelle image qui caractérise la présence de mots semblables à la requête dans les documents manipulés (Ghorbel *et al.*, 2015). La forme des fenêtres de Haar dépend des caractéristiques de l'écriture (Lettre minuscule, lettre majuscule, lettre avec hampe ou jambage, etc.). La figure 4 montre quelques exemples des filtres généralisés de Haar.

3.2.2. Estimation de la hauteur des minuscules dans des documents manuscrits

La taille des différents caractères minuscules dans les documents manuscrits fait partie de la caractérisation du style de l'écriture.

L'objectif, dans cette étape, n'est pas de détecter les lignes de texte mais plutôt d'estimer la hauteur des lignes prédominantes. On applique un ou plusieurs filtres simples de Haar ((e) de la figure 4) dans le document manipulé en faisant évoluer progressivement la taille du filtre. Selon la taille de chaque filtre appliqué, on obtient une réponse représentant la présence de lignes de textes.

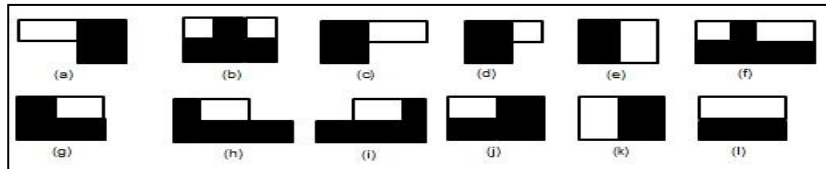


Figure 4. Quelques exemples de filtres généralisés où les rectangles noirs représentent l'écriture et les rectangles blancs représentent le fond.

La réponse est maximum quand la hauteur de la ligne est approximativement la même que celle du filtre. Ainsi, en changeant les hauteurs des filtres, on peut estimer la hauteur des lettres à partir la moyenne des hauteurs utilisées. Les résultats générés sont alors binarisés et la projection dans la direction horizontale de chaque image filtrée est calculée. A chaque échelle, la valeur médiane de la largeur des différents pics de la projection horizontale indique l'emplacement des lignes. Finalement, la valeur estimée de la hauteur des caractères minuscules ne contenant ni hampe ni jambage est calculée par l'intersection entre la courbe des valeurs médianes, fonction des tailles et la première bissectrice, points indiquant la coïncidence entre la taille du filtre et la hauteur de la ligne. La Figure 5 illustre cette technique.

3.2.3. Le principe d'accumulation des réponses aux filtres

Après l'application de ces points de vue sur l'intégralité de l'image de document, il s'agit alors de définir les zones d'intérêt du document. Le filtrage est appliqué globalement au niveau de chaque pixel de l'image du document (Ghorbel *et al.*, 2015). L'étape suivante est de fusionner les différents points de vue. Elle s'appuie sur un principe de vote. C'est un principe primordial dans notre approche. Pour chaque point de vue appliqué, un vote est exprimé à l'emplacement spatial d'une zone d'intérêt. La fusion des points de vue correspond à l'accumulation des votes. L'accumulation des hypothèses de la présence du mot nous permet de déduire la présence possible du mot dans les documents testés. L'emplacement spatial de vote associé à un pixel dépend du tableau d'index généré par le processus de codage des requêtes. Ce tableau prédit approximativement la largeur de la requête dans le document manuscrit manipulé. De plus, chaque index dans ce tableau nous aide à décider l'emplacement spatial de chaque point de vue. Egalement, le nombre et la forme généralisée des filtres de Haar qui sont appliqués sont en rapport avec ce tableau. En effet, les indices se trouvant dans ce dernier indiquent la nature de chaque filtre (un filtre pour détecter une lettre avec hampe au début du mot, une lettre avec hampe suivie par une suite de lettres minuscules sans hampes et jambages, etc.). Ainsi, ce tableau d'index joue un rôle important dans l'adaptation et la généralisation des filtres de Haar avec la requête recherchée dans le document manuscrit. La figure 6 illustre un exemple du processus de positionnement et d'accumulation de votes.

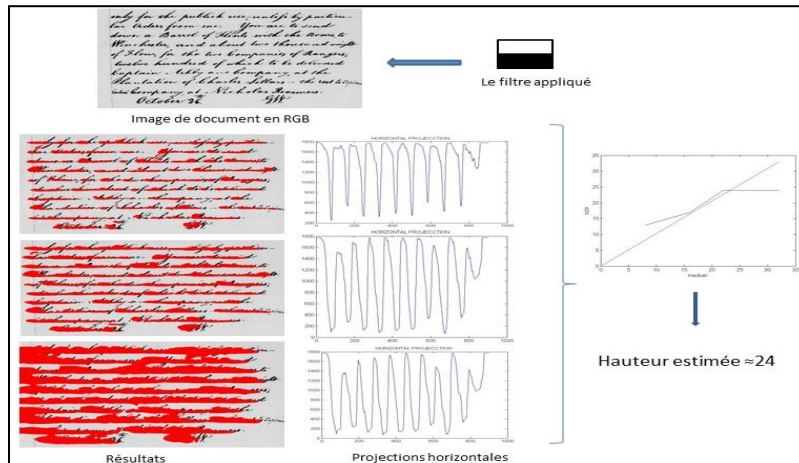


Figure 5. Le processus automatique de l'estimation de la hauteur des lettres manuscrites.

Après l'accumulation des votes, on obtient une image (Iaccum) en niveaux de gris qui met en évidence la présence possible de la requête dans le document manipulé. Ensuite, une étape de binarisation est obligatoire pour visualiser les résultats obtenus. Cette étape dépend fortement du seuil de binarisation. Ce seuil ne peut pas être fixé à l'avance à cause de la variabilité des styles d'écriture. Il faut que ce seuil soit déterminé automatiquement. Nous proposons une technique qui, à partir de l'image Iaccum, calcule tous les maximums locaux de l'image Iaccum, présentant les réponses générées par l'application des filtres généralisés de Haar à chaque pixel. Ensuite, le seuil est défini comme la valeur moyenne des maximums locaux. Une illustration de ce processus est proposée dans la figure 7. La zone d'intérêt pouvant contenir le mot recherché est déterminée en fonction de la position du maximum retenu et du rectangle englobant le modèle de mot.

3.3. Le filtrage local

Suite aux différentes étapes du filtrage global, nous déduisons la présence possible de plusieurs candidats qui ressemblent à la requête. Pour réduire le nombre de candidats, nous proposons une deuxième étape qui constitue une phase d'affinage. Nous faisons l'étude des candidats à une échelle différente. En fait, nous appliquons les points de vue sur les zones d'intérêts mais en changeant progressivement leurs tailles.

Ceci permet de bien affiner les candidats sélectionnés par le filtrage global et permet ainsi de ne retenir que les zones identiques au mot recherché. Finalement, on obtient des candidats qui correspondent le plus possible à la requête.

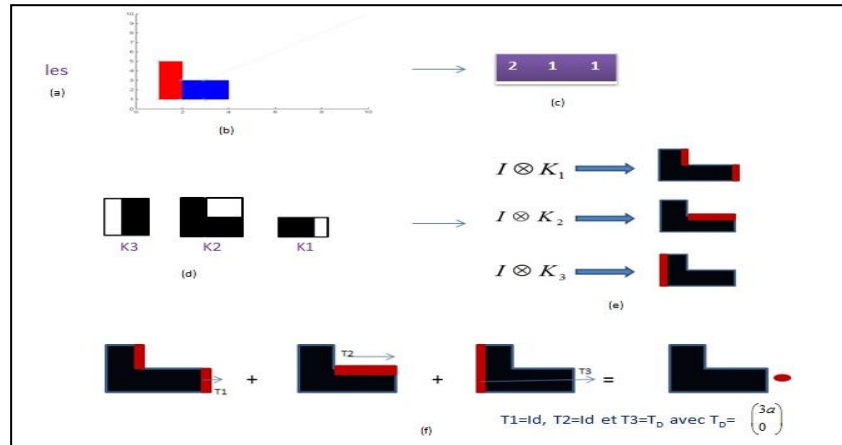


Figure 6. Un exemple illustrant le processus de translation et d'accumulation de votes. (a) La requête (b) (c) le codage de la requête (d) les filtres généralisés appliqués (e) le produit de convolution de l'image manipulée I avec le noyau K de chaque filtre. (f) les vecteurs de translation de différents votes pour réaliser l'accumulation.

4. Les expérimentations

Dans la partie expérimentale, nous avons évalué notre approche analytique de word spotting sur la base de données publique George Washington (GW) qui contient des documents du 18^{ème} siècle (Rath *et al.*, 2007) (Lavrenko *et al.*, 2004). Cette base présente les caractéristiques suivantes: 20 pages écrites en anglais par 2 écrivains, 656 lignes de textes, 4894 instances de mots, 1471 classes de mots et 82 lettres. En outre, les lettres manuscrites dans cette base ne possèdent pas la même taille vue la variabilité des styles de l'écriture et les lignes sont de différentes épaisseurs. Nos résultats issus de la phase globale sont comparés avec ceux obtenus dans les travaux de word spotting présentés dans (Rath *et al.*, 2007) et (Leydier *et al.*, 2007). Nous avons utilisé les mêmes requêtes que celles utilisées dans ces deux travaux (voir figure 8). Néanmoins, notre approche est basée QBS, donc nous n'utilisons pas ces requêtes en tant qu'images mais nous les saisissons à partir du clavier.

Dans cette partie expérimentale, aucun processus de segmentation de document ni en lignes, ni en mots et ni en caractères n'est utilisé. Ainsi, nous appliquons notre approche globalement sur les pages des documents manipulés. De plus, aucun processus d'apprentissage n'est engagé pour réaliser l'appariement entre la requête et les différents candidats.

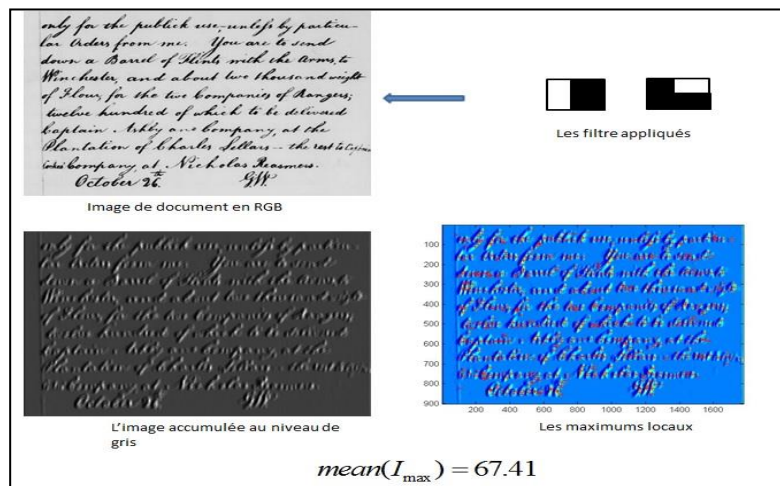


Figure 7. Le processus automatique du calcul de seuil de binarisation.

En outre, nous devons mentionner que nous n'avons supprimé aucun mot ou même filtré les ponctuations, comme cela se fait dans quelques travaux de la littérature.

L'évaluation des résultats est faite par le taux de rappel, le taux de précision et le taux de F-mesure. Pour le taux de rappel, nous obtenons une valeur de 96.6% alors que (Leydier *et al.*, 2007) (Rath *et al.*, 2007) indiquent des taux de rappel de 74.2% et 90.72% respectivement.

Les taux de précision n'étaient pas mentionnés dans ces deux articles. Pour notre part, à l'issue de la première phase de la méthode, nous générons un taux de précision de 39% et un taux de F-mesure de 55%. Les courbes de la figure 9 montrent la variation du rappel et la variation de la précision en fonction de chaque requête. Ces taux sont raisonnables pour affiner la détection du mot dans la seconde phase.

On peut remarquer que le taux de précision n'est pas bon pour certaines requêtes. Cela s'explique par le grand nombre des faux positifs qui sont détectés par notre approche. Ces faux positifs correspondent à des candidats qui ont la même forme ou

la même taille que la requête fournie en entrée (figure 10). En outre, la qualité et l'épaisseur de trait de chaque caractère entrent en jeu et affectent les résultats générés par les filtres généralisés de Haar.

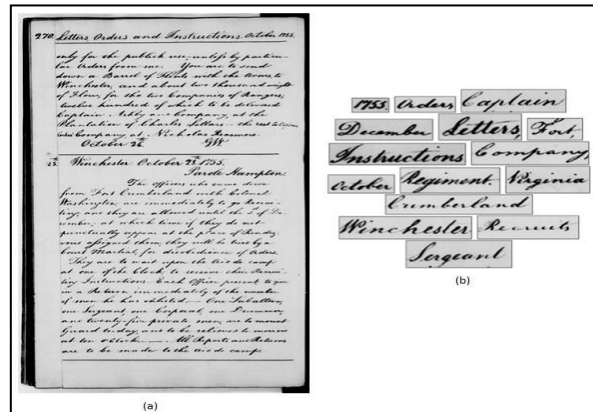


Figure 8. Une image de document de la base GW. (b) les requêtes.

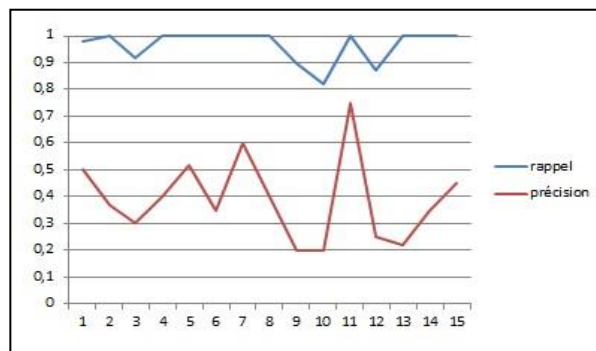


Figure 9. Les différentes valeurs du rappel et de la précision pour les requêtes utilisées.

Par ailleurs, l'approche que nous proposons, en s'adaptant au contenu des documents, est capable de détecter les occurrences des mots requête, quel que soit leur style. En revanche, (Leydier *et al.*, 2007) récupèrent seulement les occurrences qui ont le même style typographique que la requête. En outre, en terme de longueur de la requête, nous sommes capable d'obtenir de bons taux contrairement à (Rusiñol *et al.*, 2011) quelle que soit la longueur de la requête.

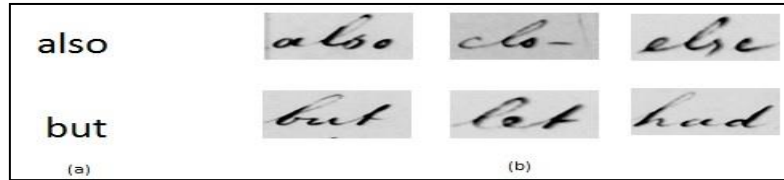


Figure 10. Un exemple montrant les différents candidats possibles obtenus pour une requête donnée. (a) 2 requêtes différentes aux requêtes d'expérimentations. (b) Les candidats possibles.

5. Conclusion

Dans cet article, nous avons proposé une approche analytique multi-échelle pour le word spotting dans les documents manuscrits. Le modèle proposé comporte deux niveaux de traitement successifs. Un module de filtrage global permettant de définir plusieurs zones candidates de la requête dans le document testé. Ensuite, l'échelle d'observation est modifiée à un niveau inférieur afin d'affiner les résultats et de sélectionner uniquement ceux qui sont pertinents. Vu que notre approche manipule des requêtes au format ASCII, nous avons proposé une technique automatique de codage permettant de représenter chaque requête par des formes rectangulaires et des indexes. Cette approche de word spotting est basée sur des familles généralisées de filtres de Haar qui s'adaptent à chaque requête pour procéder au processus de spotting et aussi sur un principe de vote qui permet de choisir l'emplacement spatial où les réponses générées par les filtres sont accumulées. De plus, une technique automatique d'estimation de la hauteur moyenne des caractères écrits dans chaque document manuscrit est décrite assurant l'adaptation aux caractéristiques des documents. Finalement, les résultats menés sur la base de donnée GW montrent que le taux moyen de rappel est très bon, alors que le taux moyen de précision devrait être amélioré.

5. Bibliographie

- Fischer. A , Keller. A, Frinken. V, Bunke. H. 2012. "Lexicon-Free Handwritten Word Spotting Using Character HMMs." *Pattern Recognition Letters Volume 33 Issue 7*, 934-42.
- Ghorbel. A, Ogier. J. M., Vincent. N. 2015. "A Segmentation Free Word Spotting for Handwritten Documents." *ICDAR 2015*, August.
- Kesidis. A. L, Galiotou. E, Gatos. B and Pratikakis. I. 2011. "A Word Spotting Framework for Historical Machine-Printed Documents." *Int. J. Doc. Anal. Recogn.*

- Gatos. B and Pratikakis. I. 2009. "Segmentation-Free Word Spotting in Historical Printed Documents." *ICDAR*.
- Aldavert. D, Rusiñol. M, Toledo. R and Lladós. J. 2013. "Integrating Visual and Textual Cues for Query-by-String Word Spotting." *In Proceedings of the Twelfth International Conference on Document Analysis and Recognition, ICDAR.*, 511–15.
- Fernandez. D, Lladós. J, Fornés. A. 2011. "Handwritten Word Spotting in Old Manuscript Images Using a Pseudo-Structural Descriptor Organized in a Hash Structure." *Iberian Conference on Pattern Recognition and Image Analysis.*, 628–35.
- Cao. H and Govindaraju. V. 2007. "Template-Free Word Spotting in LowQuality Manuscripts." *In 6th Int'l Conf. on Advances in Pattern Recognition*.
- Yalniz. I.Z, Manmatha. R. 2012. "An Efficient Framework for Searching Text in Noisy Document Images." *Document Analysis Systems, IEEE*, 48–52.
- Almazán. J, Gordo. A, Fornés. A, Valveny. E. 2012. "Efficient Exemplar Word Spotting." *BMVC*.
- Lladós. J, Rusiñol. M, Fornés. A, Fernández. D and Dutta.A . 2012. "On the Influence of Word Representation for Handwritten Word Spotting in Historical Documents." *International Journal of Pattern Recognition and Artificial Intelligence*.
- Rothfeder. J.L, Feng. S, Rath. T.M. 2003. "Using Corner Feature Correspondences to Rank Word Images by Similarity." *Conference on Computer Vision and Pattern Recognition Workshop*.
- Rusiñol. M, Aldavert. D, Toledo. R, and Lladós. J. 2011. "Browsing Heterogeneous Document Collections by a Segmentation-Free Word Spotting Method." *ICDAR*.
- Viola. P, Jones. M. 2001. "Robust Real-Time Object Detection." *2nd Int. Wkshp on Stat. & Comp. Theory of Vision*.
- Wang. P, Eglin. V, Garcia. C, Llargeron. C, Liadó. J, Fornés. A. 2014. "A Coarse-to-Fine Word Spotting Approach for Historical Handwritten Document Based on Graph Embedding and Graph Edit Distance." *In International Conference on Pattern Recognition*, 3074–79.
- Manmatha. R, Han. C, Riseman. E.M. 1996. "Word Spotting: A New Approach to Indexing Handwriting." *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 631–37.
- Shekhar. R, Jawahar. C.V. 2012. "Word Image Retrieval Using Bag of Visual Words." *International Workshop on Document Analysis Systems (DAS), 10th IAPR*, 297–301.
- Mondal. T, Ragot. N, Ramel. J.Y, Pal. U. 2015. "Performance Evaluation of DTW and Its Variants for Word Spotting in Degraded Documents." *13th International Conference on Document Analysis and Recognition (ICDAR 2015)*.
- Rath. T, Manmatha. R. 2007. "Word Spotting for Historical Documents." *Int. J. Doc. Anal. Recogn.* 9, 139–52.
- Lavrenko. V, Rath. T, Manmatha. R. 2004. "Holistic Word Recognition for Handwritten Historical Documents." *In: Proceedings Document Image Analysis for Libraries (DIAL'04)*, 278–87.
- Leydier. Y, LeBourgeois. F, Emptoz. H. 2007. "Text Search for Medieval Manuscript Images." *Pattern Recognition*.
- Liang. Y, Fairhurst. M.C, Guest. R.M. 2012. "A Synthesised Word Approach to Word Retrieval in Handwritten Documents." *Pattern Recognition Journal*.