

Indexation et recherche d'images par le contenu

1. Introduction

Nous assistons ces dernières années à une croissance énorme de l'information disponible dans des banques de données. Ainsi, nous manipulons quotidiennement un flot croissant d'information multimédia tel que le texte, le son, l'image fixe ou animée; ou une combinaison de ces différents types de données. Le problème avec ces informations est qu'elles ne sont pas stockées d'une manière structurée, par conséquent il est difficile de retrouver l'information qu'on recherche en un temps opportun. L'exemple le plus illustratif actuellement est le Web. Mais, il n'y a pas que le Web. On peut trouver des banques de données disséminées un peu partout: dans un centre de télédétection, dans un centre médical, dans un centre de police, dans une librairie digitale ; qui pour les interroger et les exploiter on a besoin de système de recherche efficace. Par conséquent, l'utilité et l'importance des systèmes automatiques d'indexation et de recherche d'informations ne sont plus à démontrer. De plus, la recherche dans une grande base de données, en particulier des bases d'images est un problème d'une importance cruciale pour de nombreuses applications telles que:

- Les applications scientifiques (Imagerie médicale, ...).
- Les applications grand public (commerce électronique....).
- L'authentification (détecter les contrefaçons, identifier un visage, ...).
- L'art, l'éducation (recherche encyclopédique d'illustrations,...).
- Le « design », la publicité (rechercher une texture spécifique pour l'industrie textile, illustrer une publicité par une image adéquate, ...).

2. Système IRIS

2.1. La phase d'indexation (hors-Ligne) :

Dans cette phase, des caractéristiques sont automatiquement extraites à partir de l'image et stockées dans un vecteur numérique appelé descripteur visuel. Grâce aux techniques de la base de données, on peut stocker ces caractéristiques et les récupérer rapidement et efficacement.

2.2. La phase recherche (On-line) :

Dans cette étape, le système analyse une ou plusieurs requêtes émises par l'utilisateur et lui donne le résultat correspond en une liste d'images ordonnées, en fonction de la similarité entre leur descripteur visuel et celui de l'image requête en utilisant une mesure de distance.

La figure 1 schématise le fonctionnement d'un système de recherche et d'indexation d'images.

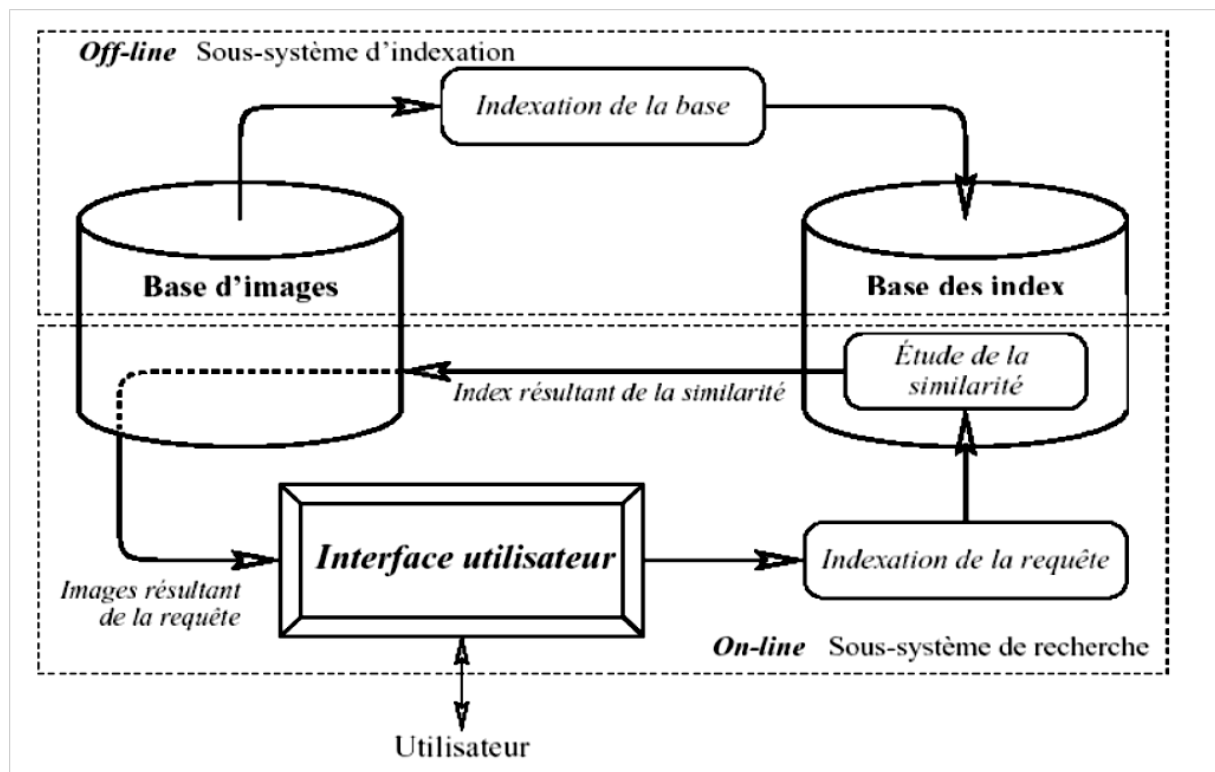


Fig.1. architecture d'un système d'indexation et recherche d'images par le contenu

3. Les requêtes

Il existe 3 façons de faire une requête dans un système d'indexation et recherche des images :

- requête par « **mots clés** »
- requête par « **esquisse** »
- requête par « **exemple** »

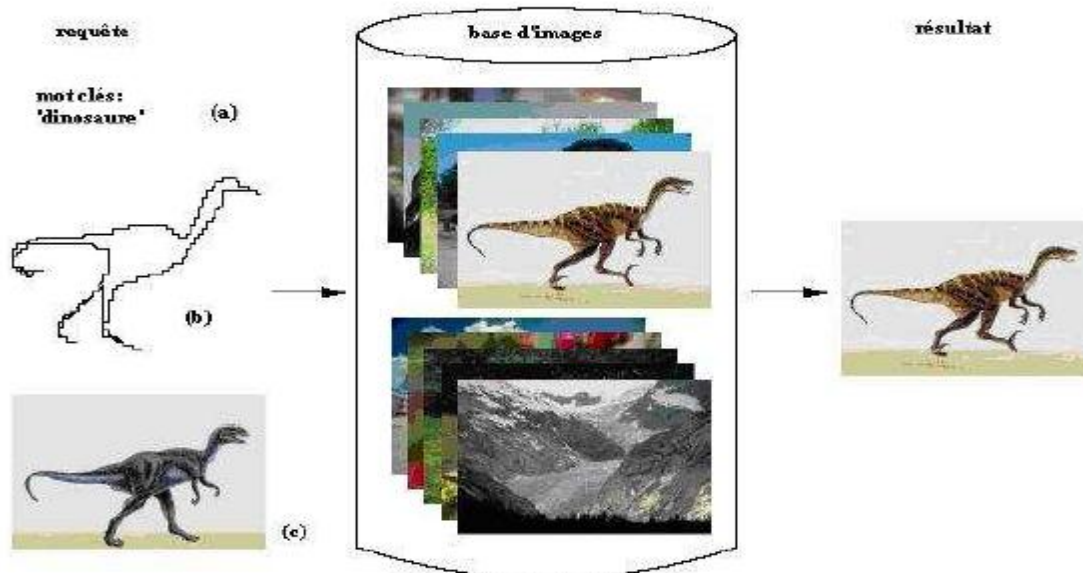


Fig.2. Trois façons pour donner une requête.

4. Exemples de systèmes IRIS :

- QBIC (Query By Image Content) d'IBM.
- Virage.
- Photobook développé par MIT Media Laboratory.
- Blobworld
- BDLP (Berkeley Digital Library Project).
- Netra, FRIC, VisualSeek, SIMPLcity, IKONA, Cortina, ...etc.

5. Caractéristiques symboliques (descripteurs)**5.1. Caractéristiques globales****5.1.1. La couleur**

La couleur est très souvent le premier descripteur qui est employé pour la recherche d'images. Ils existent de nombreuses possibilités d'attributs pour caractériser la couleur tels que l'histogramme, les moments couleur...etc.

5.1.2. L'histogramme

Les histogrammes sont faciles et rapides à calculer, et robustes à la rotation et à la translation. Cependant l'utilisation d'histogrammes pour l'indexation et la recherche d'images pose quatre problèmes.

- ils sont de grandes tailles, donc par conséquent il est difficile de créer une indexation rapide et efficace en les utilisant tels quels.
- ils ne possèdent pas d'informations spatiales sur les positions des couleurs.
- ils sont sensibles à de petits changements de luminosité, ce qui est problématique pour comparer des images similaires, mais acquises dans des conditions différentes.
- ils sont inutilisables pour la comparaison partielle des images (objet particulier dans une image), puisque calculés globalement sur toute l'image.

5.1.3. Les espaces de couleur

La deuxième approche recherche d'autres espaces de couleurs qui se basent sur la perception de couleur de l'humain. L'espace RVB est utilisé couramment mais il n'est pas forcément le mieux adapté. En effet, les trois composantes RVB sont très dépendantes les unes les autres. Un simple changement d'éclairage modifie les trois composantes, alors que les objets de la scène n'ont pas changé de couleur.

L'espace TSV (Teinte Saturation Valeur - *en anglais HSV*) est le plus utile pour la segmentation et la reconnaissance et il a été prouvé un espace très fort dans le système de recherche des images. Parce qu'il est une représentation plus physique de la couleur. Dans cet espace, on peut séparer pour un pixel : l'intensité du pixel (valeur) et la couleur du pixel (teinte et saturation).

5.1.4. Les moments statistiques

La méthode d'histogramme utilise la distribution complète de la couleur. On doit stocker de nombreuses données. Au lieu de calculer la distribution complète, dans les systèmes de recherche d'images, on calcule seulement des dominantes caractéristiques de couleur tels que :

- l'espérance,
- la variance
- et d'autres moments.

L'espérance, la variance, les moments d'ordre 3 peuvent également être calculés sur chaque composante couleur par la formule suivante :

$$E_i = \frac{1}{N} \sum_{j=1}^N p_{ij} \quad (1)$$

$$\delta_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^2 \right)^{1/2} \quad (2)$$

$$s_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^3 \right)^{1/3} \quad (3)$$

Avec i est la composante couleur (R, V, B de l'espace RVB ou T, S, V de l'espace TSV).

5.1.5. La texture

Une méthode très connue pour analyser la texture est la matrice de co-occurrences de Haralick. Haralick a proposé 14 caractéristiques statistiques extraites à partir de cette matrice. Actuellement, seulement les quatre caractéristiques les plus appropriées sont largement utilisées :

- l'énergie,
- l'entropie,
- le contraste
- le moment inverse de différence.

Il existe aussi d'autres méthodes pour analyser les textures dont celles basées sur la transformée en ondelettes de Gabor, sur les caractéristiques de Tamura, Word decomposition, Fractal...

5.2. Caractéristiques locales

Les caractéristiques présentées précédemment sont calculées de façon globale sur l'image. Soit une image composée de plusieurs objets ayant des caractéristiques très différentes, le vecteur de caractéristiques globales extrait à partir de l'image entière perd les informations

locales (les objets) et ne produit qu'une moyenne grossière du contenu de cette image. Par contre, l'analyse uniquement basée sur des caractéristiques locales risque de perdre le sens global de l'image.

Par conséquent, un compromis doit être trouvé, différent selon les applications et selon les requêtes individuelles, entre caractéristiques globales et caractéristiques locales.

Deux approches différentes peuvent être employées pour calculer les caractéristiques locales :

1. La première approche consiste à diviser une image en utilisant une grille et les caractéristiques de chaque bloc sont calculées pour chaque case de cette grille. Par exemple dans la figure 3.a, l'image est divisée en 12 zones de même taille et les caractéristiques sont calculées pour chacune de ces zones.
2. La deuxième approche consiste à segmenter l'image pour la diviser en zones locales plus proches des objets constituant l'image et ensuite calculer les caractéristiques pour chacune des régions extraites.



Fig.3. Deux découpages différents de l'image. (a) Découpage en grille de l'image. (b) Découpage par la segmentation de l'image.

Remarque :

La segmentation est une étape qu'on souhaiterait éviter, en raison de tous les problèmes pour choisir une "bonne" méthode de segmentation valide pour toutes les images de la base d'images. Cependant, la division d'une image en régions ou objets d'intérêt est souvent nécessaire pour l'extraction d'informations sémantiques à partir de l'image.

Il s'agit simplement d'oublier le découpage précis, mais plutôt de découper l'image en régions similaires du point de vue caractéristiques mesurées sur l'image.

5.3. Autres caractéristiques de l'image

Il existe plusieurs autres caractéristiques ou descripteurs, de l'image qui peuvent être utilisées pour mesurer la similarité d'une image, telles que :

- la forme,
- les points d'intérêt,
- les invariants, et plus encore.

La forme est une caractéristique visuelle importante et elle est une des caractéristiques de base pour décrire le contenu d'une image. Cependant, la description des formes est une tâche difficile.

6. Mesure de similarités entre descripteurs :

L'approche la plus simple pour mesurer la similarité entre deux images correspond aux distances de Minkowski. Cette distance est calculée entre les vecteurs descripteurs, elle est définie par :

$$L_r(V_1, V_2) = \left[\sum_{i=0}^n |V_1(i) - V_2(i)|^r \right]^{1/r} \quad (4)$$

Pour $r = 1$, on obtient la distance de Manhattan ou city block. **Pour $r = 2$** , on obtient la distance euclidienne. **Pour $r = \infty$** , on obtient la distance de Chebyshev ou la distance de maximum.

7. Mesures pour évaluer un système IRIS

Les mesures les plus courantes pour évaluer un système sont : **le temps de réponse** **l'espace utilisé**. Plus le temps de réponse est court, plus l'espace utilisé est petit, et plus le système est considéré bon. Mais avec des systèmes qui ont été faits pour la recherche d'informations, en plus de ces deux mesures, on s'intéresse à d'autres mesures.

7.1. Le rappel:

Le rappel est le rapport entre le nombre d'images pertinentes dans l'ensemble des images trouvées et le nombre d'images pertinentes dans la base d'images.

$$Rappel = \frac{|Ra|}{|R|} \quad (5)$$

7.2. La précision :

La précision est le rapport entre le nombre d'images pertinentes dans l'ensemble des images trouvées et le nombre d'images trouvées.

$$Précision = \frac{|Ra|}{|A|} \quad (6)$$

Où :

$|R|$: le nombre d'images pertinentes dans la base d'images.

$|A|$: le nombre d'images dans l'ensemble des réponses.

$|Ra|$: le nombre d'images pertinentes dans l'ensemble des réponses