

# **Vers une réduction du fossé sémantique dans le traitement des images de documents anciens à base d'ontologies : Application aux lettrines**

Mickaël Coustaty\*, Norbert Tsopze\*\*, Alain Bouju\*, Karell Bertet\*, Georges Louis\*

\*Laboratoire L3i - Université de La Rochelle  
Avenue Michel Crepeau, 17042 La Rochelle, France  
{mcoustat, abouju, kbertet, glouis}@univ-lr.fr  
<http://L3i.univ-larochelle.fr>

\*\*Département d'Informatique - Université de Yaoundé I  
BP 812 Yaoundé - Cameroun  
tsopze@uy1.uninet.cm  
<http://www.une-autre-page.html>

**Résumé.** Nous présentons dans cet article une approche de réduction du fossé sémantique dans le traitement des documents anciens, et plus particulièrement les images de lettrines. Cette approche modélise les connaissances de deux domaines (experts historiens et experts en traitement d'images) au sein d'ontologies. Cette démarche permet d'établir des liens entre ces domaines afin d'aider les historiens à interpréter ces images et à les situer dans le temps. Nous avons combiné trois ontologies (thésaurus défini par les historiens, ontologie du traitement d'image et ontologie spatiale) pour l'annotation de ces images; ainsi nous avons donc défini des règles permettant d'annoter certaines régions de l'image comme étant la lettre, une partie du corps de personnage ou encore de caractériser une lettrine comme figurative. Des expérimentations prouvent la pertinence du système dans l'annotation automatique d'images. Par ailleurs, la généralité de l'approche permet d'en envisager une exploitation dans d'autres contextes combinant analyse d'image, analyse spatiale et sémantique du domaine, comme par exemple des images de bandes dessinées Tsopze et al. (2012).

## **1 Introduction**

De grandes campagnes de numérisation sont réalisées depuis près de dix ans à travers le monde entier, générant de grands volumes d'images de documents patrimoniaux. Ces documents représentent la mémoire de nos sociétés, et de nombreux historiens s'intéressent à la création et la vie de ces documents pour retracer leur histoire et celle de la France. Il devient donc nécessaire de proposer des outils adaptés pour rechercher leur contenu, et ces grands volumes imposent d'être capable de les organiser pour permettre une recherche rapide et dédiée.

Dans le cas des images de documents anciens, cette tâche devient difficile puisqu'il faut pouvoir mettre en concurrence des données textuelles avec le contenu des images (les pixels).