

# Une approche sémantique basée sur l'apprentissage pour la recherche d'image par contenu

**Hichem Bannour**

*Département des Sciences d'informatique  
Faculté des Sciences de Monastir  
Université de Monastir, 5000 Tunisie.*

*Hichem.Bannour@issatso.rnu.tn  
Bannour.Hichem@yahoo.com*

---

*RÉSUMÉ. Cette dernière décennie témoigne un accroissement exponentiel des données multimédia (texte, image, son et vidéo). La recherche d'information au sein de cette masse de données, en particulier les images, devient un processus incertain. Aussi, le problème se pose au niveau de l'indexation puisque les techniques actuelles ne permettent pas de décrire efficacement le contenu des images. Dans ce papier, nous nous concentrons sur le problème de découverte de connaissance à partir du contenu des images et nous proposons une nouvelle approche pour l'indexation sémantique des images. Notre approche se base sur l'apprentissage pour associer des éléments symboliques (concepts) à des éléments de bas niveau.*

*ABSTRACT. Multimedia databases are witnessing an exponential growth. Information retrieval in these data collections, especially in images, becomes a hard task. Furthermore, there is a difficulty in indexing these data since the present techniques don't allow an efficient description of the content of images. In this paper, we focus on knowledge discovery in images database and we propose a new semantic approach for image indexation. Our approach is based on learning of distributions to map low level features to high level concepts.*

*MOTS-CLÉS : Fouille d'image, indexation et recherche d'image, RImC, attributs d'image, reconnaissance d'objet, segmentation d'image.*

*KEYWORDS: Image mining, image indexing and retrieval, CBIR, image features, object recognition, image segmentation.*

---

Hichem Bannour

## 1. Introduction

Il devient très difficile et imprécis, de nos jours, de faire la recherche d'information multimédia sur le Web. Les techniques d'indexation et de référencements par mots clés ne suffisent plus à pallier aux besoins croissants des utilisateurs, de plus en plus exigeants en termes de précision face à leurs requêtes.

Pour palier ce problème la communauté scientifique s'est intéressée depuis quelques années à une nouvelle approche appelé Recherche d'Image par Contenu (RImC) et qui consiste à représenter des images par des attributs de bas niveau tels que la couleur, la texture et les formes. Seulement cette approche a été confrontée à un problème important qui est le manque de formalisme précis permettant à un utilisateur d'accéder à ces informations de bas niveau. En effet, un utilisateur cherchant une image sur le web se soucie peu de ses caractéristiques de bas niveau, et formule une requête exprimant une sémantique d'un certain ordre. Cependant, les méthodes actuelles de RImC sont incapables d'abstraction, en d'autres termes, incapable d'extraire de l'information sémantique à partir des images.

Le domaine de la RImC doit donc faire face à une caractéristique importante des images qui est le manque d'un langage de représentation explicite permettant d'en exprimer la sémantique. Cet écart entre les pixels et leur signification est appelé le fossé sémantique (Semantic gap).

Dans ce contexte, l'image mining semble apporter des solutions pour la description sémantique des images. Dans le but d'une indexation sémantique des images, l'image mining consiste en la découverte de connaissances à partir du contenu des images d'un ensemble d'apprentissage, sous la forme d'associations entre des éléments sémantiques (concepts) et des éléments de bas niveau (caractéristiques des images).

Notre objectif dans ce travail est alors d'exploiter l'information contenue dans ces données, et d'en extraire une sémantique fiable pour un besoin de recherche d'information et plus particulièrement pour une indexation sémantique automatique, ou semi automatique des images dans un système de recherche d'information multimédia. Dans notre approche, nous utilisons une représentation des caractéristiques symboliques et des caractéristiques numériques dans un même espace et détectons des relations sémantiques entre les éléments de l'espace pour produire à la fin un vecteur descripteur représentant le contenu de l'image. Notre approche est basée sur l'utilisation d'une collection d'apprentissage d'images. Cette collection permettra de découvrir les régularités qui seront utilisées ultérieurement pour indexer des nouvelles images.

Le reste de ce papier est organisé comme suit, dans la section 2 nous présentons les travaux connexes dans le domaine de recherche d'image par contenu. Nous proposons notre modèle d'indexation sémantique des images dans la section 3. La section 4 introduit les résultats attendus par notre approche, puis nous discuterons les avantages et limites de cette approche dans la section 5. Enfin, nous terminerons par la conclusion et nos perspectives dans la section 6.

## 2. Travaux Connexes

Avec l'explosion des données multimédia sur le web, et en particulier les images, la mise en place d'un système pertinent de recherche d'image devient une nécessité. On distingue deux approches dans la littérature pour la recherche d'image : l'approche basée sur l'annotation textuelle et l'approche basée sur le contenu.

– La première approche est la plus utilisée, et consiste à une indexation des images par le moyen de mots clés. Néanmoins, elle nécessite un effort considérable pour une bonne description de l'image et reste encombrante et limitée.

– Une deuxième approche est la recherche d'image par le contenu (RImC) qui consiste à une indexation des images par des attributs de bas niveau tels que la couleur, la texture et les formes (Cox *et al.*, 2000, Deselaers, 2003, Flickner *et al.*, 1995).

La RImC n'a pas montré son efficacité dans le domaine de la recherche d'information (RI). Un inconvénient majeur de cet axe est qu'il est sémantiquement très faible ce qui le rend non adapté à un besoin de RI. En effet, la plupart des systèmes de recherche d'images traditionnels, tel que Photobook ou QBIC (Flickner *et al.*, 1995), se limitent à une recherche en termes de similarité visuelle.

Le domaine de l'image mining tente de trouver des solutions à ce problème. Les travaux dans cet axe essayent d'attribuer une certaine sémantique aux connaissances extraites à partir des images (Lavrenko *et al.*, 2003, Rasiwasia *et al.*, 2007, Carneiro *et al.*, 2007, Barnard *et al.*, 2003, Cheng *et al.*, 2005). Cependant les approches proposées jusqu'à nos jours ne sont pas assez développées, puisque cet axe est confronté à plusieurs obstacles :

– Le niveau d'analyse croît considérablement en fonction des connaissances à extraire - voir Figure 1.

– Le fossé sémantique <sup>1</sup> et le fossé sensoriel <sup>2</sup> définis par Smeulders dans (Smeulders *et al.*, 2000).

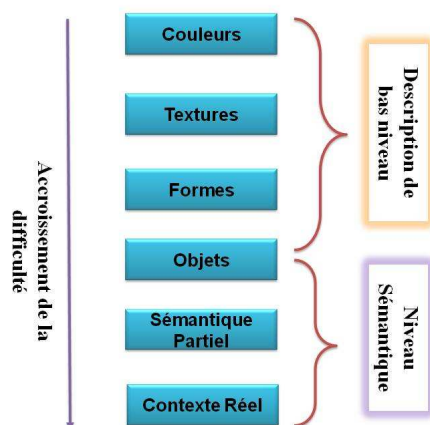
– Les problèmes liés au traitement d'image (absence de méthode universelle pour la segmentation d'image, la détection du contour, la reconnaissance d'objets, etc.).

En vue d'une description sémantique des images, Ordonez *et al.* (Ordonez *et al.*, 1999) ont appliqué un algorithme de découverte de règles d'associations à des images de synthèse. Ces images ont une description très pauvre sous forme de "tâches" (traduction de blobs). Ils n'ont utilisé que deux propriétés des images : la couleur et la texture. Une tâche est donc une région de pixels connexes cohérente au niveau de la similarité de couleur et de texture. Ces tâches sont obtenues en segmentant les images. Les règles d'associations sont exprimées entre des tâches des images. Les résultats de

1. The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.

2. The sensory gap is the gap between the object in the world and the information in a (computational) description derived from a recording of that scene.

Hichem Bannour



**Figure 1.** Niveau d'analyse pour les systèmes de RImC.

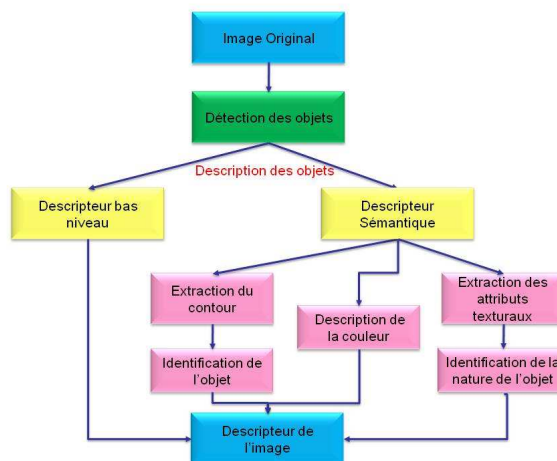
Ordonez semblent prometteurs même si les propriétés auxquelles le mining a été appliqué sont assez simples, et donc pauvre sur le plan sémantique.

Une autre approche est de combiner des données textuelles avec des données visuelles de l'image pour les inclure dans un même espace. Un exemple de cette application est présenté dans (Parsons *et al.*, 2003) où les caractéristiques des images sont extraites sous forme de termes ou descripteurs (exemple "voiture rouge", "voiture non-rouge", "toits", "arbre", "route", etc.) pour être combinées avec les termes du texte qui accompagnent les images. Ces éléments sont représentés dans le même espace en utilisant la technique d'indexation sémantique latente (LSI). Les dimensions de l'espace vectoriel obtenu avec cette technique ne sont pas étiquetées ce qui les rend sémantiquement très pauvres et la segmentation difficilement interprétable.

Dans ce papier, nous nous concentrons sur le problème de découverte de connaissances à partir du contenu des images sous la forme d'associations entre des éléments symboliques (concepts) et des éléments de bas niveau.

### 3. Modèle proposé

Le principal objectif de ce travail est de développer une méthodologie pour la représentation et l'extraction des connaissances à partir des images en utilisant des techniques d'intelligence artificielle, afin de ramener la recherche d'image à une opération d'inférence utilisant les connaissances récoltées dans les images. La solution consisterait donc à décrire d'une façon automatique le contenu des images. Ceci, permettrait de dépasser le fossé sémantique, en d'autres termes le problème lié à la subjectivité des utilisateurs de la base.



**Figure 2.** Architecture du modèle proposé.

Comme le montre la figure 2, le modèle que nous proposons utilise deux types de descripteur pour l'indexation des images : un descripteur sémantique qui permet la recherche d'image par la sémantique et un descripteur de bas niveau qui servira comme fonction d'appariement pour ordonner le résultat de la recherche. Le descripteur de bas niveau peut aussi être utilisé pour la recherche en termes de similarité visuelle ou lorsque l'utilisateur juge que le résultat de la recherche sémantique n'est pas satisfaisant.

Le principe de notre approche repose sur la description sémantique des objets contenus dans l'image. Pour aboutir au résultat souhaité plusieurs étapes sont nécessaires :

1) La première étape consiste à séparer les objets de l'image. Pour cela il est nécessaire d'utiliser un algorithme de segmentation d'image. L'algorithme de classification PFCM (Pal *et al.*, 2005) peut suffire pour obtenir un résultat satisfaisant.

2) La description de bas niveau de chaque région de l'image. Ce descripteur servira à établir un ordre de similarité pour le résultat de la recherche ou encore à faire une recherche d'image se basant sur la similarité visuelle. La méthode (Bannour *et al.*, 2009) peut être utilisée pour définir ce descripteur.

3) La description sémantique de chaque objet de l'image.

a) Afin d'identifier l'objet, il faut dans un premier lieu détecter le contour de l'objet (Canny, 1986), puis utiliser un algorithme d'apprentissage pour la reconnaissance de forme (Carpenter *et al.*, 1987). Ce descripteur a pour rôle de générer des labels comme : table, avion, personne, etc.

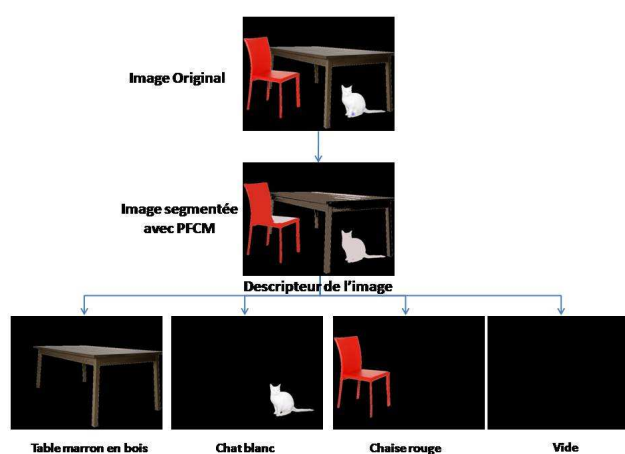
b) Nommer la couleur de l'objet (Mojsilovic, 2005). Ex. vert, noire, rouge, etc.

Hichem Bannour

c) Pour identifier la nature de l'objet, on commence par extraire les attributs texturaux de l'objet (Haralick, 1979), puis on utilise un algorithme d'apprentissage pour la reconnaissance du modèle (Carpenter *et al.*, 1987). Ce descripteur a pour rôle de générer des descriptions comme : objet en bois, en marbre, en brique, etc.

4) Enfin, construire un vecteur descripteur de l'image qui va regrouper la description de bas niveau et la description sémantique de l'image.

#### 4. Résultats



**Figure 3.** Résultat de la description sémantique par le modèle proposé.

La figure 3, illustre le résultat attendu par notre approche pour la description sémantique des images. L'indexation se base sur la description automatique des objets contenus dans l'image (appellation, couleurs et natures). La figure montre le résultat de la segmentation de l'image par PFCM, puis la description de chaque objet (Ex. objet 1 : la reconnaissance de la forme de l'objet a retourné le label "table", la reconnaissance de la texture a retourné la nature de l'objet "bois", l'identification de la couleur a donné la couleur "marron"). Une description détaillée du descripteur de bas niveau ainsi que les résultats obtenus avec est illustré dans (Bannour *et al.*, 2009).

#### 5. Discussion

Actuellement le modèle que nous proposons est en phase de développement. Nous avons déjà implémenté la partie description de bas niveau, ainsi que la partie reconnaissance de forme pour l'annotation des images. Le descripteur de bas niveau, comme nous l'avons déjà indiqué, permet de jouer le rôle de fonction d'appariement lors de la

recherche sémantique d'image, il permet aussi la recherche d'image au niveau similarité visuelle. Les résultats obtenus pour la recherche visuelle des images sont reporté dans (Bannour *et al.*, 2009). Les résultats de la recherche sémantique basée sur la reconnaissance de formes seront bientôt publiés.

Dans le cas d'une recherche sémantique des images, le modèle que nous proposons permet de traiter trois types de requêtes : les requêtes par mots clés, les requêtes par croquis et les requêtes par l'exemple. Dans le cas de la recherche d'image par l'exemple, l'image exemple sera indexée et les mots clés obtenus seront utilisés pour être comparés à ceux des images de la base. Dans le cas d'une requête par croquis, la recherche sera basée sur les formes reconnues dans le croquis et dans les images de la base. Pour la recherche par mot clés, la requête sera comparée aux mots clés des images de la base.

Cependant, les performances de notre approche dépendent du résultat de la segmentation, mais aussi de celle de la reconnaissance des formes et de la reconnaissance du modèle de textures. Un autre problème auquel nous devons faire face, est l'insuffisance de la forme pour identifier certains objets. Par exemple la forme ne permet pas de reconnaître des concepts comme "ciel", "mer", "herbes", etc., problème auquel on peut probablement pallier en se basant sur la texture aussi pour l'identification de certains objets.

Néanmoins, les premiers résultats que nous avons obtenus sur la base de COREL restent satisfaisants et prometteurs, d'autant que le modèle que nous proposons reste ouvert à plusieurs améliorations possibles. Par exemple, nous envisageons dans un futur proche d'enrichir la description des objets en définissant la relation entre eux. Une amélioration possible serait de définir des prépositions de lieu entre les objets d'une image : sur, sous, dans, devant, à côté de, entre, etc. (Ex. chat sous la table) ; ou aussi les subordonnées comparative : plus grand que, plus petit que, etc.

Une autre utilisation très intéressante de notre approche serait la génération automatique d'ontologies pour la description sémantique des images.

## 6. Conclusion

Un système de recherche d'image adapté aux besoins des utilisateurs doit être capable d'abstraction, abstraction par rapport aux simples pixels dont sont constituées les images, c'est-à-dire capable d'extraire de la sémantique d'image. Cependant, le fossé entre les attributs de bas niveau et la construction des connaissances sémantiques est le principal obstacle dans la construction d'une sémantique fiable pour la recherche d'image. Dans ce papier nous avons proposé une approche permettant de découvrir des informations sémantiques à partir des traits de bas niveau d'une image. Notre approche s'intéresse à la description sémantique des objets d'une image donnée. Comme perspective à ce travail, nous proposons de finir d'implémenter le modèle proposé, puis de l'enrichir en décrivant les relations entre les objets d'une image.

Hichem Bannour

## 7. Bibliographie

- Bannour H., AyeB B., Hlaoua L., « Toward Content Based Image Retrieval : Global versus Local Image Description », CORIA 2009 : Sixième édition de la Conférence en Recherche d'Information et Applications, Belambra de la Presqu'île de Giens, France, May, 2009.
- Barnard K., Duygulu P., Freitas N., Forsyth D., Blei D., Jordan M. I., « Matching words and pictures », *Journal of Machine Learning Research (JMLR)*, vol. 3, p. 1107-1135, 2003.
- Canny J., « A Computational Approach to Edge Detection », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, n° 6, p. 679-698, 1986.
- Carneiro G., Chan A. B., Moreno P. J., Vasconcelos N., « Supervised learning of semantic classes for image annotation and retrieval », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, p. 2007, 2007.
- Carpenter G., Grossberg S., « A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine », *Computer Vision, Graphics, and Image Processing*, vol. 37, p. 54-115, 1987.
- Cheng S. C., Chou T. C., Yang C. L., Chang H. Y., « A semantic learning for content-based image retrieval using analytical hierarchy process », *Expert Systems with Applications*, vol. 28, n° 3, p. 495-505, April, 2005.
- Cox I. J., Miller M. L., Minka T. P., Papathornas T. V., Yianilos P. N., « The bayesian image retrieval system, pichunter : Theory, implementation, and psychophysical experiments », *IEEE Transactions on Image Processing*, 2000.
- Deselaers T., Features for Image Retrievals, Diploma thesis, RWTH AachenUniversity, Aachen, Germany, 2003.
- Flickner M., Sawhney H., Niblack W., Ashley J., Huang Q., Dom B., Gorkani M., Hafner J., Lee D., Petkovic D., Steele D., Yanker P., « Query by Image and Video Content : The QBIC System », *IEEE Computer Society*, vol. 28, n° 9, p. 23-32, 1995.
- Haralick R. M., « Statistical and structural approaches to texture », *Proc. IEEE*, vol. 67, p. 786-804, 1979.
- Lavrenko V., Manmatha R., Jeon J., « A model for learning the semantics of pictures », in *NIPS*, MIT Press, 2003.
- Mojsilovic A., « A computational model for color naming and describing color composition of images », *IEEE Transactions on Image Processing*, vol. 14, n° 5, p. 690 - 699, May, 2005.
- Ordonez C., Omiecinski E., « Discovering association rules based on image content », In *Proceedings of the IEEE Advances in Digital Libraries Conference*, Mai, 1999.
- Pal N., Pal K., Keller J., Bezdek J., « A Possibilistic Fuzzy c-Means Clustering Algorithm », *Fuzzy Systems, IEEE Transactions on*, vol. 13, n° 4, p. 517-530, Aug., 2005.
- Parsons O., Carpenter G. A., « ARTMAP neural networks for information fusion and data mining : map production and target recognition methodologies », *Neural Networks*, vol. 16, n° 7, p. 1075-1089, September, 2003.
- Rasiwasia N., Moreno P., Vasconcelos N., « Bridging the Gap : Query by Semantic Example », *Multimedia, IEEE Transactions on*, vol. 9, n° 5, p. 923-938, Aug., 2007.
- Smeulders A. W., Worring M., Santini S., Gupta A., Jain R., « Content-based image retrieval at the end of the early years », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, n° 12, p. 1349-1380, December, 2000.