
Annotation Semi-automatique de Grandes BD Images : Approche par Graphes de Voisinage

Hakim Hacid

*Université Lumière Lyon 2
Laboratoire ERIC - 5, avenue Pierre Mendès-France
69676 Bron cedex - France
hhacid@eric.univ-lyon2.fr*

RÉSUMÉ. L'annotation d'images constitue l'outil principal pour l'association d'une sémantique à une image. Dans cet article nous nous intéressons à l'annotation semi-automatique d'images. En effet, avec la grande masse de données gérées à travers le monde et surtout avec l'avènement du web, l'annotation manuelle de ces images est pratiquement impossible. Cet article présente un travail préliminaire sur une démarche basée sur les graphes de voisinage. La démarche que nous proposons offre, comme montré dans la section des expérimentations, des résultats d'annotation intéressants.

ABSTRACT. The annotation of images data is the main tool for associating semantics to an image. In this article we are interested in the semi-automatic annotation of images data. Indeed, with the great mass of data managed throughout the world and especially with the Web, the manual annotation of these images is almost impossible. This article presents a preliminary work on a neighbourhood graphs based annotation approach. The approach we are proposing offers, as the experiments section shows it, interesting annotation results.

MOTS-CLÉS : Bases de données images, Graphes de voisinage, Annotation, Indexation d'images

KEYWORDS: Images databases, Neighbourhood graphs, Annotation, Images indexing

1. Introduction

L'un des défis actuels dans les bases de données images est l'association d'une sémantique à une image. En effet, les méthodes de traitement d'images ne font qu'associer à chaque image un vecteur (ou des vecteurs) de caractéristiques calculées à la base des caractéristiques d'images dites de "bas niveau" (couleur, texture, etc.). L'interrogation d'une base de données d'images se fait alors en introduisant une image requête dans le système et la comparaison des caractéristiques ainsi calculées en utilisant une mesure de similarité [VEL 00]. Aucune sémantique n'est donc associée à ce processus.

L'une des manières possibles pour l'assignation d'une sémantique à une image est l'annotation. L'annotation des données multimédias est la tâche qui consiste à assigner, à chaque document multimédia ou à des parties du document multimédia, un mot clé ou une liste de mots clés permettant de décrire son contenu sémantique. Cette fonction peut être considérée comme une manière permettant de faire une sorte de correspondance entre les aspects visuels des données multimédias et leurs caractéristiques de bas niveau.

Dans cet article, nous nous intéressons à l'annotation semi-automatique d'images pour l'assignation d'une sémantique aux images. Nous considérons l'annotation semi-automatique car elle nécessite l'intervention de l'utilisateur pour valider les décisions du système. Le reste de cet article est organisé comme suit : la Section 2 présente un état de l'art sur l'annotation d'images. La Section 3 introduit notre démarche d'annotation d'images. Nous présentons les expérimentations dans la Section 4. Nous finirons par une conclusion et des perspectives dans la Section 5.

2. Etat de l'art

Il existe trois types d'annotation d'images : manuelle, semi-automatique et automatique. La première, comme sa dénomination l'indique, est effectuée manuellement par un humain chargé d'attribuer à chaque image un ensemble de mots clé. L'annotation automatique, quand à elle, est une tâche effectuée par une machine et vise à alléger la charge de l'utilisateur. Le premier type d'annotation augmente la précision et baisse la productivité tandis que le second type baisse la précision de la tâche et augmente sa productivité. Afin de faire un compromis entre ces deux tâches, leur combinaison est devenue nécessaire. C'est ce qui est connu sous le nom "*l'annotation semi-automatique*".

Il existe des méthodes qui appliquent un partitionnement ("clustering") des images et des textes associés aux images permettant ainsi de lier le texte et l'image [BAR 01]. Avec cette démarche, il est possible de prédire l'étiquette d'une nouvelle image en calculant des probabilités. Minka et Picard [PIC 95] ont proposé un système semi-automatique d'annotation d'images dans lequel l'utilisateur choisit les régions à annoter dans une image. Une propagation des annotations est effectuée en considérant les textures. Maron et al., [MAR 98] ont abordé l'annotation automatique mais en

utilisant un seul mot clé à la fois. Mori et al., [MOR 99] ont proposé un "modèle de co-occurrences" entre l'image et les mots clés afin de trouver les mots clés les plus pertinents pour une image. L'inconvénient de ce modèle est qu'il nécessite un grand échantillon d'apprentissage pour qu'il soit efficace. Dyugulu et al., [DUY 02] ont proposé un modèle, appelé "modèle de translation", qui est une amélioration du modèle de co-occurrence proposé par Mori et al., [MOR 99], en utilisant un algorithme d'apprentissage. Des modèles probabilistes tels que "Cross Media Relevance model" [JEO 03] et "Latent Semantic Analysis" [MON 03] ont été aussi proposés. Jia et Wang [LI 03] utilisent les chaînes de markov cachées bidimensionnelles pour annoter les images. Ces travaux, comme nous pouvons le constater, opèrent sur l'annotation locale (utilisent les objets).

L'inconvénient des approches décrites ci dessus est qu'elles dépendent fortement des techniques de segmentation et du résultat des techniques de partitionnement utilisées. Ceci rend ces méthodes sensibles au problème du passage à l'échelle. Notre travail, comme celui de Barnard et al. [BAR 01], se situe au niveau de l'annotation globale afin de minimiser l'impact des techniques de traitement d'images. Nous utilisons ensuite un modèle de prédiction (graphe de voisinage) pour annoter les images.

3. Contribution

Dans cette section, nous allons tenter de répondre à la question suivante : "Ayant un ensemble d'images annotées, de quelle façon faut-il procéder afin d'annoter une nouvelle image introduite sans annotations?". Pour cela, notre contribution se base sur l'utilisation de modèles issus de l'apprentissage automatique. En effet, comme c'est montré plus loin, nous utilisons dans la démarche que nous proposons les graphes de voisinages - le graphe des voisins relatifs plus précisément - pour accomplir cette tâche.

Formellement, considérons un ensemble Ω de n images $\Omega = \{I_1, I_2, \dots, I_n\}$. Chaque image est décrite par un ensemble d'attributs $\langle f_1, f_2, \dots, f_m \rangle$ et une liste de mots clé $W = w_1, w_2, \dots, w_k$ ($m \neq k$). Une image peut alors être décrite par un vecteur $I_i = \langle f_1, f_2, \dots, f_m, W_i \rangle$. Notons par ailleurs que chaque image peut avoir un nombre de mots clés différents par rapport aux autres images disponibles dans la base de données.

A partir de là, ayant une nouvelle image I_x non étiquetée ($I_x = \langle f_1, f_2, \dots, f_n \rangle$), il s'agit de trouver un modèle capable d'affecter à l'image I_x les étiquettes pouvant décrire au mieux son aspect sémantique. En d'autres termes, passer d'une représentation sous forme $I_x = \langle f_1, f_2, \dots, f_n \rangle$ à une représentation sous la forme $I_x = \langle f_1, f_2, \dots, f_n, W_k \rangle$.

L'approche que nous proposons se compose de deux niveaux : la structuration des données (indexation) et l'annotation effective.

– *Indexation multidimensionnelle*. Il s’agit ici de la première phase du processus d’annotation. Cette phase exploite les caractéristiques de bas niveau (couleur, texture, forme, etc.) extraites des images. Les images sont alors représentées dans un espace multidimensionnel R^p . L’objectif de cette phase, en plus de la rapidité de l’accès qui est l’objectif premier d’un index, est de pouvoir repérer facilement les voisins d’une image (les images ayant des caractéristiques de bas niveau similaires).

Concrètement, durant cette phase les images de la base de données sont traitées et transformées en un ensemble de vecteurs de caractéristiques de bas niveau. Chaque image est alors représentée comme un point dans R^d et le graphe est alors construit en respectant la propriété de voisinage correspondante. A ce niveau, le système peut répondre à des requêtes de type requête par l’exemple (*Query by Image Example*) comme décrit dans [HAC 05][HAC 06].

– *annotation effective*¹. Après l’indexation ”bas niveau” de la base d’images (construction du graphe à partir des caractéristiques de bas niveau), il s’agit à ce niveau de l’attribution effective des annotations à une nouvelle image inconnue au système. Le principe de base ici est ”l’héritage”. En effet, il s’agit de faire hériter à une image, après son insertion dans le graphe de voisinage préalablement construit, des annotations de ses voisins en calculant des scores pour chaque annotation potentielle. L’une des possibilités est la prise en compte de la distance mutuelle entre les individus. Dans ce cas, nous introduisons un poids pour les connections afin de favoriser les points les plus proches.

Par la suite, l’attribution des scores se fait alors avec une sommation des poids pour chaque annotation potentielle. Les annotations satisfaisant un seuil préfixé par l’utilisateur sont sélectionnées pour constituer une annotation possible à l’image en cours.

4. Expérimentations et résultats

Nous allons montrer, en utilisant une base de données d’images, l’intérêt de la démarche que nous avons proposé pour l’annotation semi-automatique des images². Pour cela, nous utilisons la base de données décrite dans [NEN 96]. Cette base de données contient 7200 images représentant 100 objets différents pris sous diverses vues. Pour réaliser nos expérimentations, nous avons annoté manuellement toutes les images avec des mots clés décrivant la nature de l’objet (que représente-t-il ?), la couleur de l’objet, la forme de l’objet ainsi que d’autres descriptions spécifiques aux objets. Par ailleurs, nous pré-traitons les images de la base de données en appliquant quelques algorithmes de traitement d’images pour extraire certaines caractéristiques. Nous employons particulièrement la couleur, quelques caractéristiques de texture et de forme. A la fin du processus de pré-traitement, chaque image est représentée par 25 caractéristiques.

1. A cause des contraintes de taille de l’article, nous n’avons pas pu détailler quelques formules mathématiques.

2. Nous intervenons ici uniquement dans la fixation du seuil de décision. Le système valide toutes les annotations dont le score est supérieur à 75%.

téristiques de bas niveau et 3 annotations textuelles en moyenne. Une fois l'espace de représentation constitué et les annotations assignées pour chaque individu, nous divisons l'ensemble des données en deux sous-ensembles :

– *Ensemble d'apprentissage* : Cet ensemble est constitué d'images tirées aléatoirement de l'ensemble des images. Nous prenons une proportion de 70% des images pour constituer cet ensemble d'apprentissage. Cet ensemble nous servira à construire le graphe de départ. Signalons que le graphe est construit exclusivement avec les caractéristiques de bas niveau extraites des images, les annotations ne sont pas utilisées durant cette phase.

– *Ensemble de test* : Cet ensemble est constitué de 30% des individus de la base de données d'images (les individus restants), il est destiné à être annoté en utilisant notre approche. Pour annoter une image appartenant à cet ensemble, nous l'introduisons comme image requête dans le graphe et elle est positionnée (son voisinage est repéré et est mis à jour) en utilisant ses caractéristiques de bas niveau. Les annotations de départ de ces images nous servent à comparer les annotations affectées par le système et celles affectées par l'utilisateur afin d'évaluer le comportement général du graphe.

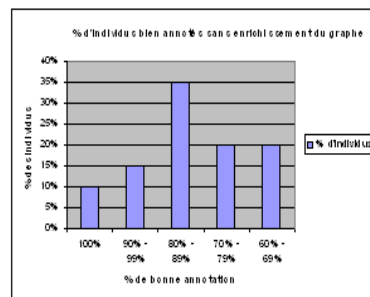


Figure 1. Statistiques de bonne annotation

Le graphique de la Figure 1 illustre le pourcentage d'individus appartenant à l'ensemble de test en fonction du taux de bonne annotation. La première observation est le fait que la majorité des individus ont un taux de bonne annotation qui varie entre 80% et 90% ce qui constitue à ce stade un bon résultat d'annotation.

5. Conclusion et perspectives

L'annotation constitue l'outil principal pour l'association d'une sémantique à une image. Dans cet article nous nous sommes intéressés à l'annotation semi-automatique d'images. En effet, avec la grande masse de données gérées à travers le monde et surtout avec l'avènement du Web, l'annotation manuelle de ces images est quasiment impossible. Nous avons alors proposé une démarche basée sur les graphes de voisinage offrant plusieurs possibilités : *interrogation par le contenu*, *interrogation par des mots clés* et *l'annotation avec des techniques de vote*. La démarche que nous avons

proposé offre des résultats d'annotation très intéressants tout en satisfaisant le critère du passage à l'échelle qui est un point très crucial dans un contexte où la masse de données est très importante.

En tant que perspectives, nous projetons tout d'abord d'appliquer notre méthode sur une plus grande base de données d'images plus complexes que celle utilisées dans notre travail et de comparer nos résultats avec les résultats d'autres méthodes telles que celles citées dans l'état de l'art. L'absence de données images annotées nous empêche, pour le moment, d'effectuer ces comparaisons. Par ailleurs, nous travaillons sur l'intégration des connaissances de l'utilisateur dans le processus d'annotation (*Relevance Feedback*) afin d'automatiser tout le processus.

6. Bibliographie

- [BAR 01] BARNARD K., FORSYTH D. A., « Learning the Semantics of Words and Pictures. », *ICCV*, 2001, p. 408-415.
- [DUY 02] DUYGULU P., BARNARD K., DE FREITAS J. F. G., FORSYTH D. A., « Object Recognition as Machine Translation : Learning a Lexicon for a Fixed Image Vocabulary. », *ECCV (4)*, 2002, p. 97-112.
- [HAC 05] HACID H., ZIGHED D., « Neighborhood Graphs for Image Databases Indexing and Content-Based Retrieval », *First IEEE International Workshop on Mining Complex Data (IEEE MCD05), Texas, USA*, 2005.
- [HAC 06] HACID H., ZIGHED D., « Content-Based Image Retrieval in Large Image Databases », *IEEE International Conference on Granular Computing (GrC 2006), Atlanta, USA*, May 2006.
- [JEO 03] JEON J., LAVRENKO V., MANMATHA R., « Automatic image annotation and retrieval using cross-media relevance models. », *SIGIR*, 2003, p. 119-126.
- [LI 03] LI J., WANG J. Z., « Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach. », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, n° 9, 2003, p. 1075-1088.
- [MAR 98] MARON O., RATAN A. L., « Multiple-Instance Learning for Natural Scene Classification. », *ICML*, 1998, p. 341-349.
- [MON 03] MONAY F., GATICA-PEREZ D., « On image auto-annotation with latent space models. », *ACM Multimedia*, 2003, p. 275-278.
- [MOR 99] MORI YASUhide T. H., RYUICHI O., « Image-to-Word transformation Based on dividing and vector quantizing image with words », *In MISRM'99, First international workshop on multimedia intelligent storage and retrieval management.*, , 1999.
- [NEN 96] NENE S. A., NAYAR S. K., MURASE H., « Columbia Object Image Library (COIL-100) », rapport, February 1996, Technical Report CUCS-006-96.
- [PIC 95] PICARD R. W., MINKA T. P., « Vision Texture for Annotation. », *Multimedia Syst.*, vol. 3, n° 1, 1995, p. 3-14.
- [VEL 00] VELTKAMP R. C., TANASE M., « Content-based image retrieval systems : A survey », rapport n° UU-CS-2000-34, 2000, Department of Computing Science, Utrecht University.