

Le projet ANITA : ANalyse d'Images et de Tampons Anciens

Frédéric NICOLIER, Gilles MILLON, Étienne BAUDRIER

LAM - URCA, IUT de Troyes
9 rue de Québec
10 026 Troyes Cedex
33 (0)3 25 42 71 01
E-mail : f.nicolier@iut-troyes.univ-reims.fr

Résumé : Le projet ANITA du LAM-URCA (EA 2075) consiste à exploiter le fond de documents anciens de la Médiathèque de l'Agglomération Troyenne. Ce projet comporte plusieurs volets :

- impression virtuelle de tampons
- reconnaissance d'impressions issues du même tampon
- reconnaissance d'impressions issues de tampons différents mais qui représentent la même scène.
- duplication de tampons

La nature de la collaboration entre des chercheurs en traitement d'image, en représentation des connaissances, ceux en histoire du Livre et la Médiathèque de l'Agglomération Troyenne est détaillée. Les résultats actuellement obtenus et les méthodes envisagées sont présentés.

Mots clés : Reconnaissance de formes, appariement d'images, 3D.

I. CONTEXTE

Des banques de données numérisées concernant les illustrations de livres et documents anciens sont actuellement en cours de constitution. Elles nécessitent des outils informatiques permettant de les exploiter efficacement. Des problèmes de traitement d'image spécifiques à ces bases de données sont présentés avec les solutions envisagées.

Nous présentons ici un ensemble de travaux qui recouvrent des aspects scientifiques, économiques et littéraires dans les domaines du traitement d'images et de l'étude des livres et documents anciens. Ils intéressent donc les chercheurs du CEPLECA (Centre d'Étude du Patrimoine Linguistique et Ethnologique de Champagne-Ardenne), ceux du groupe de traitement d'Image du LAM (Laboratoire d'Automatique et de Micro-électronique) mais aussi la Médiathèque de l'Agglomération Troyenne (MAT). Il s'agit de travaux essentiellement pluridisciplinaires [7].

Un programme de numérisation de documents de la MAT (miniatures, livrets de colportage de la Bibliothèque Bleue, microfilm de manuscrits médiévaux) est en cours depuis 1998 et a pour but de mieux conserver les originaux et de fournir aux chercheurs des outils pour leurs travaux. Dans

cette optique, il est intéressant de chercher à profiter de ce programme pour développer des outils d'aide à la comparaison, à l'archivage, à la recherche et à l'indexation de ces documents..

L'identification des illustrations anciennes est un outil important des recherches dans le domaine de l'Histoire du livre [6, 8]. En effet, les tampons en bois qui ont servi à imprimer ces images ont pu circuler, être refaits ou bien être copiés. On dispose là de tout un matériel permettant d'attribuer des impressions anonymes à leur producteur effectif et d'explicitier des relations entre ateliers typographiques. Les opérations de numérisation des fonds de livres anciens actuellement en cours fournissent des bases de données d'images. Un des aspects du programme de la BMVR de Troyes porte sur la numérisation des livrets de colportage de la Bibliothèque bleue et de l'iconographie troyenne.

Les problèmes à résoudre sont multiples. Ils vont de la numérisation de tampons en bois gravés pour constituer des banques de données jusqu'à la mise en oeuvre d'outils logiciels de traitement d'image pour une exploitation efficace de ces banques d'images.

II. IMPRESSIONS VIRTUELLES

Les bibliothèques et les musées conservent des tampons en bois gravés pour lesquels on ne dispose pas, a priori, de l'image qu'ils reproduisent. En effet, bien souvent, le tampon conservé par une bibliothèque a produit des images sur des documents conservés par d'autres bibliothèques. Le problème est qu'aucune d'entre elles n'a connaissance de l'existence du tampon ou de son image dans les autres. Il faut donc développer un ensemble d'outils matériels et logiciels qui formerait un banc de reconstitution des images issues des tampons en bois gravés. En d'autres termes, il s'agit d'extraire l'image produite par un tampon à partir du bois gravé lui-même. Cela contribuerait à la constitution de banques de données permettant la mise en correspondance automatique des illustrations et des tampons quels que soient leur origine et/ou leur lieu de conservation actuel. La figure 1 représente un de ces tampons en bois.

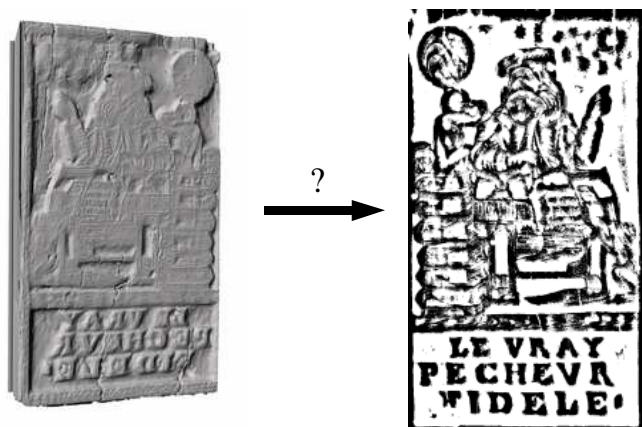


Fig.1. Exemple d'une impression virtuelle.

Les moyens matériels et logiciels envisagés pour constituer ce banc d'acquisition dépendent bien sûr du problème de traitement d'images mais aussi et surtout des contraintes imposées par les bibliothèques. Les conditions de conservation (température, humidité...) et de manipulation sont strictes. Il est, par exemple, exclu de ré-encre les tampons pour les réutiliser sous des presses typographiques. Les tampons en bois ont été fragilisés au cours du temps (fentes, bois vermoulu, dégradations par l'humidité ou l'impression...) au point de devoir les manipuler avec beaucoup de précautions. De même, on ne peut pas envisager d'enduire le bois d'un quelconque liquide ou de saupoudrer la surface gravée de talc, par exemple, pour faciliter le processus d'extraction de l'image produite par le tampon. La technique d'extraction de l'image produite doit donc rester très respectueuse de l'état du bois.

Des travaux préliminaires ont montré que l'utilisation d'une image de luminance ne permet pas d'obtenir une information exploitable. La technique finalement retenue est donc de partir d'une image de profondeur représentative de la géométrie du tampon [1]. Un traitement simple (seuillage local) permet alors d'obtenir une image imprimable. La figure 1 contient un exemple de résultat obtenu. Un paramètre k variable permet de fixer le taux « d'encrage » de l'impression virtuelle. Il reste à valider la méthode à grande échelle et à comparer des impressions réelles aux impressions virtuelles. Pour cette application nous collaborons avec le Le2i (IUT du Creusot) en ce qui concerne l'acquisition et les traitements 3D [9]. L'objectif est de développer une plateforme complète à coût réduit. Il s'agit donc d'utiliser des méthodes de stéréovision pour développer une acquisition 3D. La résolution de l'acquisition de la plateforme sera réduite. Le but final étant (pour le moment) essentiellement une analyse visuelle, une résolution réduite devrait être suffisante.

III. RECONNAISSANCE D'IMPRESSIONS ISSUES DU MÊME TAMPON

Pour le second problème, il s'agit, à partir de plusieurs tirages, ou reproductions de gravure, d'identifier ceux qui

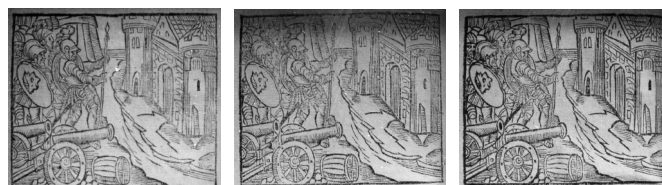


Fig.2. Les images de trois impressions d'un même tampon.

ont été obtenus à partir d'un même tampon, ou qui, au contraire proviennent de tampons copiés. Le problème est complexe dans la mesure où un même tampon ne donne pas toujours des tirages identiques (cf. figure 2) : les différences d'encrage, les phénomènes de retrait liés à l'humidité plus ou moins importante du papier support, les défauts de tirages, les « accidents » subis par le bois au cours de son existence (fentes, trous de vers, retouches...) mènent à des épreuves d'aspect parfois fort différent qu'il est souvent difficile de différencier d'épreuves obtenues à partir de tampons copiés.

Une des difficultés réside dans le fait que les images à comparer peuvent provenir de fonds numérisés différents. Dans ce cas, les conditions d'éclairage et d'acquisition ne sont pas maîtrisables par le logiciel. Il est donc nécessaire de procéder à des pré-traitements. Les opérations nécessaires sont donc :

1. Binarisation (seuillage)
2. Recalage en translation, rotation et homothétie
3. Comparaison des images recalées.

La comparaison des images recalées fera appel à des calculs de distances adaptées du type distance de Hausdorff. Il s'agit d'une distance qui mesure un écart entre deux courbes [5]. Ces mesures seront suivies d'une étape de classification.

IV. RECONNAISSANCE D'IMPRESSIONS REPRÉSENTANT LA MÊME SCÈNE

Le troisième problème identifié consiste à reconnaître dans une banque d'images les illustrations de la même scène. En effet, les études bibliographiques relatent quelquefois les mêmes événements. Elles ont été rédigées à diverses époques et par conséquent utilisent des tampons différents ou bien copiés, ou encore un même tampon dégradé. Une même scène peut donc être représentée par de nombreuses images différentes. Par exemple, la figure ci-dessous reprend quatre illustrations de l'histoire des quatre fils Aymon. On y reconnaît bien sûr la même scène, mais les images diffèrent par certains détails. Les tampons ont été produits par des graveurs à travers toute la France sur une période s'étalant de 1613 à 1860. L'objectif est de mettre en évidence un degré de ressemblance entre les scènes. Un expert devra valider les résultats de l'outil développé (A. Robert et/ou M.D. Leclerc du CEPLECA).



Fig.3. Les images de quatre impressions réalisées à partir de quatre tampons différents.

Le problème consiste donc à reconnaître dans une banque d'images les illustrations de la même scène. Une même scène peut donc être représentée par de nombreuses images différentes. La problématique réside dans l'association des images qui possèdent le même contenu sémantique.

Il s'agit de construire une représentation sémantique des images et de comparer des représentations pour réaliser l'appariement. Dans ce cadre, nous travaillons sur l'adaptation aux images, des techniques employées en traitement automatique du langage. En particulier, l'adaptation des graphes conceptuels à la reconnaissance de forme est une voie qui est explorée. En effet, les graphes conceptuels constituent un technique "souple" de représentation des connaissances par réseaux qui se rapproche de l'efficacité descriptive du langage naturel [2, 10]. Les graphes conceptuels sont un outil puissant et bien étudié pour le langage écrit. Il existe d'ailleurs des logiciels de manipulation de graphes déjà développés par d'autres équipes.

Nos travaux, qui constituent une recherche innovante et très peu explorée à ce jour, ont donc trois objectifs :

- Développement d'une sémantique graphique. Quels sont les graphes conceptuels qui permettent de représenter au mieux les illustrations étudiées ?
- Construction automatique du graphe conceptuel d'une illustration donnée.
- Comparaison des graphes construits et mesure de similitude des illustrations.

La principale difficulté dans ce genre de problème consiste à construire un codage ou une représentation efficace des images analysées. L'approche que nous cherchons à mettre en oeuvre en appliquant la théorie des graphes conceptuels pour construire une sémantique graphique est originale mais s'inscrit dans une thématique de recherche nationale.

La construction et la comparaison de graphes risquant d'être une opération qui explose combinatoirement, nous travaillons sur une approche multirésolution. Il s'agit de construire par itération le graphe d'une image en partant d'une version basse résolution et en ajoutant des détails successifs. Dans cette optique nous utilisons une extension non-linéaire de la théorie des ondelettes qui permet d'associer des opérateurs de morphologie mathématique aux analyse multirésolution [3, 4].

L'intérêt scientifique de ce projet est de contribuer au développement et à la mise au point de méthodes de recherche d'images dans une base de données en se basant sur le contenu. Les méthodes développées seront généralisables et permettront à terme de rechercher dans une banque les images "proches" d'une image de référence. Tout l'intérêt et l'originalité de notre projet tient dans le fait que la mesure de similarité est sémantique. Les méthodes développées pourront être exploitées dans des domaines variés (fonds documentaires anciens, moteurs de recherche d'images pour internet, indexation d'images par le contenu, et plus généralement reconnaissance de formes).

V. DUPLICATION DE TAMPONS

Une dernière application consiste à réaliser des duplicata des tampons anciens. Il s'agit de pouvoir produire une copie des tampons pour éviter la perte du tampon en cas de destruction. Cela permettra également l'envoi et/ou la vente de copie à des chercheurs éloignés ou à des visiteurs.

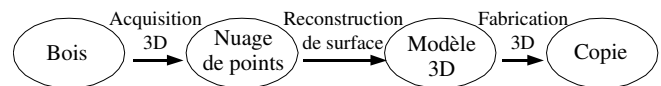


Fig.4. Schéma de principe de la duplication de tampons.

La figure 4 contient un schéma de principe de la duplication. Celle-ci se décompose en 3 étapes :

1. Acquisition 3D : obtention d'un nuage de point brut
2. Reconstruction d'une surface en vue d'obtenir un modèle 3D adapté aux contraintes de fabrication (qui sont différentes des contraintes de visualisation)
3. Fabrication 3D : plusieurs méthodes sont envisageables (dépôt de surface ou fraissage).

Ce travail est réalisé en collaboration avec le Le2i.

Bibliographie

- [1] E. BAUDRIER, G. MILLON, F. NICOLIER, « Impression virtuelle de tampons en bois gravé anciens : binarisation et recalage d'images de profondeur », colloque "Maquette virtuelle et Patrimoine", Cluny, mars 2003.
- [2] K. CHIBOUT, A. VILNAT, X. BRIFFAULT, « Sémantique du lexique verbal : un modèle en arborescence avec les graphes conceptuels », Revue Traitement Automatique des Langues (TAL), vol 42, no 3, 2002, p.691-727.

- [3] J. GOUTSIAS, H. HEIJMANS, « Nonlinear Multiresolution Signal Decomposition Schemes - Part I : Morphological Pyramids», IEEE Trans. on Image Proc., vol 9, no 11, nov. 2000, p. 1862-1876
- [4] J. GOUTSIAS, H. HEIJMANS, « Nonlinear Multiresolution Signal Decomposition Schemes - Part II : Morphological Wavelets», IEEE Trans. on Image Proc., vol 9, no 11, nov. 2000, p. 1897-1913
- [5] D. P. HUTTENLOCHER, « Comparing Images Using the Hausdorff Distance », IEEE Trans. on PAMI, vol 15, no 9, sept. 1993. p.850-863
- [6] M.-D. LECLERC, « A la recherche du livre perdu : Identification de quelques bois gravés dans la bibliothèque bleue de Troyes », in *La Bibliothèque Bleue et les littératures de colportage*, Étude réunies par T. Delcourt et E. Parinet, Études et rencontres de l' École des chartes, n°7, Paris - Troyes, École des chartes - La maison du Boulanger, 2000, p. 109 - 129
- [7] G. MILLON, F. NICOLIER, A. ROBERT, « Études d' images médiévales », Colloque national sur la recherche universitaire dans les IUT CNR' IUT2002, Le Creusot, mai 2002, p. 66-70.
- [8] A. ROBERT, « Classification et bibliographie sommaire des éditions des Quatre Fils Aymon dans la Bibliothèque Bleue », in *Entre épopée et légende : Les Quatre fils Aymon ou Renaud de Montauban*, Etudes publiées sous la direction de D. Quéruel, tome II, Langres, D. Gueniot, 2000, p. 137-147.
- [9] R. SEULIN, O. MOREL, G. MILLON, F. NICOLIER, « Range Image Binarization : Applications to Wooden Stamps Analysis », QCAV2003, mai 2003, Gatlinburg, USA.
- [10] J. SOWA, « Knowledge Reprerstation : Logical, Philosophical and Computational Foundations», Brooks/Cole Ed., Pacific Grove, USA, 2000