Etude de la dynamique du ductus. Analyse et classification des écritures.

Mots clés :

- **Directeur de thèse** : nicole VINCENT
- **Co-encadrant(s)** :
- **Unité de recherche** : Laboratoire d'Informatique PAris DEscartes
- **Ecole doctorale** : École Doctorale Informatique, Télécommunications, Électronique de Paris
- **Domaine scientifique principal**: Divers

Résumé du projet de recherche (Langue 1)

Our work is part of the ANR GRAPHEM project. This is a multidisciplinary project aiming at helping paleographers in their expert work of Latin Middle Age handwritings analysis. Our main goal as computer scientists is to develop automatic (or semi automatic) processes of handwriting styles classification by the constitution of shapes code books. On such manuscripts, we face several constraints due to the bad conditions of handwriting preservation and especially of ink degradations. Furthermore, the execution rules of writing in paleography are very strict: some letters and letter combination can only be produced by a unique dynamic movement. It is necessary to take into account all these constraints and execution specificities of writings to decompose the handwritten manuscripts into coherent patterns (also called graphemes) which can be considered as small handwriting strokes by the detection of the center line (the skeleton) directly on the grayscale images while avoiding some backward pen movements. This initial decomposition leads to a first vectorial description of each handwriting. The definition of a well adapted similarity criterion to compare patterns is then realized for the classification step on each handwriting sample: it produces a robust and flexible shapes Codebook which is a representative of the graphemes distribution (in shapes and frequencies) by the use of a relevant classification process based on Graph Coloring that has never been exploited in such a context. These Codebooks represent basic signatures of handwritings that are consequently exploited in different automatized tasks like styles classification, recognition and shapes spotting for content information retrieval.