

Un ordinateur reconnaît les actions humaines dans les films

16 mars 2012, 01:00

Pascal Pollet

Un ordinateur capable de reconnaître visuellement les actions humaines dans les images vidéo ouvre de nouvelles perspectives pour de nombreuses applications industrielles, comme la surveillance mais aussi pour le contrôle de qualité.

Avec la croissance exponentielle des images vidéo des systèmes de caméras en circuit fermé, de la télé, des films, YouTube et d'autres sources, il devient nécessaire de trouver des méthodes pour analyser automatiquement le contenu de ces vidéos. La reconnaissance visuelle des actions humaines dans différentes séquences vidéo réalistes se heurtait jusqu'à présent à plusieurs obstacles, comme le manque de jeux de données vidéo réalistes. Grâce aux développements des chercheurs de l'institut français INRIA (Institut national de recherche en informatique et en automatique) et de l'université d'Oxford, les ordinateurs sont maintenant en mesure de reconnaître certaines actions humaines naturelles, comme se lever et s'asseoir, donner la main, marcher, s'embrasser, se faire la bise.

[Acceptez les cookies marketing pour regarder cette vidéo.](#) Pour les utilisateurs de FireFox, désactivez la « protection renforcée contre le suivi » de votre navigateur pour visionner cette vidéo.

Scripts de film et classification

[Acceptez les cookies marketing pour regarder cette vidéo.](#) Pour les utilisateurs de FireFox, désactivez la « protection renforcée contre le suivi » de votre navigateur pour visionner cette vidéo.

Les chercheurs français ont d'abord développé une manière de collecter automatiquement des exemples réalistes d'actions humaines sur base de scripts de films. Les jeux de données existants pour la reconnaissance visuelle des actions humaines comprennent des exemples d'entraînement de seulement quelques classes d'actions dans des situations contrôlées et simplifiées. Les chercheurs ont voulu s'attaquer à cette limitation. Ils ont commencé par réunir de bonnes images vidéo. A côté de ces fragments, ils ont placé les scripts de film correspondants. Ces textes décrivent le contenu du film : les scènes, les personnages, la transcription des dialogues et les actions humaines. Pour assurer la synchronisation avec la vidéo, les chercheurs ont utilisé des sous-titres avec indication des temps.



(Source : INRIA)

De la sorte, ils disposaient d'un jeu nettement plus important de bons exemples d'entraînement. Ensuite, ils ont développé une nouvelle description mathématique, les 'space-time interest points', pour les images vidéo. Ces points d'intérêt spatio-temporels permettent de décrire automatiquement des actions spécifiques dans une image en tant que vecteurs. Les vidéos peuvent alors être décrites comme un jeu de vecteurs. Ensuite, un système de classification a été entraîné à reconnaître la signification des points d'intérêts spatio-temporels sur la base d'exemples d'actions.

Les chercheurs de l'INRIA ont obtenu une nette amélioration avec la nouvelle méthode par comparaison à d'autres résultats récents sur le même jeu de données. Ils pouvaient ainsi reconnaître correctement huit actions humaines.

La méthode d'Oxford

Le système des chercheurs de l'université d'Oxford reconnaît les actions humaines sur la base de l'orientation de la tête, de la distance entre les personnes et de leur mobilité. Ces trois caractéristiques permettent de déceler avec une assez bonne précision le type d'action présent dans une image.

Les chercheurs ont utilisé un détecteur pour déceler le haut du corps des personnes dans chaque image d'une vidéo.



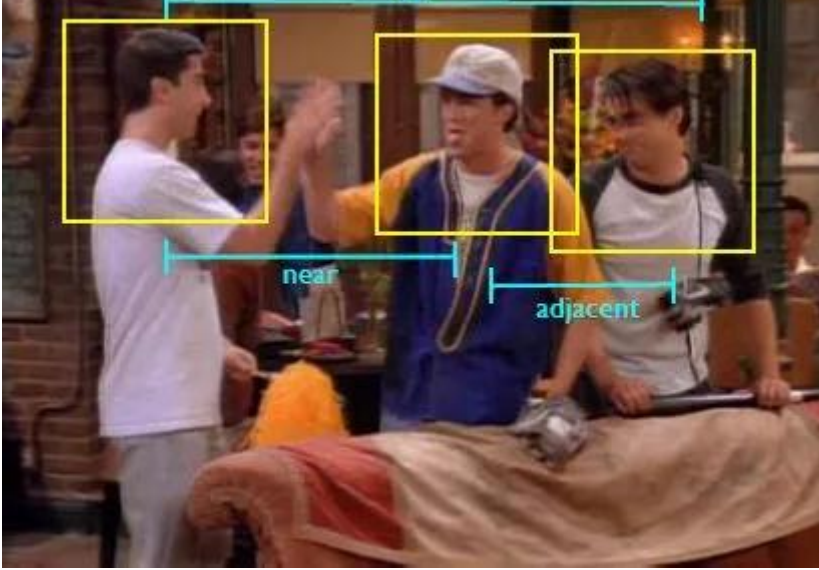
(Source : Robotics Research Group, Department of Engineering Science, University of Oxford)

On a ensuite appris à un système de classification pour l'orientation de la tête à reconnaître cinq positions (profil de gauche, profil de droite, frontal gauche, frontal droite, côté arrière).



(Source : Robotics Research Group, Department of Engineering Science, University of Oxford)

Ensuite, ils ont mesuré le déplacement local autour de chaque personne et la distance entre les personnes. La distance entre les personnes est une indication de la sorte d'interaction en cours ou à venir. Il sera plus plausible que deux personnes qui se rapprochent l'une de l'autre vont s'embrasser ou se donner la main que si elles restent éloignées l'une de l'autre.



(Source : Robotics Research Group, Department of Engineering Science, University of Oxford)

À l'aide des techniques d'apprentissage automatique, ils ont pu entraîner un ordinateur à reconnaître les bonnes actions sur base des informations extraites de l'orientation, de la distance et de la mobilité.

Potentiel d'application de la technologie de reconnaissance

Ces développements sont-ils utiles pour l'industrie ? De nombreuses tâches d'inspection industrielle par caméra reviennent encore généralement à compter les pixels au-delà d'un seuil donné.

Ces techniques informatiques pourraient trouver des applications par exemple dans la surveillance, mais aussi dans le contrôle qualité : en observant les opérateurs, on pourrait s'assurer qu'ils ont bien effectué toutes les opérations manuelles prescrites pendant la production d'un produit en évitant que les opérateurs cochent les listes de contrôle mécaniquement.

Sources

- <http://www.di.ens.fr/~laptev/actions/>
- http://www.robots.ox.ac.uk/~alonso/projects/human_interactions_project.html
- http://www.robots.ox.ac.uk/~alonso/tv_human_interactions.html

Auteurs



Pascal Pollet