

Résumé

L'analyse et la comparaison intelligentes d'images sont parmi les sujets suscitant le plus d'intérêt dans les milieux académiques autant qu'industriels. Décrire et comparer automatiquement les images est en effet un enjeu critique pour le plein développement de la «société de l'information».

Les moteurs de recherche fonctionnant sur le texte ont prouvé leur utilité de façon éclatante mais à l'heure actuelle il n'existe aucun système équivalent fonctionnant uniquement sur les images. Une explication possible est que nous ne disposons pas de langage permettant de décrire les images et que les comparaisons pertinentes sont ainsi beaucoup plus difficiles que dans le cas du texte. Cependant, le cas du texte nous montre qu'il n'est pas nécessaire que les machines comprennent ce qu'elles analysent pour renvoyer des résultats pertinents. Des méthodes simples d'analyse syntaxique associées à des règles de composition suffisent à piloter des moteurs de recherche d'une grande efficacité.

Pour permettre à des machines de simuler l'interprétation des images, il faudrait donc créer des descripteurs faisant office de mots et des règles pour les regrouper, ce qui permettrait de comparer des scènes comme on compare des phrases. On dispose d'ores et déjà de nombreuses méthodes pour détecter automatiquement de petits objets et des régions dans des images, par leur couleur commune, leur mouvement identique, etc. Poursuivant l'analogie, on pourrait comparer ces petits objets à des syllabes. La difficulté consiste à les grouper en mots, puis en phrases et comparer celles-ci, tout en étant robuste face aux perturbations.

Pour ce faire, nous utilisons des graphes pour stocker ces objets et leurs relations. Ces relations peuvent être de voisinage ou d'inclusion, ce qui conduit les graphes à être respectivement des graphes plans ou des arbres. Nous verrons ainsi plusieurs méthodes permettant de construire l'un ou l'autre type de représentation, ainsi que leurs avantages et inconvénients.

Dans une première étape, nous avons utilisé les algorithmes d'appariement de graphes développés par Cristina Gomila à la fin de sa thèse au CMM (1998-2001). Profitant du projet européen MASCOT étudiant l'utilisation de «métadonnées» pour faciliter le codage vidéo, nous avons étudié en détail les forces et faiblesses de cette approche. Nous avons d'abord testé le remplacement de l'algorithme au cœur de l'appariement de graphes. Nous avons obtenu une légère amélioration de la stabilité et également de meilleurs temps de calcul. Puis nous avons cherché à améliorer notre robustesse face aux variations de segmentation en utilisant une projection dans le domaine spectral. Malgré de bons résultats sur des images simples, nos essais sur des images plus difficiles n'ont pas été couronnés de succès. Pour pallier cette fragilité dès que les graphes ne sont plus similaires, nous avons préféré revenir à notre matériau source, les images.

La seconde étape de ce travail a porté sur le développement de techniques basées sur l'image pour réduire la sensibilité de nos algorithmes de segmentation au bruit et aux petites variations. Pour ce faire, nous avons développé une classe d'opérateurs de filtrage adaptatifs, les «amibes morphologiques», extrêmement efficaces pour réduire le bruit dans les images. Par ailleurs, nous avons également développé un opérateur de gradient couleur robuste permettant de mieux détecter les contours

dans les images bruitées. Ces deux opérateurs ont amélioré de façon parfois impressionnante la stabilité de nos modélisations, puis de nos graphes et donc des résultats globaux.

L'étape suivante dans ce travail a porté sur le développement de modélisations d'objets indépendamment du reste de l'image. La motivation derrière cette approche est de considérer que, dans certains scénarios, le contenu de l'image, hors de certains objets bien définis, n'est pas informatif. Il faut donc analyser directement et de la façon la plus précise possible les objets eux-mêmes. Nous avons dans un premier temps supposé que les segmentations des objets étaient connues, afin de nous concentrer sur le calcul d'une signature robuste de chaque objet. Pour l'obtenir, nous avons modifié un algorithme de ligne de partage des eaux pour effectuer une resegmentation «top-down» d'un espace d'échelle morphologique basé sur des nivellements. Ceci a donné lieu à une nouvelle modélisation robuste utilisant des arbres de régions imbriquées. Nous avons également développé une distance entre ces arbres et nous l'avons testée sur une base d'images classique dans le domaine de l'indexation.

La dernière étape est centrée sur l'aspect applicatif. En premier lieu en comparant les différentes approches présentées dans ce travail, notamment aux niveaux de leur robustesse et de leur vitesse d'exécution.

Enfin, nous avons cherché la meilleure combinaison de techniques pour concevoir une application de vidéosurveillance. En particulier, nous avons développé des techniques rapides et robustes de segmentation dans le cadre du projet PS26-27 «Environnement Intelligent» en collaboration avec ST Microelectronics et le groupe ORION de l'INRIA. Ce projet visait à construire un démonstrateur de technologies de vidéosurveillance appliquées à la détection d'accidents dans les cadres domestique et hospitalier. Notre part du travail consistait à la mise au point d'algorithmes de détection de silhouettes en mouvement dans des séquences vidéo.

Ainsi, en couplant ces techniques à nos descripteurs d'objets par arbres, nous avons pu définir des signatures robustes de personnes, qui pourront être utilisées avec un grande efficacité dans des systèmes automatisés de vidéosurveillance.

Mots-clefs : segmentation, espaces d'échelle morphologiques, graphes, arbres, réduction de bruit, amibes morphologiques