

# Représentation d'images par des chaînes de symboles : application à l'indexation d'images\*

J.M. Jolion<sup>1</sup>

I. Simand<sup>1</sup>

<sup>1</sup> LIRIS, FRE CNRS 2672 INSA

Bât. J. Verne

INSA Lyon, 69621 Villeurbanne cedex - France

{Jean-Michel.Jolion, Isabelle.Simand}@liris.cnrs.fr

## Résumé

*Nous introduisons dans cet article une nouvelle représentation pour des images basée sur la notion de chaîne. De nombreux travaux en indexation d'images font référence à la notion de points d'intérêt comme base d'une représentation concise des images mais leur structure spatiale de même que leur exacte signification sont rarement prises en compte si ce n'est comme un paramètre supplémentaire. Nous proposons une représentation d'images sous la forme d'une liste triée, une chaîne, d'un ensemble de points extraits de l'image. La valuation de chaque point est extraite d'une analyse du contraste de l'image et se traduit sous une forme symbolique et non pas numérique. La comparaison de deux représentations se fait en utilisant des outils de la reconnaissance des formes structurelle et plus particulièrement une distance dérivée de la distance d'édition. Des illustrations permettent de mesurer les apports d'une telle technique.*

## Mots clefs

Indexation d'images, chaînes de symboles, analyse contraste.

## 1 Introduction

Le domaine de l'indexation d'images a connu un essor très rapide ces dernières années sans que l'on ait pu réellement voir émerger des services opérationnels. Après une phase très largement exploratoire, les travaux relevant de ce domaine sont maintenant principalement orientés vers des optimisations sur un corpus de méthodologies et techniques de plus en plus stable. Dans le cadre de notre collaboration avec France Télécom, nous avons souhaité revenir à des travaux plus exploratoires relevant de l'approche structurelle de la reconnaissance des formes.

Les motivations principales de cette étude trouvent leur source dans plusieurs travaux. Tout d'abord, les études sur la vision humaine ont mis en évidence que le parcours de

l'oeil analysant une image a peu de corrélation avec les lieux réputés les plus informatifs d'un point de vue du traicteur d'images hors mis la notion de contraste [1]. La puissance du code lié à l'information extraite par l'oeil est donc ailleurs et une conjecture actuelle situe cette puissance dans la prise en compte de la notion d'ordre dans les images [2]. Par ailleurs, grâce à la loi de Naka-Rushton, on connaît l'importance de la notion de contraste dans l'analyse visuelle.

Nos travaux ont donc pour but de traduire ces deux éléments (ordre et contraste) dans une nouvelle définition d'un code d'images pour la classification / reconnaissance / indexation d'images.

## 2 Architecture générale

Au delà du critère qui guide les points de focalisation de l'oeil, il est maintenant admis que réduire une image à un ensemble de points d'intérêt est à la fois une approche peu coûteuse mais aussi très performante (et ceci depuis les travaux de Schmid et Mohr [3]). Notre approche reprend cette analyse tout en la combinant avec les notions de contraste et d'ordre. La couleur et la multirésolution, bien que déjà mises en oeuvre dans notre système, ne seront pas exposées ici par manque de place.

La figure 1 détaille les différentes étapes du processus.

Ce processus permet de coder chaque image sous la forme d'une chaîne, ce que nous allons détailler dans la suite.

### 2.1 Binarisation d'images

Nous utilisons dans ce premier aspect une technique que nous avons proposée il y a quelques années pour le rehaussement de contraste [4]. Son principe est le suivant :

- représentation multirésolution d'une image sous la forme d'une pyramide ;
- calcul d'une pyramide de contraste par ratio de deux niveaux successifs ;
- rehaussement itératif du contraste de l'image à tous les niveaux de résolution
- reconstitution de l'image rehaussée par projection *top-down*.

\* Cette étude est financée par France Télécom R&D dans le cadre du contrat ECAV3 42568725

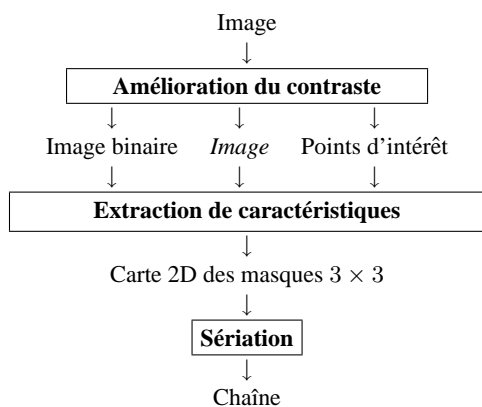


Figure 1 – Schéma général de l'extraction du code à partir d'une image.

Cette méthode itérative a un point fixe pour chaque image. Celui-ci correspond à une image qui conserve au mieux l'information de contraste tout en réduisant l'entropie de l'image. On obtient donc ainsi une image binaire. La figure 2 montre un exemple de cette convergence.



Figure 2 – Plusieurs niveaux de la pyramide initiale (à gauche) et de sa binarisation obtenue en 7 itérations.

## 2.2 Points d'intérêt

Il a été montré dans [5] que lors de la première itération de ce mécanisme, on pouvait extraire un ensemble de points d'intérêt dans l'image, définis comme étant les maxima locaux de la mesure de contraste cumulée dans la base de la pyramide.

La figure 3 montre un exemple d'extraction de points d'intérêt. Ces points servent de filtre sur l'image binaire issue de l'étape de binarisation. En effet, seuls les points d'intérêt et leur voisinage immédiat sont retenus (cf figure 4).



Figure 3 – Une image originale avec superposition des points d'intérêt.



Figure 4 – Masques binaires 3x3 correspondant aux points d'intérêt de l'image de la figure 3.

## 2.3 Sériation

Les deux étapes précédentes ont permis de résumer l'image de départ à une carte 2D de masques binaires (cf figure 4). Il est actuellement impossible de garder une telle carte comme code de l'image car la comparaison de ces structures 2D reste complexe. C'est pourquoi nous proposons de traduire cette structure 2D en une structure 1D par un processus de sériation. Il existe de très nombreuses approches de la sériation : approche spectrale de décomposition du graphe de voisinage, parcours de l'image, parcours du graphe... Nous avons, dans un premier temps, opté pour une approche simple reposant sur un ordonnancement des points en regard d'un critère scalaire. La mesure retenue est l'énergie de contraste qui est disponible sans calcul supplémentaire puisqu'elle a déjà servie à extraire les points d'intérêt. Les points d'intérêt sont donc triés selon un ordre décroissant de cette mesure ce qui nous permet d'obtenir un code avec une complexité de calcul de l'ordre de  $\mathcal{O}(n \log(n))$  où  $n$  est le nombre de points d'intérêt. Par contre, cette approche ne permet pas la prise en compte de la répartition spatiale des points. Ceci constitue une des



avec des poses différentes au sein d'une base de visages, tous acquis dans les mêmes conditions).

- Base 4 : classification de symboles binaires (des chiffres).
- Base 5 : similarité à une requête image dans une base très généraliste (images TV en Mpeg1).
- Base 6 : robustesse à la perte de généralité (robustesse à l'ajout de bruit, *i.e.* de 300 à 1500 images quelconques, dans une base de 20 classes représentant 300 images).

Par manque de place, nous ne présenterons ici que le cas de la base 1. La figure 5 montre un échantillon d'images issues de cette base.

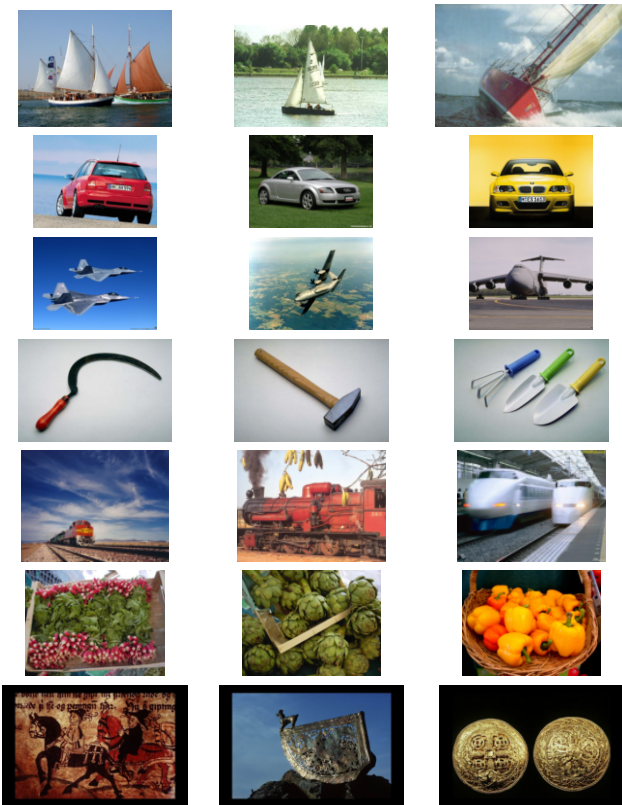


Figure 5 – Images utilisées dans la base 1. Elles sont classées en 7 catégories (en ligne) : bateaux, voitures, avions, outils, trains, légumes, objets vikings

La figure 6 présente les résultats obtenus avec notre approche. Par comparaison, une classification simple sur les histogrammes couleurs permet d'atteindre 68.5% de bonne classification. Les distances de distribution (+) et d'ordre (●) donnent des résultats comparables et de même qualité qu'une approche globale par histogramme couleur (information qui n'est pas prise en compte par notre code sur cet exemple). Par contre, grâce à un apprentissage (×), on peut obtenir une bien meilleure classification qui témoigne de la capacité de ce nouveau code à s'adapter à une configuration donnée.

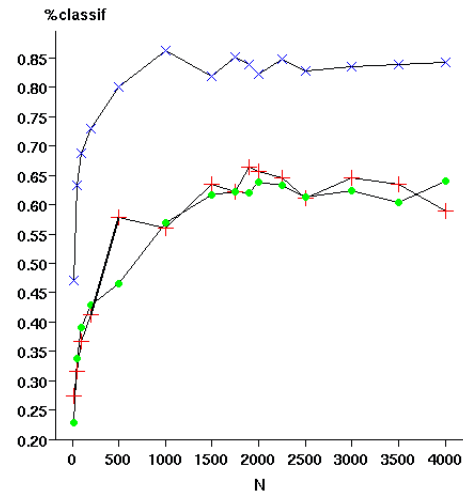


Figure 6 – Pourcentage de bonne classification en fonction de la taille des chaînes extraites et pour trois distances de chaînes : distance entre distribution de symboles (+), distance prenant l'ordre en compte (●) et avec apprentissage (×).

## 5 Conclusion

Dans le cadre de cette collaboration avec France Télécom R&D, nous avons proposé une approche structurale de la construction d'un code d'images en vue de comparaison, classification et indexation. Cette approche est nouvelle en ce sens qu'elle redonne sa pleine expression à des travaux anciens sur la reconnaissance de formes structurale, plus à même de rapprocher le codage d'une image de son contenu sémantique.

Sur ces premiers travaux encourageants, il nous reste à bâtir un code plus complet qui intègre la notion de couleurs (travaux en cours) et de multirésolution.

## Références

- [1] D.J.Parkhurst et E. Niebur. Scene content selected by active vision. *Spatial Vision*, 16(2) :125–154, 2003.
- [2] S.J. Thorpe, A. Delorme, et R. VanRullen. Spike-based strategies for rapide processing. *Neural Networks*, 14 :715–725, 2001.
- [3] C. Schmid et R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. on PAMI*, 19(5) :530–535, 1997.
- [4] J.M. Jolion. Analyse multirésolution du contraste dans les images numériques. *Traitement du Signal*, 11(3) :245–255, 1994.
- [5] S. Bres et J.M. Jolion. Detection of interest points for image indexation. Dans D.P. Huijsmans et A.W. Smeulders, éditeurs, *Visual Information and Information Systems*, pages 427–434. LNCS, 1614, Springer, 1999.