

SYSTEME DE RECHERCHE ET DE NAVIGATION DANS UNE BASE D'ARABESQUES PAR DES DESCRIPTEURS LOCAUX

Souissi Abdelmoghith (*), Tabout Hassan (*), Sbihi Abderahmane (*)
ab.souissi@menara.ma, h.tabout@caramail.com, sbihi@univ-ibntofail.ac.ma

Université Ibn Tofaïl, Faculté des Sciences / Laboratoire Image et reconnaissance des Formes
BP 133 Kénitra 14000 Morocco Tél. +212 61 67 92 12 Fax: +212 37 37 27 70

Mots clefs :

Système RIPC, navigation automatique, classification supervisé, classification non supervisé

Keywords:

CBIR system, automatic navigation, classification supervised, classification not supervised

Pala bras clave:

BIPC Sistema, Navegación automática, Clasificación supervisado, Clasificación no supervisado

Résumé

Nous présentons dans cet article, un système de recherche des arabesques par le contenu dans les bases d'images spécifiques à ce domaine, la recherche est basée sur une approche locale, à l'aide de l'indexation des régions d'intérêts identifiées dans l'image, les descripteurs de ce dernier sont basés sur des attributs de bas niveaux de la couleur et de la texture permettant de s'approcher du sens sémantique porté pour chaque région d'intérêt. Le système propose également un mécanisme de navigation sur les différents modèles d'arabesques. Pour l'évaluation d'une telle approche, nous avons utilisé un corpus des arabesques de 2000 images rassemblées de plusieurs modèles des artisans.

Abstract

In this paper, we present an arabesque images retrieval system based on their contents. The proposed technique relies on a local approach. The image is segmented into interest regions and each region is characterized by a signature based on color and texture features. The system also proposes a mechanism of navigation on various kinds of arabesque images. To evaluate our system we make use of 2000 images collected by artisans.

1- Introduction

De nos jours, le domaine de l'indexation et de la recherche des images par le contenu a connu une pleine expansion, avec l'omniprésence de l'image dans plusieurs domaines, l'explosion d'Internet et aussi le développement à grande échelle de la photographie numérique, de plus la diminution des coûts et la montée en puissance et en stockage des ordinateurs, tous ces éléments ont abouti à la création du besoin d'une telle approche, également ce n'est pas rare d'avoir des bases d'images numériques contenant des milliers d'images, que ce soit des bases spécifiques pour un domaine d'activité professionnelle (journalisme, tourisme, éducation, médecine, musées, ...) ou tout simplement pour les particuliers qui accumulent d'immenses bases de photographies numériques (souvenirs, voyages, famille, événements, ...).

Pour gérer et exploiter efficacement ces bases d'images, un système d'indexation et de recherche d'images par le contenu s'est avéré indispensable. Actuellement la majorité des systèmes disponibles sont basés sur la recherche d'images soit par le nom des images, ce qui donne en générale des résultats non pertinents, soit par l'utilisation des annotations manuelles, ce qui est fastidieux dans le cas des bases de données gigantesque.

Il existe deux niveaux d'interprétation d'une image : symbolique, ou bas niveau, et sémantique, ou haut niveau. L'information portée par les structures d'interprétation symbolique n'est pas assez riche pour pouvoir être utilisée individuellement. Un pixel ne contient qu'un niveau de gris (ou un triplet de valeurs RGB dans le cas d'images couleurs) ce qui ne permet pas, en général, d'inférer directement des informations sur le contenu de l'image. C'est pourquoi les pixels sont généralement utilisés non pas seuls, mais par l'intermédiaire d'indices visuels, qui sont des groupes de pixels combinés au sein d'une même structure. L'objectif de la recherche étant d'obtenir une interprétation de l'image au niveau sémantique, afin de communiquer des résultats fidèles aux attentes des utilisateurs

Nous nous intéressons dans cet article à l'interprétation du contenu d'une base des images contenant des modèles arabesques à des fins d'indexation et de recherche à partir du contenu. Dans notre cas, nous utilisons une représentation locale des images au moyen des descripteurs calculés à partir des attributs de bas niveaux (couleur et texture) dans un contexte de mise en correspondance de région d'intérêt appelée aussi cluster, un descripteur de la forme n'est pas nécessaire, en effet que les modèles à chercher ne sont pas discriminés par des formes déterminées. Chaque cluster de l'image sera présenté par un descripteur de la couleur basé sur une analyse statistique des deux composantes H et S de l'espace HSL. Quant à la texture on va utiliser un descripteur basé sur le spectre de la luminance.

Pour la formation des régions d'intérêt ont fait recours à la méthode de classification **median online**

2- Formation des régions d'intérêt

La formation des régions ou objets d'intérêt est souvent nécessaire pour pouvoir s'attaquer ensuite à l'extraction d'information sémantique à partir de l'images, il consiste à trouver des structures de dépendances en commun entre pixels ou bloc de pixels et les regrouper sur des classes visuelles, ce processus est réalisé généralement, par une méthode d'apprentissage automatique des données visuelles de bas niveau.

Notre approche de clustering est réalisée par la méthode **median online** qui fonctionne selon un algorithme d'apprentissage non supervisé, les paramètres d'apprentissages hybridés qu'on a adoptés, sont composés des attributs de la couleur et de la texture (**Figure 1**).

Pour la couleur on a utilisé les moyens des zones de 8 x 8 des composantes H et S et pour la texture, on a calculé l'entropie du spectre des mêmes zones de la composante L de l'espace de couleur HSL.

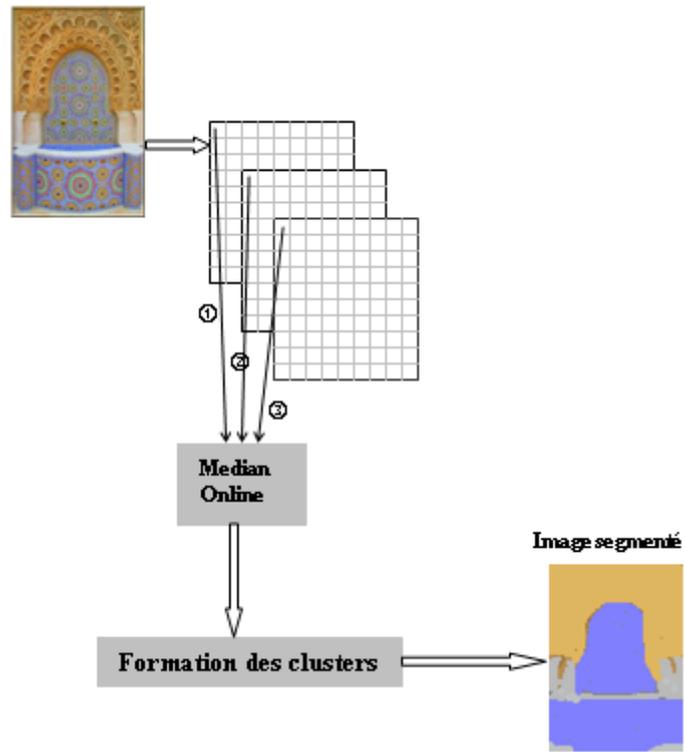


Figure 1 : déroulement de l'opération du Formation des clusters

$$Moy_H(k) = \frac{1}{n \times m} \sum_1^{1+n} \sum_c^{c+m} H(i,j) \quad \textcircled{1}$$

$$Moy_S(k) = \frac{1}{n \times m} \sum_1^{1+n} \sum_c^{c+m} S(i,j) \quad \textcircled{2}$$

$$Entp_L(k) = \sum_1^{1+n} \sum_c^{c+m} L(u,v) * \text{Log}(L(u,v)) \quad \textcircled{3}$$

$$L(u,v) = \frac{1}{n \times m} \sum_1^{1+n} \sum_c^{c+m} L(i,j) * \text{Exp}[-2j\pi (Un/n + Vm/m)] \quad \textcircled{4}$$

Avec $Moy_H(k)$ et $Moy_S(k)$ les moyennes des zones $n \times m$ des

matrices des couleurs H et S, $Entp_1(k)$ est l'entropie du spectre des zones n x m de la matrice de la luminance L et $L(u,v)$ est le transformé de Fourier discret des mêmes zones de la matrice de la luminance.

L'algorithme **online median** [1], proposé par R. Mettu et C. Plaxton est une solution *online* du problème *k-median*. Au lieu d'optimiser globalement le placement de *k* centres (*k* fixé), ceux-ci sont placés un par un, selon le principe des *Facility Location*.

L'algorithme **online median** place successivement, et de façon définitive, les centres du clustering. Le processus s'arrête lorsqu'un critère d'arrêt est vérifié, ou lorsque toutes les données ont été choisies comme centre.

L'algorithme dispose de paramètres fixés α , β , γ et δ propres à l'algorithme. Définissons tout d'abord la valeur d'une boule A de centre x et de rayon r.

$$val(A) = \sum_{y \in A} (r - d(x, y)) \quad \textcircled{5}$$

Le fils d'une boule A de centre x et de rayon r est un point y qui vérifie $d(x, y) < r\beta$. Rappelons que $d(x, y)$ représente $\min_{x \in X} d(x, y)$

Chaque étape d'ajout d'un centre se fait ainsi :

- (1) Calcul de la valeur de toutes les boules centrées en un point x des données D et de rayon $\frac{d(x,Z)}{\gamma}$ où Z est l'ensemble des centres déjà placés. Si $Z = \emptyset$; (cas du premier centre), le rayon choisi est $\max d(x, y)$.
- (2) Sélection de la boule A_0 de valeur maximale.
- (3) Tant que A_i contient plusieurs fils

Considérons les boules centrées en y vérifiant $d(x, y) \leq \beta r_i$, de rayon $r_{i+1} = \frac{r_i}{\alpha}$. Soit A_{i+1} la boule de valeur maximale, et x_{i+1} le centre correspondant. Son rayon est le r_{i+1} précédent.

- Lorsque la boule A_n n'a qu'un seul fils, on le choisit comme nouveau centre du cluster

Cela revient à mettre à jour une boule dont le rayon diminue à chaque itération (par $r_{i+1} = \frac{r_i}{\alpha}$) et qui se déplace à chaque étape vers la région de « plus forte densité », estimée à travers la valeur.

3- Indexation et recherche par le contenu

La recherche des images par des descripteurs locaux constitue une autre vision de recherche des images par le contenu. Cette nouvelle réflexion sur ce domaine a redéfini le problème au point de vue recherche, car on ne s'intéresse plus à reconnaître les éléments constituant l'image, mais on s'intéresse plutôt à la similarité entre les images [2], ou plus précisément, entre les objets et les régions qui se trouvent dans les images.

Nous partons du principe que pour chaque image **I**, nous disposons d'un ensemble de régions. Chaque région **R_j** d'une image **I** est représentée par un descripteur **D(R_j)** de l'espace visuel multidimensionnel.

Par la suite, nous allons associer à chaque image un ensemble de descripteurs visuels comme suit :

$$C(\mathbf{R}) = \{ D(\mathbf{R}_1), D(\mathbf{R}_2), \dots, D(\mathbf{R}_j) \} \text{ et } D(\mathbf{R}_j) = \{ D_c, D_T \}$$

Chaque descripteur est composé de 42 paramètres (Figure 2), 16 de la couleur H, 16 de la couleur S et 10 de la luminance L

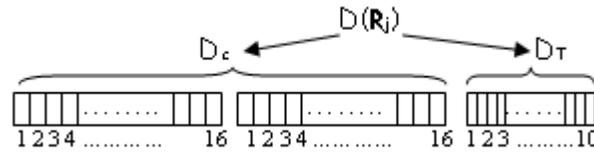


Figure 2 : modèle d'un descripteur d'une région Rj

3-1 Descripteur de la couleur

La description de la couleur de chaque région identifiée est réalisée par l'histogramme des pourcentages des couleurs [3] ou des groupements de couleurs 'Pin' de composante H et S, les valeurs sont directement calculer après l'étape de segmentation détaillé en 2. La distribution couleur de chaque région sera alors décrite par le vecteur :

$$D_c = (\delta(P_1), \delta(P_2), \dots, \delta(P_n))$$

$$\text{Avec } \delta(P) = \frac{1}{S(R_i)} \sum_{p=1}^n \varphi_p^C \quad \text{⑥}$$

$S(R_i)$ représente la surface d'une Région R_p^C , φ est une couleur d'un groupement P et de composante C . Dans notre cas les composantes de couleurs utilisés sont H et S et le nombre de Pin est fixé sur 16 pour obtenir les 32 paramètres de couleurs.

3-2 Descripteur de la texture

La notion de texture est utilisée pour traduire un aspect homogène de la surface d'un objet sur une image. La texture se manifeste donc par une information visuelle qui permet de la décrire qualitativement à l'aide des adjectifs suivants: grossière, fine, lisse, tachetée, granuleuse, marbrée, régulière ou irrégulière. Il existe plusieurs descripteurs de la texture, les plus utilisés sont les fonctions de corrélation à travers la matrice de cooccurrence, **Tamura** [4] a proposé 6 paramètres possèdent une signification visuelle la caractérisant directement.

Certains auteurs combinent plusieurs approches afin de tirer profit des différents types d'analyse, matrice de cooccurrence, et décomposition en sous bandes [5] et [6], ondelette et réseaux de neurones [7].

Pour notre cas, on a utilisé une analyse spectrale locale de la texture vue la nature des images d'arabesque qui présentent beaucoup des motifs répétitifs. D'après l'étude de **Jernigan** et **D'Astous** [8], ils ont proposé des paramètres extraits du domaine spectral. Ces paramètres mesurent des caractéristiques spécifiques de la distribution des composantes du spectre de puissance en fonction des fréquences spatiales. Les paramètres proposés quantifient des propriétés telles que la régularité, la direction, la linéarité, la finesse. Ils sont évalués à partir de l'étude du pic central et de la forme particulière du spectre de puissance en fonction de la texture. Les entropies mesurent régionalement la présence des composantes fréquentielles dans certaines bandes fréquentielles. Notons aussi que l'entropie mesurée par (Formule (3)) est invariante par rotation et par translation. En fait, cela permet de voir si une texture contient des motifs répétitifs, et elle constitue donc un bon moyen d'analyse pour ce type d'image.

Chaque région identifiée est caractérisée directement après l'étape de clustering par les 5 fort et les 5 faibles valeurs de l'entropie du spectre des zones de 8 x 8 de l'ensemble Z_R d'une région R, i.e.

$$D_T = (5 \max_{k \in Z_R} (Entp_L(k)), 5 \min_{k \in Z_R} (Entp_L(k)))$$

Les textures bruitées qui ont beaucoup de composantes fréquentielles donnent une plus grande valeur de l'entropie. Au contraire, les textures hautement structurées ont une faible entropie. Ces paramètres pris séparément ont été jugés trop peu discriminant pour des textures naturelles car ils sont sensibles au bruit.

3-2 Mesure de la similarité

Pour estimer la ressemblance entre deux régions, il faut mesurer les similarités entre leurs descripteurs correspondant aux mêmes attributs. Il existe deux façons pour combiner les deux similarités. La première

$$D_G = w_c \cdot d_c + w_T \cdot d_T \quad \textcircled{7}$$

Où d_i et w_i représentent respectivement la distance normalisée et le poids associé à l'attribut i (c pour la couleur et T pour la texture). Le poids peut être fixé interactivement par l'utilisateur ou évalué automatiquement par le système en utilisant des images de tests connue. La distance de Kullback est utilisée pour la couleur [9] tandis qu'il fait appel à la distance L1 pour la texture. La seconde approche est hiérarchique, les attribut sont alors considérés successivement par ordre décroissant d'importance, en évaluant les distances d_c et d_T

$$d_{c(R,I)} = \sum_{n=1}^H \sum_{m=1}^T (R_{n,m} \ I_{n,m}) \text{Log} \left(\frac{R_{n,m} + I}{I_{n,m} + I} \right) \quad \textcircled{8}$$

$$d_{T(R,I)} = \sum_{n=1}^P \left| Entp_R(j) - Entp_I(j) \right| \quad \textcircled{9}$$

Où H est le nombre d'éléments de l'histogramme de chaque composante de couleur, T est le nombre de composante de couleur (notre cas $H=16$ et $T=2$), et R et I , sont les pourcentages de la $m^{\text{ème}}$ composante du $n^{\text{ème}}$ couleur.

$Entp_R$ et $Entp_I$ sont respectivement les paramètres d'entropies gagnant d'une région de l'image clé et celle de l'image de la base.

La mesure de la similarité se fait pour toutes les régions détectées des images de la base, le système retourne les images dont au moins une région est pertinente. Aussi le système propose une mode de recherche personnalisée qui donne aux utilisateurs la possibilité de préciser les régions ainsi que le nombre de région à chercher dans les images de la base.

4 Système de recherche et de navigation

Le système proposé présente deux niveaux de recherche un niveau basé sur un environnement de navigation hiérarchique intelligent et un autre par une recherche directe, soit par une image entière, soit par une région de l'image clé.

4.1 L'outil de navigation

L'interface de navigation fonctionne selon une structure hiérarchique, il donne à l'utilisateur un mécanisme interactif de Zoom entre les classe n et n-1 et vice versa, jusqu'à qu'il atteigne la région ou l'image recherchée.

La difficulté rencontrée dans un tel environnement est le mode de recherche par région ou par image entière. [10] propose d'effectuer une projection de sammon sur la base, afin de projeter celle-ci sur le plan le plus discriminant. La base ensuite classée de manière hiérarchique. Dans [11] l'auteur a proposé une classification hiérarchique de la base par nuée dynamique (k-means) pour réduire le problème d'images mal placées il a utilisé une quantification floue (Fuzzy k-means) après celle effectuée par nuée dynamique. Notre approche dans cet article consiste à effectuer une classification supervisée par la méthode **K-PP** [12] pour classer les images selon les 4 classes principales **C** (Carrelage, plâtre, Bois et pierre), ensuite une classification non supervisée **CAH** (classification ascendante hiérarchique) [13] est procédée pour regrouper des modèles adjacents dans chaque classe précitée (Figure 3). Le critère d'agrégation utilisé est le lien moyen (ou distance moyenne), la distance entre deux classes est la moyenne entre les éléments des deux classes.

On distingue que les classes de l'ensemble **C** sont dissemblables, car aucune région **R_i** ne peut être affectée à deux classes à la fois :

$$\forall k \neq k' \text{ tel que } C_k(R_i) \cap C_{k'}(R_i) = \emptyset$$

Tandis que pour une image **I_j** contenant plus qu'une région peut appartenir à deux classes différentes.

$$\exists k \neq k' \text{ tel que } C_k(I_j) \cap C_{k'}(I_j) \neq \emptyset$$

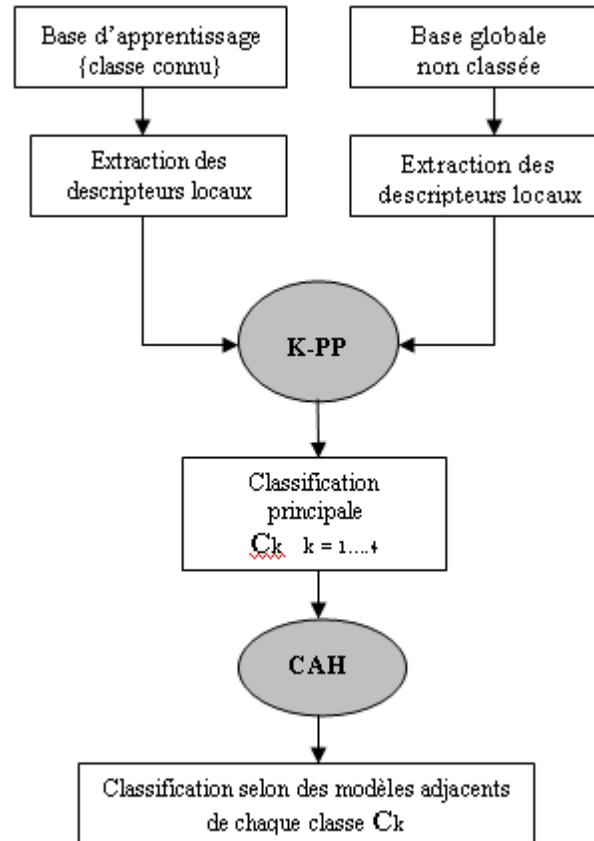


Figure 3 : Déroulement de l'opération de la classification automatique.

4.2 Outil de recherche direct

La fenêtre 4 montre l'interface du système de recherche direct.

La zone A représente La zone de l'image ou la région requête, B est le composante hiérarchique de navigation et C est la zone d'affichage des images similaires.

La recherche est effectuée soit globalement sur la base, soit Le système propose à l'utilisateur les classes de recherches par l'algorithme K-PP [12], en mesurant la similarité avec les images les plus proches aux barycentres de chaque classe C_k .

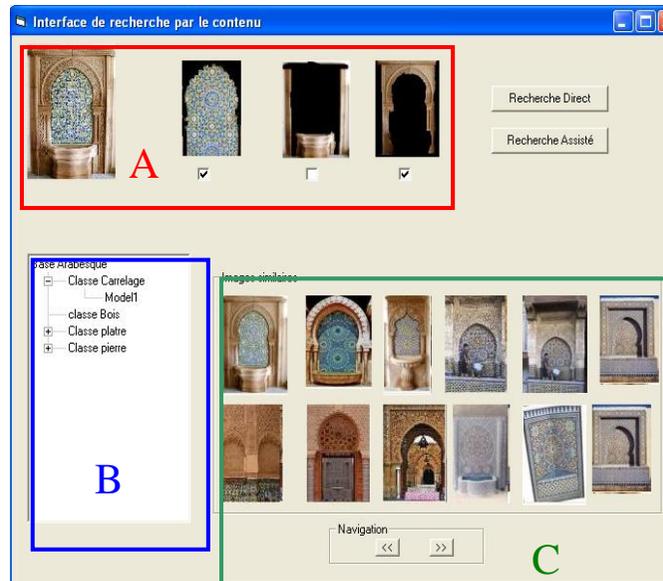


Figure4 : Interface de recherche

5. Résultats et Evaluation

Pour l'évaluation des performances du système proposée que ce soit dans la partie de la recherche ou celle de navigation, on a utilisé une base de 2000 images des arabesques, on a choisi parmi celle-ci 80 images de références qui ont été utilisées dans l'étape de l'apprentissage supervisé.

Pour l'évaluation de l'outil de la recherche pour les deux modes global et assisté (Figure 5), on a utilisé le paramètre de Rappel précision [14], ce paramètre est connu dans la littérature de la recherche par le contenu. Elle est définie comme suit:

$$\text{Rappel} = N / Q, \text{ Précision} = N / A$$

Avec N le nombre de modèles pertinents dans les A premiers modèles en réponses. Q est le nombre de modèles pertinents dans la collection qui correspondent au nombre de modèles de la classe auquel le modèle requête appartient.

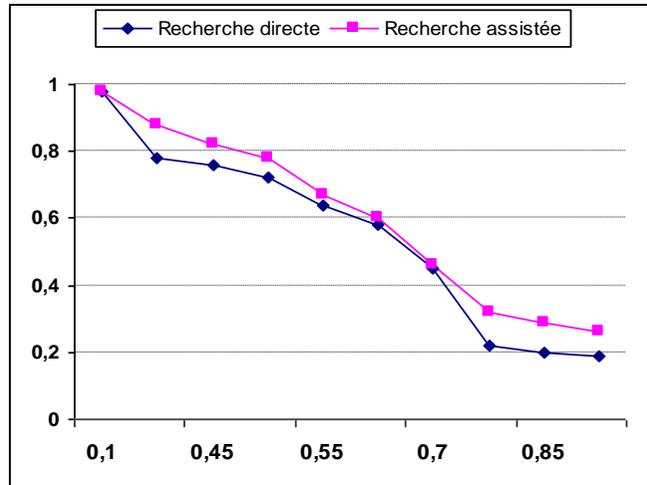


Figure 5 : Les courbes de Rappel-Précision

Ces résultats montrent que le mode de recherche assisté donne des résultats meilleurs par rapport au mode direct (pour un Rappel de 50% on a obtenu 0,78 pour le mode assisté et 0,72 pour le mode recherche direct) et avec un gain très important du temps de recherche.

Pour l'évaluation de l'outil de navigation, la performance peut être estimée en fonction du nombre d'images appartenant à une classe et présentes dans le nœud associé à cette classe

La Figure 6 montre les résultats obtenus pour chacune des 4 classes utilisées dans la base, en fonction des pondérations relatives aux descripteurs de la couleur et de la texture, le meilleur résultat réalisé pour les classes plâtre, pierre et carrelage est avec une pondération de 70% de la couleur et 30 % de la texture, quant à la classe du bois le meilleur résultat est réalisé par 80% de la couleur et 20% de la texture.

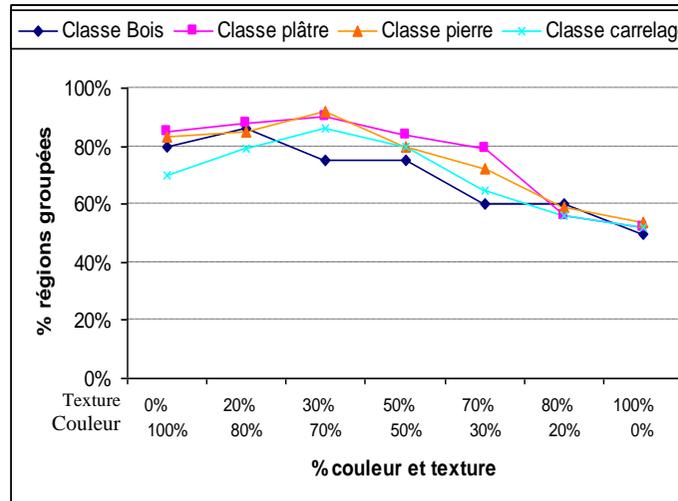


Figure 6 : Les courbes des pourcentages des régions groupés

6. Conclusion et perspective

Les résultats obtenus par ce système montrent que les descripteurs locaux sont performant dans les deux modes d'accès utilisés, la caractérisation des régions d'intérêts contenus dans l'image permet une nette amélioration de la connaissance de la sémantique de l'image. De plus le processus de classification automatique à jouer un rôle important, d'une part, une petite amélioration de la pertinence des images recherchées, et d'autre part l'efficacité du temps de recherche. Notre recherche va se concentrer sur la sémantique des images, et développer davantage l'interactivité de ce système dans la partie navigation automatique.

Bibliographie

- [1] R. R. Mettu and C. G. Plaxton : *The online median problem*, In *FOCS '00: Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, page 339, 2000.
- [2] ALAIN BOUCHER ET THI-LAN: *Indexation sémantique des images* 3rd International Conference: Sciences of Electronic, Technologies of Information and Telecommunications March 27-31, Setit 2005 – TUNISIA
- [3] SWAIN, M. J. ET BALLARD, D. H.: *Color indexing. International Journal of Computer Vision*, 7(1) :11–32. 1991. pages 12, 13
- [4] R. Haralik: *Statistical and structural approach to textures*, Proceedings of the IEEE, vol 67 NO. 5, pp 786-804, 1979
- [5] K. S THYAGARAJAN, T. NGAYEN, C. PERSONS: *Maximun like lihood approach to texture classification using Wavelet transform*, In Proc, IEEE International conference on image Processing 1994
- [6] A. KUNDU, J.L. CHEN: *Texture classification using qnf bank-based Subband decomposition* CVGIP
- [7] C. BUSH, M. H. GROSS : *Interactive neural network texture analysis and visualization for surface reconstruction in medical imagin*, In R. J. Hubbold and R. Juan, editors, *Enrographics 93* pp 49-60, Oxford, UK Eurographics, Backwell Publishers, 1993
- [8] M.E.JERNIGAN AND F.D'ASTOUS: *Entropy-based texture analysis in the spatial frequency domain*, IEEE Trans. on pattern analysis and machine intelligence 1984, vol.PAMI6
- [9] S. Kullback : *Information theory and statistics*. New York: John Winley & Sons, 1959.
- [10] Z. PECENOVIC, M. DO, M. VETTERLY, P. PU: *Integrated browsing and searching of large collection*, 2000
- [11] K. IDRISSE, G LAVOUE, J RICARD: *Système de Recherche d'Images et de Navigation Visuelle Basé sur une Approche locale*. Int. Conf. on Image and Signal Processing '03 2003.
- [12] T. HASTIE, R. TIBSHIRANI, AND J. FRIDMAN: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, 2001.
- [13] WILLIAMS, WT. 1967 *A general theory of classificatory sorting strategies : I. Hierarchical systèmes*. Computer Journal. 9, 378 - 380
- [14] J.R. SMITH: *Image Retrieval Evaluation, CBAIVL'98*: Proceedings of the IEEE Workshop on Content - Based Access of Image and Video Libraries, pp. 112, 1998.