

Indexation et recherche d'images par le contenu visuel

- Pour trouver l'information qu'il recherche, l'utilisateur dispose d'au moins trois types de moyens:
- les requêtes par exemple visuel (1) (*query by example*). L'utilisateur fournit un exemple, le système en extrait des descripteurs visuels, puis recherche les images ayant des descripteurs similaires. L'exemple peut être :
 - une image : la requête de l'utilisateur est composée d'une image entière
 - une région : l'utilisateur recherche des régions d'images qui ressemble à la région de l'image sélectionnée par l'utilisateur
 - une composition logique de catégories de régions : l'utilisateur choisit les régions parmi des régions construites à partir d'un résumé (thésaurus photométrique) des régions de la base.
 - une composition spatiale de régions : l'utilisateur recherche une image ayant la région en haut de la même couleur que l'objet A, et la région à droite de la même texture que l'objet B

- un croquis dessiné par l'utilisateur : le croquis est un cas particulier de requête par l'exemple, car dans ce cas là, le principal descripteur utilisé est la forme
- les requêtes directes (2) (*direct query*). Les requêtes directes sont composées de descripteurs visuels. On distingue :
 - les requêtes par caractéristiques visuelles (par exemple, 25% de rouge, 30% de bleu et 45% de vert)
 - les requêtes par combinaison pondérée de caractéristiques visuelles (par exemple, la couleur a un poids de 75% et la texture de 25%).
- la navigation (4) (*browsing*) dans les images.

Extraction des information visuelles

- Les valeurs des pixels d'une image ou d'une région d'images ne peuvent pas être exploitées directement. C'est pourquoi on extrait à l'aide d'algorithmes plus ou moins complexes des descripteurs visuels afin d'obtenir une représentation plus facile à utiliser, et qui correspondent si possible au contenu sémantique
 - . L'extraction des informations visuelles des images doit être effectuée aussi bien pour les images de la base que pour la requête.
 - Elle est généralement constituée de trois étapes:
 - d'abord, le système extrait des parties de l'images (régions, zones d'intérêt, points d'intérêt...) choisies en fonction de l'information qu'elles contiennent (4) et (5)
 - ensuite, le système extrait les descripteurs visuels, c'est-à-dire les caractéristiques de couleurs, de textures, de formes... de chacune des parties (phase de caractérisation) (6) et (7).
 - enfin, une représentation (appelée parfois signature ou index) est parfois nécessaire pour résumer les descripteurs visuels en une forme encore plus exploitable par le système (phase d'indexation) (8) et (9).
- La représentation est classiquement un vecteur (histogramme, distribution...), mais peut également être un ensemble de vecteurs, des graphes...

Descripteurs visuels

- Puisque l'on souhaite construire des systèmes de recherche d'images qui soient utilisables par l'être humain, et que c'est l'être humain qui donne un sens à ce qu'il voit, il peut être intéressant de s'inspirer du système de perception humain pour choisir les espaces visuels afin de s'approcher le plus possible de la compréhension qu'à l'être humain de l'image, et ainsi réduire le fossé sémantique.

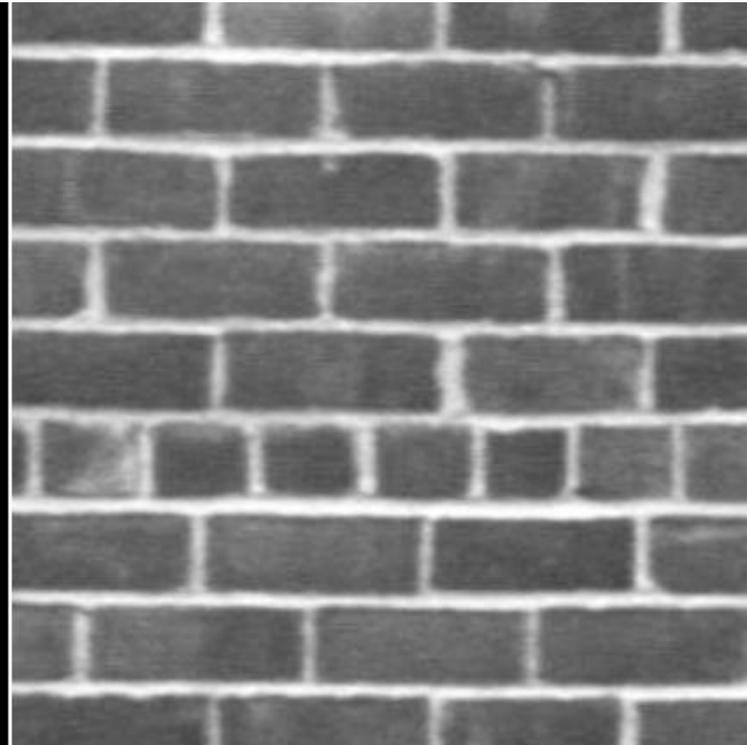
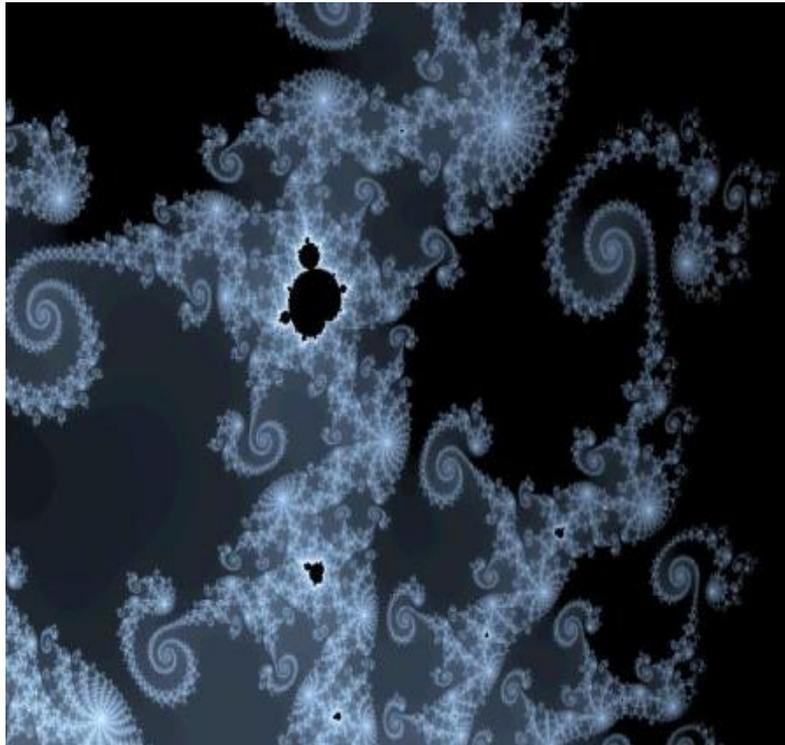
1.Couleur

- C'est la caractéristique visuelle la plus utilisée dans les CBIR car elle correspond la description intuitive la plus courante chez l'humain. La représentation la plus usuelle de la couleur est l'espace RVB (Rouge, Vert, Bleu). L'espace RVB est celui employé par les caméras. Cependant, il ne correspond pas au processus de perception humaine. L'espace HSV (Hue-Saturation-Value) sépare les informations relatives à la teinte (Hue), la saturation (Saturation) et l'intensité (Value). Il correspond à la façon dont nous percevons les couleurs. D'autres espace couleur sont aussi considéré dans différents CBIR (Lab., Luv, ...). L'espace HSV est le plus efficace pour la recherche d'images par le contenu.

2. La texture

- Il n'existe pas de définition pertinente de la texture. Cependant, une définition de sens commun est la suivante :
- la texture est la répétition d'éléments de base construits à partir de pixels qui respectent un certain ordre. Le sable, l'eau, l'herbe, la peau sont autant d'exemples de textures.
- L'aléatoire joue un rôle particulier dans les textures. On peut distinguer deux types extrêmes de textures, entre lesquels se positionnent toutes les textures :

Il existe deux types de textures: texture périodiques (Fig.1)
définies par la répétition d'un motif de base.



Textures dites aléatoires qui ont un aspect anarchiques (Fig.2) mais qui apparaissent homogènes vis-à-vis de leur environnement.



Les outils classiques des statistiques (moyenne, variance, moment, énergie, entropie...) peuvent être utilisés pour analyser les textures. Par exemple, le moment d'ordre 3 (*skewness*) donne une indication sur la symétrie ; le moment d'ordre 4 (*kurtosis*) donne une indication sur l'aplatissement. Les matrices de cooccurrences sont également des outils adaptés pour l'analyse de texture.

3. La forme

L'utilisation de descripteurs de forme (*shape*) n'a de sens que sur une image segmentée. Pour extraire les descripteurs d'une forme, la première chose à faire est de définir sa fonction caractéristique. En général, elle est représentée sous la forme d'un masque dans lequel chaque pixel est représenté par le numéro de la région à laquelle il appartient. C'est à partir de cette fonction, que sont calculés la plupart des descripteurs de formes, soit à partir de la région entière, soit à partir des contours seulement.

On peut citer quelques descripteurs classiques :

les moments d'inertie. Ils ont la propriété d'être invariants par rotation. Ils décrivent bien l'allongement de formes régulières comme des ellipses ou des distributions gaussiennes. Ils sont plus ambigus sur des formes complexes ;

la recherche des boîtes englobantes ou minimales de la région ;

les moments invariants (nommés moments de Hilbert), invariants par translation, rotation et changement d'échelle ;

les polygones de Guzman (1968). Cette approche consiste à envelopper l'objet à reconstruire dans des boîtes de formes de plus en plus précisément adaptées ;

descripteurs de Fourier ;



4.Segmentation et points d'intérêt

L'extraction de descripteurs visuels sur l'image entière (descripteurs globaux) permet de réduire le nombre de calculs nécessaires, la taille de la base de données ainsi que le coût des recherches des images les plus similaires. Cependant, l'approche globale ne permet pas une recherche efficace d'objets (au sens large) dans l'image. A l'inverse, les descripteurs extraits d'une partie de l'image (descripteurs locaux) sont efficaces, mais coûteux. Les descripteurs locaux peuvent être :

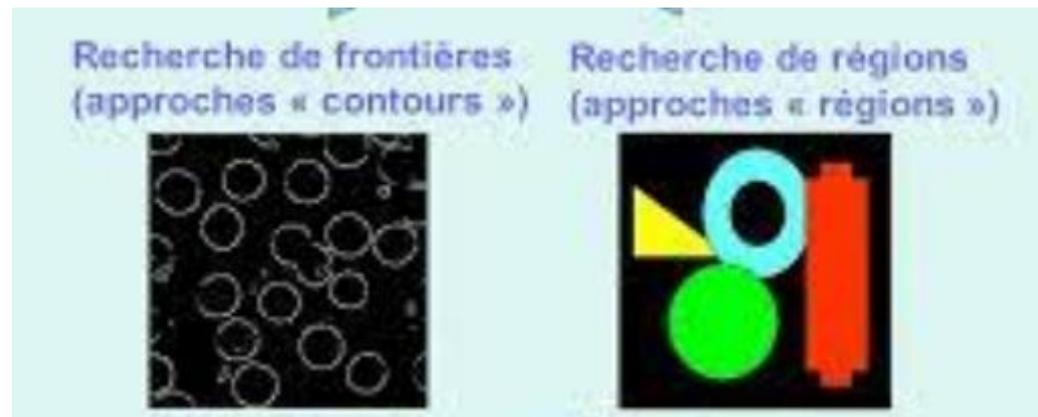
- des régions de l'image obtenues soit par segmentation de l'image entière, soit par recherche de régions d'intérêt,
- des points d'intérêt.

4.1. Segmentation

on distingue cependant deux grandes familles d'algorithmes :

la segmentation par approche «contour» ou «frontière» (*edge-based segmentation*). Un contour est une frontière entre deux milieux différents (2 couleurs, 2 niveaux de gris. Un algorithme classique d'extraction du contours est l'algorithme de Canny

la segmentation par approche «région» (*region-based segmentation*). Le principe est de trouver les régions en regroupant les pixels ayant des caractéristiques similaires et en séparant ceux qui sont différents (techniques division-fusion, par accroissements de régions, par statistiques bayésienne...).



4.2 Points d'intérêt

Les points d'intérêt d'une image sont les points qui seront trouvés similaires dans les images similaires. Une manière de les déterminer est de prendre en compte les zones où le signal change.

Par exemple, les points d'intérêt peuvent être les coins, les jonctions en T ou les points de fortes variations de texture.

Trois types d'approche pour l'extraction de points d'intérêt :

approches contours : les contours d'une image sont d'abord détectés, puis les points d'intérêt sont extraits le long des contours en considérant les points de courbures maximales ainsi que les intersections de contours.

approches intensité : la fonction d'intensité est utilisée pour extraire directement des images les points de discontinuité.

approche à base de modèle : les points d'intérêt sont identifiés dans l'image par mise en correspondance de la fonction d'intensité avec un modèle théorique.

Merci