

Les algorithmes de recherche de visages

Plutôt que de définir explicitement les éléments qui constituent un visage, des chercheurs de l'Université Carnegie Mellon ont «entraîné» un réseau de neurones à l'aide d'une grande variété d'images légendées représentant des visages et d'autres images sans visages. Finalement, le réseau a appris à distinguer les yeux, le nez et la bouche. Une interface avec le réseau Internet permet à tout utilisateur de soumettre des images au programme. Un échantillon des centaines d'images où celui-ci a cru reconnaître des visages est donné ci-contre.



images trouvées étaient des planches à voile, mais le système afficha aussi d'autres images sans rapport avec celles-ci, et il ne trouva pas certaines images de planche à voile qui étaient pourtant dans la base de données. En matière de recherche d'informations, un compromis entre précision et échec est inévitable : si l'on impose une définition trop stricte, on ne trouve que les images du type cherché, mais certaines images cherchées sont omises ; et si l'on assouplit la définition, toutes les images cherchées sont trouvées, mais d'autres images sans intérêt font partie du lot. Ainsi, un puissant moteur de recherche par motifs retrouve assez bien des images contenant des objets particuliers, à condition que l'utilisateur convertisse mentalement la recherche d'un objet en recherche de motifs.

Dans l'état actuel de notre programme, l'utilisateur peut chercher directement un petit nombre d'objets : des chevaux, des bateaux à voile ou des personnages nus. Pour spécifier les motifs, il indique les proportions des différentes couleurs qu'il pense trouver dans une image, ainsi que le nombre et la taille des régions de couleurs devant apparaître. Il peut aussi préciser d'autres caractéristiques dérivées du contenu de l'image, par exemple si une image contient ou non un horizon. Enfin, il peut rechercher des «métadonnées», c'est-à-dire des informations additionnelles stockées avec l'image, comme sa légende, le nom du photographe ou la date de la photographie.

Les écueils de la reconnaissance d'objets

Quand on donne en outre des connaissances à l'ordinateur, ce dernier peut déduire, d'après les motifs présents dans l'image, quels objets s'y trouvent. L'identification de chevaux, de personnes ou de n'importe quel objet dans une image est une discipline nommée reconnaissance d'objets. Au cœur des recherches sur les systèmes de vision depuis plusieurs décennies, elle est étudiée aussi bien par les laboratoires publics que par les laboratoires privés, par les civils et par les militaires. Cependant, les techniques actuelles ne s'appliquent qu'aux images contenant un petit nombre d'objets dont les formes sont précisément connues et peuvent être vues selon un domaine limité d'angles de vue. Elles ne reconnaissent même pas des objets aussi «ordinaires» que des personnes ou des chevaux. Les objets ordinaires présentent une telle variété de tailles, de volumes ou de marques apparentes qu'ils déroutent les programmes de vision automatisée : pourtant une personne reste une personne, quels que soient sa taille, son poids ou son habillement.

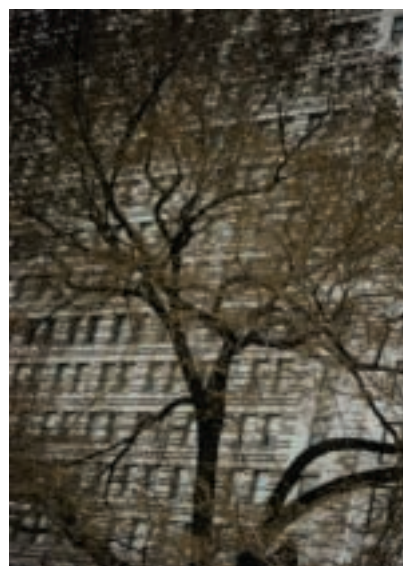
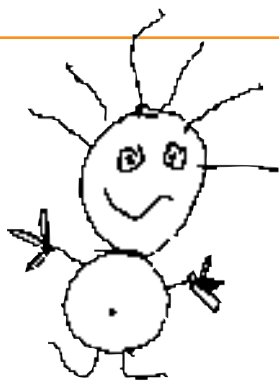
Comment, en outre, «extraire» les objets de l'arrière-plan de l'image ? C'est un problème difficile, même, dans certains cas, pour un observateur humain. Considérons un léopard dans l'ombre de la jungle. Afin de tenter une identification, le programme doit déterminer quelles parties de l'image sont associées, trouver que les taches appartiennent au léopard, les feuilles aux arbres, etc.

Autrement dit, le programme doit regrouper les motifs en objets.

Ce problème du regroupement perceptif a été beaucoup étudié par l'école gestaltiste. Les psychophysiciens ont mis en évidence plusieurs facteurs qui aident le regroupement. La similarité de couleur ou de texture est très efficace : nous formons facilement des groupes d'après des parties unicolores, telle une tache rouge, ou de texture uniforme, tel un tissu écossais. Des repérages plus complexes ont également lieu : les régions d'une image qui sont symétriques par rapport à un axe sont considérées comme la projection à deux dimensions d'un objet doué d'une symétrie axiale, tel un vase.

Depuis des années, les spécialistes de vision automatisée cherchent à traduire de telles règles empiriques en algorithmes efficaces. Serge Belongie, C. Carson, Hayit Greenspan et l'un de nous (J. Malik) ont mis au point un programme qui, bien que très rudimentaire par rapport à la vision humaine, décompose l'image en petits ensembles de régions de même couleur ou de même texture. Chaque «bloc» cohérent est caractérisé par sa position, sa forme, sa couleur et sa texture. Ainsi les images sont des ensembles de blocs.

À l'aide de cette décomposition, nous pouvons maintenant trouver des images qui représentent des avions sur fond de ciel en cherchant une tache grise à l'intérieur d'une zone bleue ; nous trouvons aussi des photographies



De gauche à droite et haut en bas : H. Fujii, K. Moeller, K.-P. Zauner, P. Quattrough, De Vinci, K.-P. Zauner, S.J. Brown, M. Locke, Carnegie Mellon University, H.K. Konishi



de tigres en recherchant un bloc jaune rayé de noir et un bloc ayant la couleur et la texture de l'herbe, typique de l'habitat du tigre. Pour chaque caractéristique des blocs (couleur, texture, forme ou position), l'utilisateur choisit le degré de précision qu'il désire.

La représentation des objets comme des ensembles de blocs se prête bien aux techniques d'apprentissage automatique. Ainsi on peut éviter à l'utilisateur d'identifier les aspects les plus caractéristiques d'une scène contenant des tigres ou des avions. Nous avons utilisé des techniques d'apprentissage statistique, afin de réaliser un logiciel de classement statistique auquel nous avons présenté des photographies de plusieurs catégories visuelles (comme des scènes comportant des avions ou des tigres), et avons ainsi entraîné le logiciel à classer de nouvelles photographies dans ces catégories.

Aujourd'hui, nos programmes classent des scènes en analysant uniquement la couleur et la texture ; avec des informations appropriées sur la forme des blocs, ils distingueront les ensembles de blocs qui constituent des objets et ceux qui n'en forment pas. L'apprentissage fournit un cadre naturel pour traiter les variations non pertinentes au sein d'une classe d'objets, parce qu'un logiciel de classement peut s'ajuster de façon à ignorer ces variations (le ciel peut avoir diverses nuances de bleu, les nuages ont des formes toujours changeantes).

L'apprentissage automatique

Takeo Kanade et ses collègues de l'Université Carnegie Mellon ont mis au point un logiciel de reconnaissance de visages qui applique les techniques d'apprentissage automatique. Comme les programmes classiques détectent difficilement les yeux, les nez et les bouches individuellement, T. Kanade et ses collègues ont entraîné un logiciel du type réseau de neurones à détecter la présence simultanée de tous ces éléments dans une configuration adéquate. Ils ont aussi associé cette reconnaissance de forme à une reconnaissance de sons pour analyser des extraits de films. Un ordinateur peut ainsi apprendre à reconnaître un discours sur une bande-son de journal télévisé, par exemple, et associer la mention du nom d'une célébrité avec le visage détecté au même moment dans les données vidéo.

Si la couleur et la texture aident à découper l'image en blocs, plusieurs caractéristiques supplémentaires permettent le difficile regroupement des blocs en objet. Tout d'abord, beaucoup d'objets sont composés de parties aux formes géométriques régulières ; de surcroît, ces formes tridimensionnelles de base se présentent comme des blocs

d'images de forme simple : par exemple, un cylindre apparaît presque toujours comme une région rectangulaire, aisément identifiable par le programme.

Lorsqu'il existe des relations géométriques entre les blocs, de la même façon que l'amplitude de rotation des articulations chez l'homme et chez l'animal limite les attitudes possibles entre les différentes parties de leur corps, on peut conclure à l'appartenance de deux blocs au même objet. En bref, l'analyse de chaque bloc engendre des hypothèses sur son identité, et la répétition de regroupements conduit finalement à l'analyse des images. La théorie de l'apprentissage statistique fournit des mécanismes grâce auxquels on détermine quels regroupements sont possibles. Lorsqu'un groupe suffisamment grand a été ainsi assemblé, l'objet est considéré comme identifié.

Margaret Fleck, de l'Université de l'Iowa, et l'un de nous (D. Forsyth) ont utilisé ces observations pour construire deux programmes qui reconnaissent des objets particuliers parmi de nombreuses images. Le premier programme trouve les images qui contiennent des êtres humains. Notre version actuelle ne fonctionne que pour des gens qui portent peu ou pas de vêtements ; les couleurs et les textures de la peau sont très limitées, comparées à celles des vêtements. Le programme divise d'abord l'image en régions correspondant à la peau (ce qui peut être fait de façon assez précise) et rejette les images comportant peu ou pas de peau.