

Computer herkent menselijke handelingen in films

16 maart 2012, 01:00

Pascal Pollet

Een computer die in staat is menselijke handelingen in videobeelden visueel te herkennen opent nieuwe perspectieven voor tal van industriële toepassingen, zoals bewaking, maar ook kwaliteitsborging.

Met de exponentiële groei van videomateriaal via CCTV-camera's, tv, films, YouTube en andere bronnen is er een groeiende behoefte aan methoden om automatisch de inhoud van dergelijke video's te analyseren. Het visueel herkennen van menselijke handelingen in verschillende realistische videobeelden stootte tot nu toe altijd op verschillende hindernissen, zoals een gebrek aan realistische videodatasets. Onderzoekers aan het Franse INRIA (Institut national de recherche en informatique et en automatique) en aan de universiteit van Oxford zijn nu een stap verder gekomen in ontwikkelingen die ervoor zorgen dat computers verschillende natuurlijke menselijke handelingen, zoals opstaan, gaan zitten, een hand geven, lopen, knuffelen en kussen, kunnen herkennen.

[Aanvaard marketingcookies om deze video te bekijken.](#) FireFox-gebruikers moeten de 'Verbeterde trackingbescherming' van hun browser uitschakelen om deze video te bekijken.

Filmscripts en classificatie

De Franse onderzoekers ontwikkelden eerst een manier om automatisch realistische trainingsvoorbeelden van menselijke acties uit films te verzamelen op basis van filmscripts. De bestaande datasets voor visuele herkenning van menselijke handelingen bevatten trainingsvoorbeelden van slechts enkele handelingsklassen in gecontroleerde en vereenvoudigde situaties. Deze beperking wilden de onderzoekers aanpakken. Ze begonnen met het verzamelen van interessante videobeelden. Naast deze fragmenten legden ze de bijbehorende filmscripts, die tekstuele omschrijvingen bevatten van de inhoud van de film op het vlak van scènes, personages, getranscribeerde dialogen en menselijke handelingen. Om een synchronisatie met de video te bereiken, maakten de onderzoekers gebruik van ondertitels met tijdsaanduiding.



(Bron: INRIA)

Op die manier konden ze beschikken over een veel grotere set van goede trainingsvoorbeelden. Daarna werd een nieuwe wiskundige beschrijving, 'space-time interest points', voor videobeelden ontwikkeld. Met 'space-time interest points' kunnen specifieke acties in een beeld automatisch worden beschreven als een vector. Op die manier kunnen video's worden beschreven als een set van vectoren. Vervolgens werd een classificatiesysteem getraind die op basis van de voorbeeldacties leerde wat de 'space-time interest points' betekenen.

De INRIA-onderzoekers bereikten een aanzienlijke verbetering met de nieuwe methode in vergelijking met andere recente resultaten op dezelfde dataset. Zo konden ze acht verschillende types van menselijke handelingen correct herkennen.

Detectie van positie en beweging

Het systeem van de onderzoekers aan Oxford University herkent menselijke handelingen op basis van de oriëntatie van het hoofd, de afstand tussen de personen en hun bewegelijkheid. Met deze drie kenmerken kon men vrij goed berekenen welke soort actie in een beeld voorkomt.

De onderzoekers gebruikten een detector voor bovenlichamen om mensen te vinden in elk frame van de video.



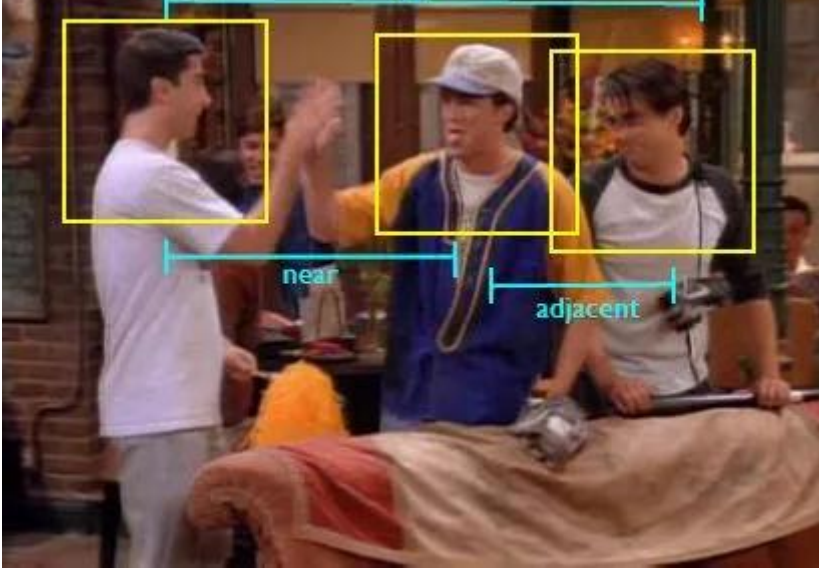
(Bron: Robotics Research Group, Department of Engineering Science, University of Oxford)

Een classificatiesysteem voor de oriëntatie van hoofden werd vervolgens vijf houdingen aangeleerd (profiel van links, profiel van rechts, frontaal van links, frontaal van rechts, achterkant).



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

Daarna worden de plaatselijke beweging rond elk persoon en de afstand tussen personen gemeten. De afstand tussen mensen vormt een aanwijzing van het soort interactie dat plaatsvindt of zal plaatsvinden. Zo is het bijvoorbeeld aannemelijker dat twee mensen die dicht bij elkaar zijn elkaar een knuffel of een hand zullen geven dan wanneer ze verder van elkaar verwijderd zijn.



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

Met behulp van machine-learning-technieken kon de computer op basis van de geëxtraheerde informatie over oriëntatie, afstand en beweeglijkheid getraind worden om de juiste acties te herkennen.

Toepassingspotentieel van herkenningstechnologie

Wat betekenen dergelijke ontwikkelingen nu voor de industrie? Veel industriële inspectietaken met camera's komen typisch nog neer op het simpelweg tellen van pixels boven een ingestelde threshold.

Dit onderzoek toont dat computers steeds meer cognitieve taken aankunnen die tot voor kort alleen maar door mensen konden worden verricht.

Mogelijke toepassingen van deze computertechnieken situeren zich bijvoorbeeld in bewaking, maar ook in kwaliteitsborging: zo zal men door het observeren van operators zekerheid hebben of alle voorgeschreven manuele operaties tijdens de productie van een product effectief uitgevoerd werden en vermijden dat operatoren blindelings checklists aankruisen.

Bronnen

- <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