

PROPOSITION DE STAGE

Campagne 2012

(*) Supprimer la mention inutile



Laboratoire L3i

Sujet Stage :

Recherche d'images par le contenu : Réduction de la complexité des techniques de mise en correspondance de « sacs de mots visuels » à l'aide de générateurs minimaux

Résumé du travail proposé :

Dès qu'il s'agit de traiter des documents multimédia (image, sons, vidéo...), les moteurs de recherche traditionnels voient leurs performances s'effondrer car plutôt que de traiter réellement le contenu des documents, ils s'appuient sur la recherche textuelle concernant les éléments contextuels du document recherché. Les systèmes de recherche d'images par le contenu (CBIR) ont été développés dans l'idée d'indexer directement les images [Veltkamp,2002] en construisant une représentation de l'image basée sur son contenu et de mettre en correspondance cette représentation et la sémantique associée. Les techniques actuelles s'appuient sur l'extraction de nombreuses informations locales de type SIFT [Lowe,1999] ou SURF [Bay,2008] qui génèrent en moyenne plus de 1000 points par image. Chacun de ces points est ensuite caractérisé par un dictionnaire de mots visuels (pattern visuel caractéristique d'une portion de l'image). Chaque image est ainsi représentée par un « sac de mots visuels ». Les techniques d'indexation consistent à comparer chaque « sac de mot » d'une image avec l'ensemble des autres. Toutefois, le coût computationnel de telles techniques les rendent difficiles à utiliser dans un cadre où de nombreuses images sont disponibles. De leur côté, les techniques d'extraction de règles d'association visent à extraire des implications entre caractéristiques. Récemment, des méthodes d'extraction de « bases de règles informatives » ont été introduites dans le but de minimiser les règles extraites à partir du treillis des fermés [IGB, 04]. Composées de « générateurs minimaux » en prémisses, ces règles forment des bases génériques en lien avec la logique propositionnelle. Dans ce travail, on se propose d'adapter ces techniques d'extraction de générateurs minimaux aux mots visuels et de les intégrer dans un système de recherche d'image par le contenu avec retour de pertinence.

Mots clés :

Recherche d'images par le contenu, CBIR, treillis, sacs de mots

Informations complémentaires :

Encadrant(s) : Karell BERTET et Arnaud REVEL

Axe thématique : « Données complexes, Images et Documents »

Axe stratégique : « Pertinence Interaction-Contenu »

Cadre de coopération :

Date de début du stage :

Durée du contrat :

Contexte de l'étude:

L'essor d'Internet a fait se multiplier le nombre et la nature des sources d'informations disponibles. Les moteurs de recherche ont permis d'offrir un accès facile à des données fortement structurées. Cependant, dès qu'il s'agit de traiter des documents multimédia (image, sons, vidéo...), les moteurs de recherche traditionnels voient leurs performances s'effondrer. En effet, plutôt que de traiter réellement le contenu des documents, les techniques de recherche s'appuient sur la recherche textuelle concernant les éléments contextuels du document recherché (page Web associée, méta-informations du fichier multimédia...). Or ce texte associé aux documents peut avoir une sémantique très différente du contenu réel du document ou de ce que l'on peut en inférer. Les systèmes de recherche d'images par le contenu (« Content-based Image Retrieval », CBIR) ont été développés pour répondre à cette problématique [Veltkamp,2002]. L'idée générale est de construire une représentation de l'image basée sur son contenu et de mettre en correspondance cette représentation et la sémantique associée à l'image grâce à des techniques d'apprentissage automatique. Les systèmes à bouclage de pertinence sont en particulier les méthodes ayant donné les meilleurs résultats [Wood, 1998; Huang,2001].

Description du sujet :

Les techniques actuelles s'appuient essentiellement sur l'extraction de nombreuses informations locales obtenues grâce à des détecteurs spécialisés de type SIFT [Lowe,1999] ou SURF [Bay,2008] qui génèrent en moyenne plus de 1000 points par image. Chacun de ces points est ensuite caractérisé et un dictionnaire de mots visuels (pattern visuel caractéristique d'une portion de l'image) est construit. Chaque image est ainsi représentée par un « sac de mots visuels ». Les techniques d'indexation consistent dans un premier temps à apprendre à classifier les images en fonction de ces sacs de mots puis à comparer chaque « sac de mot » d'une image avec l'ensemble des autres images afin d'en inférer la classe. Les techniques les plus courantes sont celles basées sur des classifieurs SVM [Cortes,1995] utilisant des noyaux composés (noyaux mineurs et majeur). Toutefois, le coût computationnel de telles techniques les rendent difficiles à utiliser dans un cadre où de nombreuses images sont disponibles.

De leur côté, les techniques d'extraction de règles d'association visent à extraire des implications entre caractéristiques sous forme de règles de type « prémisses -> conclusion », où prémisses et conclusion sont des ensembles de caractéristiques. Bien qu'initialement introduites dans un objectif de description, les règles d'association s'utilisent également dans un cadre prédictif où la conclusion est un attribut de classe à prédire. Chaque classe est alors décrite par un ensemble de prémisses qui peut s'apparenter à un sac de mots. A-Priori, algorithme fondateur d'extraction des règles [Agrawal et al. 94], cherche à extraire les règles de façon exhaustive, induisant un risque d'explosion combinatoire inhérent au nombre potentiellement élevé des règles extraites. Par la

suite, des approches visant à réduire le nombre de règles extraites tout en garantissant leur informativité ont été proposées. Ces approches exploitent des résultats issus de la théorie des treillis dans le but de réduire les combinaisons de caractéristiques considérées aux seules « fermetures » d'un treillis des fermés, et qui s'avèrent être suffisamment informatives [Pasquier 2000]. Plus récemment, des méthodes d'extraction de « bases de règles informatives » ont été introduites dans le but de minimiser les règles extraites à partir du treillis des fermés [IGB, 04]. Composées de « générateurs minimaux » en prémisse, ces règles forment des bases génériques en lien avec la logique propositionnelle.

Dans ce travail, on se propose d'adapter ces techniques d'extraction de générateurs minimaux aux mots visuels et de les intégrer dans un système de recherche d'image par le contenu avec retour de pertinence.

Prérequis et contraintes particulières :

Connaissances du C++/Java ou bon niveau en programmation

Références bibliographiques :

[Veltkamp, 2002] R. Veltkamp, "Content-based image retrieval system: A survey," University of Utrecht, Tech. Rep., 2002.

[Wood,1998] M. Wood, N. Campbell, and B. Thomas, "Iterative refinement by relevance feedback in content-based digital image retrieval," in ACM Multimedia 98, Bristol, UK, September 1998, pp. 13–20.

[Huang,2001] T. Huang and X. Zhou, "Image retrieval with relevance feedback: From heuristic weight adjustment to optimal learning methods," in International Conference in Image Processing (ICIP'01), vol. 3, Thessaloniki, Greece, October 2001, pp. 2–5.

[Lowe,1999] David G. Lowe, « Object recognition from local scale-invariant features », dans Proceedings of the International Conference on Computer Vision, vol. 2, 1999, p. 1150–1157.

[Bay,2008] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008

[Cortes,1995] Corinna Cortes and V. Vapnik, "Support-Vector Networks", Machine Learning, 20, 1995

[Agrawal,1994] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.

Contacts – liens :

Email : {kbertet,arevel}@univ-lr.fr

Lien vers le fichier de description : (PDF)

Présentation libre :

Cette partie ne sera visible QUE sur le fichier PDF dont le lien est présenté ci-dessus. Cette partie permet de présenter certains aspects non développés dans la partie WEB qui reste nécessairement globale.

L'essor d'Internet a fait se multiplier le nombre et la nature des sources d'informations disponibles. Les moteurs de recherche ont permis d'offrir un accès facile à des données fortement structurées. Cependant, dès qu'il s'agit de traiter des documents multimédia (image, sons, vidéo...), les moteurs de recherche traditionnels voient leurs performances s'effondrer. En effet, plutôt que de traiter réellement le contenu des documents, les techniques de recherche s'appuient sur la recherche textuelle concernant les éléments contextuels du document recherché (page Web associée, méta-informations du fichier multimédia...). Or ce texte associé aux documents peut avoir une sémantique très différente du contenu réel du document ou de ce que l'on peut en inférer. Les systèmes de recherche d'images par le contenu (« Content-based Image Retrieval », CBIR) ont été développés pour répondre à cette problématique [Veltkamp,2002]. L'idée générale est de construire une représentation de l'image basée sur son contenu et de mettre en correspondance cette représentation et la sémantique associée à l'image grâce à des techniques d'apprentissage automatique. Les systèmes à bouclage de pertinence sont en particulier les méthodes ayant donné les meilleurs résultats [Wood, 1998; Huang,2001].

Les techniques actuelles s'appuient essentiellement sur l'extraction de nombreuses informations locales obtenues grâce à des détecteurs spécialisés de type SIFT [Lowe,1999] ou SURF [Bay,2008] qui génèrent en moyenne plus de 1000 points par image. Chacun de ces points est ensuite caractérisé et un dictionnaire de mots visuels (pattern visuel caractéristique d'une portion de l'image) est construit. Chaque image est ainsi représentée par un « sac de mots visuels ». Les techniques d'indexation consistent dans un premier temps à apprendre à classifier les images en fonction de ces sacs de mots puis à comparer chaque « sac de mot » d'une image avec l'ensemble des autres images afin d'en inférer la classe. Les techniques les plus courantes sont celles basées sur des classifieurs SVM [Cortes,1995] utilisant des noyaux composés (noyaux mineurs et majeur). Toutefois, le coût computationnel de telles techniques les rendent difficiles à utiliser dans un cadre où de nombreuses images sont disponibles.

De leur côté, les techniques d'extraction de règles d'association visent à extraire des implications entre caractéristiques sous forme de règles de type « prémisses -> conclusion », où prémisses et conclusion sont des ensembles de caractéristiques. Bien qu'initialement introduites dans un objectif de description, les règles d'association s'utilisent également dans un cadre prédictif où la conclusion est un attribut de classe à prédire. Chaque classe est alors décrite par un ensemble de prémisses qui peut s'apparenter à un sac de mots. A-Priori, algorithme fondateur d'extraction des règles [Agrawal et al. 94], cherche à extraire les règles de façon exhaustive, induisant un risque d'explosion combinatoire inhérent au nombre potentiellement élevé des règles extraites. Par la suite, des approches visant à réduire le nombre de règles extraites tout en garantissant leur informativité ont été proposées. Ces approches exploitent des résultats issus de la théorie des treillis dans le but de réduire les combinaisons de caractéristiques considérées aux seules

« fermetures » d'un treillis des fermés, et qui s'avèrent être suffisamment informatives [Pasquier 2000]. Plus récemment, des méthodes d'extraction de « bases de règles informatives » ont été introduites dans le but de minimiser les règles extraites à partir du treillis des fermés [IGB, 04]. Composées de « générateurs minimaux » en prémisse, ces règles forment des bases génériques en lien avec la logique propositionnelle.

Dans ce travail, on se propose d'adapter ces techniques d'extraction de générateurs minimaux aux mots visuels et de les intégrer dans un système de recherche d'image par le contenu avec retour de pertinence.