

d'hui couramment utilisé par plusieurs facteurs pour prévoir la place et la dimension des trous. Nous avons aussi, en réalisant des calculs *a priori*, proposé des innovations que les facteurs n'auraient jamais imaginées : des tubes placés près de l'embouchure d'une flûte ou d'une clarinette permettent de jouer des quarts de tons, fréquents en musique contemporaine, sans doubler le nombre de trous.

Les calculs qui ont produit ces résultats ne prennent toutefois en compte que l'instrument seul, sans l'instrumentiste, et leur capacité prédictive ne concerne que la justesse des notes. En simulant les oscillations elles-mêmes, nous commençons à prévoir aussi la «qualité» des notes en termes de timbre et de facilité d'émission.

Au-delà, la compréhension du fonctionnement des instruments conduit vers une nouvelle synthèse musicale : on ne simule plus le son, mais la source du son. Les synthétiseurs sont commandés par des capteurs artificiels, donc par un véritable instrumentiste, ce qui produit des sons musicalement réalistes, c'est-à-dire évoluant avec des causes perçues comme «physiques». Là, ils se rapprochent du musicien, qui peut intervenir finement sur son instrument et réagir immédiatement sur le résultat obtenu. C'est cette interactivité et les micro-variations de justesse, de timbre, de phrasé qui permettent d'exprimer les émotions les plus subtiles.

Jean Pierre DALMONT maître de conférence, Joël GILBERT, chargé de recherche, et Jean KERGMARD, directeur de recherche, travaillent au Laboratoire d'acoustique (UMR CNRS 6613) de l'Institut d'acoustique et de mécanique de l'Université du Maine, au Mans.

Les instruments de l'orchestre, Pour la Science, 1995.

H. BOUASSE, *Instruments à vent*, Librairie Delagrave, Paris, 1986.

N.H. FLETCHER et T.D. ROSSING, *The Physics of Musical Instruments*, Springer Verlag, New York, 1990.

Mechanics of Musical Instruments, sous la direction de A. Hirschberg, J. Kergomard et G. Weinreich, CISM Courses and Lectures No.355, International Center for Mechanical Sciences, Springer-Verlag, Vienne-New York, 1996.

C. NEDERVEEN, *Acoustical Aspects of Woodwind Instruments*, Northern Illinois University Press, De Kalb, Illinois, 1997.



Pour la recherche française, l'Amazonie n'est ni une terre inconnue, ni une mode passagère : la coopération scientifique franco-brésilienne s'est déployée en Amazonie depuis le début des années 80, et a obtenu des résultats significatifs, tant pour la connaissance de l'environnement que sur le type de développement - ou de non-développement - qui s'y produit.

Pour contribuer au «programme-pilote de préservation des forêts tropicales du Brésil», la France a choisi de favoriser le renforcement du partenariat ancien et précieux qui s'est progressivement construit entre sa recherche et la recherche brésilienne. Ce livre tire les premiers enseignements de l'effort spécial de coopération qui a été mené dans ce cadre.

code 1532 - 208 pages
Prix : 130F



Voir bon de commande p. 98

La recherche d'images numériques

DAVID FORSYTH • JITENDRA MALIK • ROBERT WILENSKY

Les programmes de recherche d'images repéreront les éléments spécifiques d'une personne, d'un lieu ou d'un objet dans les archives photographiques.

Le réseau Internet et les bibliothèques numériques qu'il relie contiennent une profusion d'informations, mais la recherche d'une information précise et, en particulier, la recherche des images restent difficiles. Aujourd'hui, un utilisateur qui cherche une image aussi simple qu'un cheval au bord de l'eau ne trouve

son bonheur que si les images sur ce sujet ont été correctement légendées.

L'expérience acquise dans les principaux centres d'archives montre qu'il est impossible de prévoir toutes les demandes. En outre, la réponse à chaque nouvelle question impose une nouvelle recherche complète dans la base de données. Si, ayant trouvé l'image d'un

cheval au bord de l'eau, on demande des images qui comprennent un cheval et de l'herbe, le réexamen de la collection entière est nécessaire.

Les utilisateurs cherchent des images dans des contextes variés et avec des objectifs différents : l'un demandera aux archives d'un musée d'art combien de fois une couleur par-



1. LA RECHERCHE DE PHOTOGRAPHIES à l'aide de programmes de recherche d'images est difficile, car le mécanisme de la reconnaissance d'un objet reste mal compris. Pour trouver une photographie de fleurs pourpres dans une base de données, un

utilisateur a utilisé un programme prototype de recherche sur des images et demandé les images contenant de petites taches violacées ; quelques-uns des résultats de cette recherche sont montrés dans la rangée supérieure. Beaucoup d'images trouvées

ticulière a été utilisée pour peindre la peau dans les portraits, un autre interrogera une base de données d'images satellite pour estimer la prochaine récolte de blé ; un autre encore cherchera dans les réserves d'agences photographiques une représentation d'un coucher de soleil pour la couverture d'une brochure touristique ; un autre enfin recherchera une série d'images des rivières du Midi afin de comparer l'impact des inondations récentes par rapport à celles des années précédentes...

On ne satisfera ces utilisateurs du réseau que si l'on met au point des méthodes automatiques d'analyse des images. L'un des moyens les plus simples de traiter ces diverses demandes consiste à utiliser un programme qui discerne si une image donnée contient un objet spécifique, tel un cheval ou une pièce d'eau, quels que soient sa couleur, sa taille, son apparence ou l'angle de vue. Un tel programme n'existe pas encore, mais les informaticiens utilisent des méthodes variées pour le créer : ici

la psychophysiologie de la vision s'allie au raisonnement automatique pour améliorer l'analyse des caractéristiques des images.

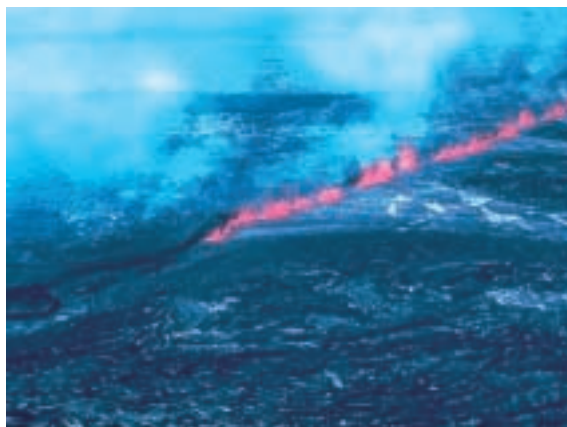
Les programmes mis au point analysent le contenu des images de plusieurs manières. Certains recherchent des images qui s'apparentent à un modèle préenregistré. D'autres se fondent sur l'aspect général : pour identifier un champ de fleurs sauvages, ils cherchent des images qui contiennent un grand nombre de points jaunes ; pour trouver un feu, un coucher de soleil ou une voiture de sport, ils cherchent des images dont une région centrale est rouge vif. Enfin, d'autres programmes cherchent un objet à l'identité définie, comme de l'eau ou un cheval. Tous ces programmes sont ainsi des mises en œuvre de trois techniques classiques : trouver des images similaires à un modèle, analyser les images en termes de motifs (des régions de couleur ou de texture à peu près constante) et, enfin, identifier les objets à la manière de l'être humain (la première tech-

nique est décrite dans *Les circuits configurables*, par John Villasenor et William Mangione-Smith, page 62).

Facile, mais inutile

La simplicité de chacune de ces techniques est inversement proportionnelle à leur efficacité. La comparaison directe de deux images est un jeu d'enfant, mais ne trouve que des ressemblances superficielles. Des changements de position, de composition ou de configuration des objets cherchés feront presque toujours échouer la comparaison : une vue de face d'un cheval ne ressemble pas à une vue de profil.

L'analyse de motifs, plus prometteuse, ne nécessite pas de calculs excessivement complexes : en une fraction de seconde, un programme détermine la proportion de pixels (éléments d'image) rouges, bleus ou verts d'une image. Comme les images sont composées de motifs, cette technique est utile au début d'une recherche : la plupart des logiciels de recherche d'images l'utilisent. C'est le cas, par



satisfont l'utilisateur, car l'objet recherché (les fleurs) y apparaissait comme des taches violettes. Pour trouver des objets plus complexes, tels des tigres, on utilise un algorithme qui recherche des zones de couleurs et de textures homogènes. Les

résultats de l'une de ces recherches sur le tigre sont montrés dans la rangée inférieure d'images. L'algorithme qui a trouvé ces images est beaucoup plus précis que ceux ignorent les formes. Ces deux programmes ont été mis au point à l'Université de Berkeley.

Trouver un cheval morceau par morceau

Les programmes de reconnaissance d'objets regroupent des éléments d'image en régions de taille et de complexité croissantes ; puis ils engendrent des hypothèses sur ce que ces régions composites représentent. Un programme reconnaissant des animaux à quatre pattes commence par analyser une image (a), à la recherche de zones qui pourraient être de la couleur des chevaux (b). Il définit les contours de ces régions (c), puis il cherche les régions ayant à peu près la forme du corps d'un cheval (le corps et les jambes sont approximativement des rectangles) (d). Le programme supprime ensuite les régions présentant des relations spatiales impossibles (le corps et les membres doivent être perpendiculaires) (e). Les classifications finales ne peuvent parfois être faites qu'avec une connaissance détaillée de la couleur et de la texture de la surface des objets ; même un être humain peut avoir du mal à distinguer sur la seule base de leur forme les silhouettes d'un éléphant à la trompe enroulée (en haut à gauche) et d'un ours (en haut à droite).



PHOTOGRAPHIE ORIGINALE



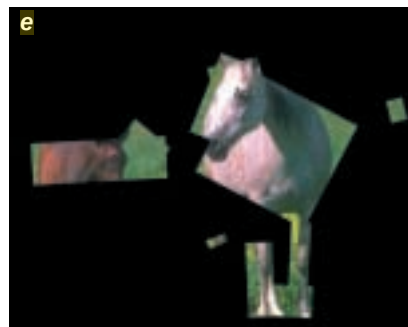
MOTIF (OMBRAGE DES ÉLÉMENTS DE CHEVAL)



CONTOURS DES MOTIFS



TOUS LES BLOCS CORPORELS POSSIBLES



BLOCS CONFIGURÉS COMME UN CORPS DE CHEVAL

D. Forsyth et M. Fleck

exemple, de QBIC (*Query by Image Content*, soit «recherche par contenu d'images»), l'un des plus anciens programmes de recherche d'images, mis au point par des chercheurs de la Société IBM : il permet à l'opérateur de spécifier l'image recherchée en termes de couleurs, de position des couleurs et de textures (celles-ci allant de la simple alternance de deux couleurs, comme les rayures du zèbre, à des motifs multicolores plus complexes). Le programme propose alors une sélection d'images ordonnées selon la concordance avec les critères demandés.

Un autre programme, *Photobook*, a été mis au point par Alex Pentland et ses collègues de l'Institut de technologie du Massachusetts : comme QBIC, il considère les images comme des juxtapositions de régions uniformes, mais il traite des textures plus

complexes et décompose automatiquement l'image en blocs.

Malgré leurs résultats impressionnants, ces programmes sont insuffisants. Par exemple, une recherche fondée uniquement sur la proportion de couleurs dans une image ne distingue pas le drapeau français du drapeau britannique. Plus généralement, les utilisateurs sont intéressés par le contenu, lequel n'est exprimé ni par un seul modèle ni par une similitude de motif.

En vue d'améliorer les recherches dans la bibliothèque numérique de l'Université de Berkeley, nous avons mis au point un système de recherche d'images qui traite le contenu. Le projet de bibliothèque, commencé par Ginger Ogle et Chad Carson, contiendra plus de 600 000 photographies. Environ 50 000 images provenant de nombreuses sources, y compris

des photographies aériennes et terrestres, sont déjà accessibles sur le réseau Internet.

Notre programme utilise les diverses techniques de recherche, de sorte que les utilisateurs peuvent à la fois rechercher des objets et des motifs. Cependant, la recherche par objet est difficile, et il est souvent plus efficace de demander des photographies qui contiennent un motif à partir duquel les objets pourront être retrouvés. Les interrogations par motifs les plus efficaces sont celles qui servent à construire des représentations des objets.

Lorsque l'un des concepteurs du programme chercha des images de planche à voile, il demanda des photographies contenant au minimum 30 pour cent de bleu-vert représentant le ciel ou la mer et au moins une petite partie de jaune pour la rayure jaune d'une voile. Une grande partie des



2. LA DISTINCTION d'un même objet vu sous des angles variés met au défi les programmes de reconnaissance d'objets. Le logiciel mis

au point par l'un des auteurs reconnaît parfois si une image comporte des chevaux quels que soient leur posture et l'arrière-plan.

La reproduction de la mémoire visuelle

Les spécialistes de la vision humaine ont établi qu'un être humain mémorise environ 100 000 images qu'il peut d'identifier en une fraction de seconde. Un robot doué de performances analogues pourrait se déplacer dans un environnement en se fondant sur sa vision : plutôt que de se repérer sur un plan qu'il est difficile de mettre en correspondance avec une image numérique, la machine reconnaîtrait les images de lieux déjà visualisés.



À Grenoble, nous avons mis au point un programme qui reconnaît les images. Ici la recherche se limite à des objets déjà vus plutôt qu'à des formes génériques comme des chevaux ou des humains.

Le principe de cette recherche est fondé sur un calcul de propriétés locales qui sont invariantes lorsque les prises de vues varient. Pour chaque image, en un certain nombre de points déterminés automatiquement, le programme calcule des propriétés invariantes par rotation et par d'autres transformations comme le changement d'éclairage. Il détermine ainsi une «signature» caractéristiques de l'image. En comparant les signatures, le programme trouve les images «déjà vues».

La figure de gauche montre la recherche dans une base de 1 000 images contenant des tableaux, des objets pris sous différents angles, des images aériennes. Les images recherchées le sont toujours à partir d'images prises dans des conditions différentes, par exemple suite à un déplacement de l'avion pour des images aériennes. Ainsi, lorsque l'on recherche l'image du haut, le programme trouve l'image du bas qui représente la même tour de Marseille, vue sous un angle différent. Ci-contre, on présente à l'ordinateur une image dans laquelle se trouve une partie de la queue d'un tyrannosaure, sur un fond de forêt. Là encore, comme dans la quasi-totalité (99 pour cent) des cas, le programme trouve une image qui correspond (le tyrannosaure en incrustation).

Roger MOHR et Cordélia SCHMID,
INPG, IMAG, INRIA, Grenoble

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 psychophysiologistes 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 écosais. 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