

Classification et sélection automatique de caractéristiques de textures

Marine Campedel*, Eric Moulines**

Ecole Nationale Supérieure des Télécommunications
Laboratoire de Traitement du Signal et des Images
46, rue Barrault, 75013 Paris

* Marine.Campedel@enst.fr,
<http://www.tsi.enst.fr/~campedel>

** Eric.Moulines@enst.fr
<http://www.tsi.enst.fr/~moulines>

Cette étude est financée par le Centre National d'Etude Spatiale (CNES).

Résumé. Choisir a priori les caractéristiques pertinentes pour une application donnée n'est pas aisé. En particulier, les outils de classification d'images utilisent des modèles variés pour représenter les textures. Nous proposons de choisir les modèles de texture les plus pertinents à l'aide d'une procédure automatique de sélection de caractéristiques. Nous comparons pour cela l'efficacité de plusieurs algorithmes récents en évaluant les performances de différents classificateurs exploitant ces sélections. Nous démontrons l'intérêt d'une telle procédure de sélection à partir des images de Brodatz.

1 Introduction

Face à l'accroissement rapide des tailles des bases de données, en particulier des bases d'images, il est nécessaire de développer de nouveaux algorithmes de traitement facilitant à la fois le stockage et l'indexation de ces données. Nous nous intéressons dans ce travail aux algorithmes de sélection de caractéristiques (appelées aussi descripteurs), qui permettent d'extraire une information non redondante et pertinente, en vue d'une exploitation efficace des bases de données. Ces algorithmes font l'objet d'une littérature abondante depuis une dizaine d'années [Kohavi et John, 1997, Blum et Langley, 1997, Guyon et Elisseeff, 2003] .

Les algorithmes de sélection sont répartis en deux groupes principaux : les 'filters' et les 'wrappers'. Les premiers exploitent les propriétés intrinsèques des caractéristiques utilisées, sans référence à une quelconque application. Les seconds, au contraire, définissent la pertinence des caractéristiques par l'intermédiaire d'une prédiction de la performance du système final. Les premiers sont généralement moins coûteux mais aussi moins efficaces que les seconds.

Les performances des algorithmes dépendent de l'évaluation de la pertinence des descripteurs mais aussi de la stratégie de recherche du sous-ensemble de caractéristiques. Cette recherche ne peut pas être exhaustive (il faudrait explorer 2^D combinaisons