

# Approche complexe de l'analyse de documents anciens

Mickael Coustaty\*, Giap NGuyen\*, Vincent Courboulay\*, Jean-Marc Ogier\*

\*Laboratoire L3i - Université de La Rochelle  
Pôle Science et Technologie, Avenue Michel Crépeau, 17042 La Rochelle  
{mcoustat,vcourbou,giap.nguyen,jmogier}@univ-lr.fr

**Résumé.** Cet article présente une méthode complexe pour la caractérisation et l'indexation d'images graphiques de documents anciens. A partir d'un bref état de l'art, une méthode pour décrire ces images en tenant compte de leur complexité est proposée. Trois étapes principales de ce traitement sont détaillées dont une méthode novatrice d'analyse, de segmentation et de description des traits. Les résultats sont issus de travaux en cours et sont encourageants.

## 1 Introduction

L'indexation de documents issus du patrimoine représente un enjeu actuel pour la sauvegarde de nos mémoires. Dans cet article, nous nous intéressons précisément aux images contenues dans ces documents pour leur complexité et leurs caractéristiques particulières. Nous présentons tout d'abord quelques méthodes existantes pour l'analyse de ces images. Puis nous nous intéressons plus particulièrement dans les parties 3 et 4 à la complexité des ces documents et à la manière de décrire leurs éléments constitutants. Enfin, la dernière partie de cet article présente notre représentation de la complexité dans les documents. Pour cela nous nous appuyons sur une description des images en "*unités et échelles fonctionnelles*". La dernière unité fonctionnelle proposée est novatrice puisqu'elle propose de caractériser les images non plus en utilisant les pixels mais des primitives de plus haut niveau : des traits. Les travaux présentés dans cet article sont réalisés dans le cadre du projet ANR NaviDoMass (2010) et les premiers résultats sont encourageants.

## 2 Les documents anciens

### 2.1 Un enjeu actuel

De plus en plus de bibliothèques nationales, de projets nationaux, européens ou mondiaux (NaviDoMass (2010); Europeana; Passe-Partout; Google Livres; Open Content Alliance,...) cherchent à préserver leur patrimoine documentaire. Dans cette optique, de grandes campagnes de numérisation sont actuellement menées par ces différents projets qui cherchent à sauvegarder en masse, au format image, des copies de ces documents. Ces campagnes numérisent de grandes masses de documents et nécessitent de plus en plus souvent des services de navigation pour permettre aux usagers de retrouver les documents. Ces services impliquent la nécessité