Intelligente oren en ogen

Wetenschappers aan de UvA werken aan bewakingssystemen die abnormaal gedrag herkennen. Een intelligente surveillancecamera is veiliger en zou zelfs de privacy beter kunnen beschermen. **Marc Hijink**

De computer herkent aan de hand van camerabeelden de lichaamshouding van verschillende personen op het perron, door de omtrekken van hun armen, hoofd en torso te volgen. FOTO UVA

ALLES LIJKT RUSTIG in de centrale hal van de Haagse voetbalvereniging ADO. Mensen wandelen rond, kletsen wat met elkaar. Plotseling klinkt een schot en stuift de menigte uiteen: een aanslag!

De paniek is onecht: de commotie wordt door diverse professionele acteurs en tientallen figuranten nagespeeld. Er staan vijftien camera's en evenveel microfoons opgesteld om de scène van alle kanten in detail te registreren. Hier vinden opnames plaats van het ADABTS-project (voluit: *Automatic detection of abnormal behaviour and threats in crowded spaces*). Dit project probeert met intelligente sensoren abnormaal gedrag te detecteren. Bijvoorbeeld terrorisme, rellen, of criminaliteit zoals zakkenrollen en drugs dealen. De onderzoeksgroep van hoogleraar Dariu Gavrila aan de Universiteit van Amsterdam (UvA) ontwikkelt binnen ADABTS algoritmes die in videobeelden personen detecteren en hun gedragspatronen analyseren. "We willen een systeem ontwikkelen dat beveiligingsmensen in de controlekamer waarschuwt als er iets mis is", aldus Gavrila.

En dat lukt steeds beter: computers worden steeds slimmer als het gaat om het beoordelen van beelden. Voor bewakers is dat belangrijk: "Voor een mens is het onmogelijk om uren achtereen aandachtig al die monitoren te volgen waarop meestal niets bijzonders te zien is. Als je je vervelt, ga je al gauw zinloos op mensen inzoomen. Met het risico dat echt belangrijke ontwikkelingen aan het oog ontsnappen. Maar zodra de computer helpt bij het saaie detectiewerk, kan dit de aandacht van menselijke bewaker in de controlekamer in de juiste richting leiden."

SUPPORTERS Het vorige onderzoeksproject van Gavrila, Cassandra genaamd, was bedoeld om de camera te leren agressief gedrag herkennen. Als studiemateriaal dienden video- en audio-opnames van acteurs op het perron van station Amsterdam Amstel, die elkaar op verzoek van de UvA-wetenschappers afwisselend vriendelijk begroetten of flink te lijf gingen. Abrupte bewegingen, mensen die wild met de armen maaien en schreeuwen; dat bleek goed te herkennen door de ontwikkelde software. Zelfs als er ondertussen een trein door de scène reed. Gavrila: "Cassandra kon goed de rustige groepen van onrustige groepen onderscheiden. Wat moeilijker bleek: onderscheid te maken tussen een stel juichende of agressieve supporters." Die subtielere verschillen kunnen juist wel goed door een bewaker in de controlekamer beoordeeld worden.

In het nieuwe ADABTS-project ligt de lat hoger. De camera moet straks niet alleen specifieke handelingen herkennen die van tevoren zijn vastgelegd in algoritmes, maar ook met nieuwe situaties overweg kunnen. Want een systeemontwikkelaar kan niet alle mogelijke scenario's voor een zakkenroller of overvaller programmeren. Daarom ligt de nadruk binnen het ADABTS ook meer op het in kaart brengen van normaal gedrag, zodat alles wat hiervan afwijkt door het systeem als 'abnormaal' gesignaleerd kan worden – zonder het verder te moeten benoemen. De uiteindelijke beoordeling vindt plaats in de controlekamer door een mens.

De recente video-opnames bij ADO Den Haag – ettelijke terabytes data – vormen de basis voor het onderzoek. Bij beeldherkenning keek de computer tot voor kort alleen naar pixelbewegingen, in verticale en horizontale richting. Bij ADABTS worden er pixelgroepen gevormd die vertaald worden naar 3D-positie en houding van personen; de computer herkent menselijke torso's, armen en benen. Met die gegevens is beter te bepalen welke persoon er beweegt en met welke bedoeling dat lijkt te gebeuren.

Deze technologie heeft wel wat weg van Kinect, de nieuwe controller van Microsofts spelcomputer de Xbox. Kinect gebruikt echter een speciale infraroodcamera, die afstandsmetingen uitvoert om een 3D-beeld te creëren van de speler; dit beeld wordt rechtstreeks gebruikt om een spel aan te sturen. Daarentegen werkt ADABTS met conventionele videocamera's en gaat het erom gedragspatronen over een langer tijdsinterval te herkennen. Ook geluid speelt een belangrijke rol: TNO werkt in dit onderzoek mee met 'luisterende' camera's die afwijkende pieken en timbres in het geluid herkennen en zo agressie of angst detecteren.

MOGELIJKHEDEN "Het wordt nu pas leuk", zegt Ben Kröse. Hij is lector Digital Life bij de Hogeschool van Amsterdam en bijzonder hoogleraar aan de UvA, en al jaren bezig met beelddetectie. "We zijn nu zover dat computers snel genoeg zijn om *in real time*, dus zonder vertraging, beelden te analyseren. Dat biedt veel nieuwe mogelijkheden." En niet alleen op het gebied van beveiliging.

Kröse herinnert zich hoe Nederland bijna dertig jaar geleden kennismaakte met automatische nummerbordenherkenning. "Foto's van auto's die te snel reden werden ingescand met een voor die tijd enorme resolutie van 512 bij 512 pixels, grote computers deden de beeldherkenning." Inmiddels bestaan er gewone digitale camera's die automatisch afdrukken als er witte tanden te zien zijn. In fabrieken controleren camera's of er glasresten zitten in de flessen die op de lopende band voorbijrazen. En straks volgt de virtuele passpiegel in de kledingzaak. Kröse: "Je gaat voor de spiegel staan en de kleren worden automatisch in de juiste maat geprojecteerd."

Ook leuk voor de winkel: de camera houdt de loop- en kooppatronen van winkelend publiek in de gaten. Op die manier kan een winkelier zijn zaak efficiënter inrichten en de weekaanbieding op een betere plek neerzetten. Een intelligent bewakingssysteem kan potentiële autodieven herkennen, door de looppatronen van individuen te analyseren. Kröse: "De meeste mensen lopen op een parkeerterrein rechtstreeks op hun auto af. Maar een dief die op zoek is heeft een heel ander patroon."

Door acteurs nagespeelde panieksituatie in de centrale hal bij ADO Den Haag. FOTO UVA

VALPARTIJEN Beelddetectie kan ook binnenshuis van pas komen. In het project 'De mens voor de lens' bestudeert Kröse het gebruik van intelligente camerasystemen voor de gezondheidszorg. Hij laat voorbeelden zien van 'valdetectie' in zorginstellingen: een systeem dat in de gaten houdt of patiënten vallen. Ze worden op hun kamer in de gaten gehouden door een camera in het plafond. Zodra iemand struikelt of zijn balans verliest, slaat het systeem alarm en krijgt het verzorgend personeel een waarschuwing. In het lab van de HvA is een bejaardenflat nagebouwd waar acteurs op commando vallen. Zo kunnen de camera's getraind worden op realistische scenario's. Kröse: "We moeten daarbij ook infraroodcamera's gebruiken, want de meeste valpartijen vinden 's nachts plaats."

De algoritmes worden zo geprogrammeerd dat ze nauwelijks valse meldingen geven. Je kunt beter 10 procent van de valpartijen missen dan dat verzorgingspersoneel om de haverklap voor niets komt opdraven, zegt Kröse: "Want als er dan eens echt iets misgaat, komen ze niet."

Maar vormen al die camera's in huis, aan talloze gebouwen en in uitgaanscentra in Nederlandse steden geen grote bedreiging voor de privacy? Kröse denkt dat dat met intelligente detectiesystemen wel meevalt, zeker als er geen gebruikgemaakt wordt van gezichtsherkenning. "Je moet er wel voor zorgen dat alle data lokaal en tijdelijk worden opgeslagen. En dat de gegevens alleen door een computer worden beoordeeld, niet door een mens."

Ook Dariu Gavrila denkt dat de privacy niet per se gevaar hoeft te lopen. Geautomatiseerde beveiligingscamera's die geen identificatie uitvoeren kunnen volgens hem zelfs beter voor de privacy zijn. "Ze zijn 'objectiever' dan menselijke bewakers die eerder geneigd zijn personen eruit te pikken op basis van uiterlijk."

Maar volgens de UvA-wetenschappers staat gezichtsherkenning sowieso nog in de kinderschoenen. Afgelopen zomer startte het regionale vervoersbedrijf RET in Rotterdam een project dat alle reizigers op 'probleemlijn' 2 controleert. Dertien trams zijn uitgerust met speciale camera's die biometrische gegevens vergelijken met een database van bekende raddraaiers. De foutmarges zijn nog erg hoog (meer dan 20 procent) omdat mensen een muts of een baard kunnen dragen en de lichtomstandigheden onvoorspelbaar zijn, zeker in een rijdende tram.

STEREOCAMERA'S Bij de UvA proberen ze slechte lichtomstandigheden te compenseren door twee camera's te gebruiken. Zo'n stereo-opstelling is robuuster doordat lichtinformatie van twee bronnen met elkaar vergeleken wordt. Daarnaast zitten er waardevolle gegevens in de dieptemeting van twee invalshoeken. Kröse: "Met stereocamera's is het ook eenvoudiger om de lichaamsdelen van mensen te herkennen en is een betere hoofddetectie mogelijk."

Dat is met name nuttig voor een andere toepassing van bewakingssystemen: het tellen van menigten. Zo kan de politie betere schattingen doen over de aantallen mensen die een voetbalwedstrijd, een popconcert of een demonstratie bijwonen. Met deze technologie had wellicht de ramp op de Loveparade, afgelopen zomer in Duisburg, voorkomen kunnen worden. Door een verkeerde inschatting van het aantal bezoekers kwamen 19 mensen om het leven in een paniekerige mensenmassa.

Marcel Worring, onderzoeker aan de UvA, houdt zich regelmatig met politiezaken bezig. Hij perfectioneert een systeem om grote hoeveelheden videogegevens te analyseren. Dat kan veel tijd besparen in het forensisch onderzoek, vertelt hij: "Neem de terroristische aanslagen op de metro in Londen, in 2005. Agenten zijn twee jaar bezig geweest om tienduizend uur videomateriaal terug te kijken om de daders te volgen door de stad."

CAMPUS Voor zijn onderzoek gebruikt Worring een Amerikaanse voorbeeldvideo die ook in andere laboratoria gebruikt wordt als dataset. In de film lopen personages over een Amerikaanse campus, gefilmd vanuit verschillende hoeken. Een van de personen geeft de ander een pakje en loopt weer uit het beeld. Later verschijnt hij weer op een andere camera.

Het is de bedoeling dat de software uit deze beelden patronen samenstelt. Relevante resultaten worden ter beoordeling voorgelegd aan een menselijke specialist. Die geautomatiseerde voorselectie moet duizenden manuren besparen. Zeker in Groot-Brittannië: dat is wereldwijd koploper met vier miljoen beveiligingscamera's. Londen heeft een *ring of steel* van honderdduizenden videocamera's. Dat levert bij een eventuele volgende calamiteit nog veel meer beeldmateriaal op.

Terwijl Worring zich buigt over analyse van videobeelden achteraf, werkt Dariu Gavrila bij autofabrikant Daimler aan toepassingen voor intelligente auto's waarbij het juist om de razendsnelle interpretatie van videobeelden draait. *Pre-crash* systemen die de auto laten remmen bij een dreigende kop-staartbotsing bestaan al langer: de radar detecteert naderende obstakels.

Daimler, maker van de Mercedes-Benz, werkt nu aan een veiligheidssysteem dat ook overstekende voetgangers detecteert. Als die achter een geparkeerde auto vandaan lopen, moet de auto in staat zijn snel af te remmen om de gevolgen van een eventuele botsing te verminderen.

Bij gewone beveiligingsbeelden is er een stabiele camera waarbij de achtergrond stilstaat en objecten gedetecteerd worden omdat ze bewegen. Maar in een bewegende auto gelden andere wetten, legt Gavrila uit: „De achtergrond verandert continu en de auto schudt een beetje. We proberen dat effect te verminderen door twee camera's te gebruiken."

Maar de voetgangers zijn lastiger te signaleren dan de grote auto's. Voetgangers variëren enorm in verschijning – in bijvoorbeeld houding of kleding. Daarom worden overstekende mensen met behulp van statistische methodes zoals neurale netwerken gedetecteerd. Die moeten worden getraind en daarvoor heeft Gavrila met zijn team meer dan een miljoen voorbeelden van voetgangers en niet-voetgangers verzameld.

VOETGANGER Bij Mercedes zit het systeem nog niet in productieauto's, maar er is wel een prototype dat op tijd remt als er een voetganger tevoorschijn komt. Daimler is bezig met de volgende uitdaging: het systeem zo intelligent maken dat de auto ook kan besluiten om uit te wijken in plaats van alleen te remmen. Dat kan nog meer levens sparen. De onderzoeker laat een filmpje zien waarbij dummypop (bij Daimler noemen ze hem *Der Hansie*) oversteekt, het Mercedes-stuur uit zichzelf naar links draait en weer naar de oorspronkelijke route stuurt.

Maar dan is er nog een probleem: er moet wel ruimte zijn op de andere weghelft om de ingreep uit te voeren. Want automatisch uitwijken om vervolgens op de bumper van een tegenligger te belanden – dat kan natuurlijk niet.

Maar gelukkig zit er, hoe intelligent de camera ook is, voorlopig altijd nog een mens achter die de definitieve beslissingen neemt. ●

Beeld van het computerprogramma Cassandra waarbij de computer situaties analyseerde op het Amstelstation. FOTO UVA

'Der Hansie' steekt plotseling over vanachter een busje, maar de naderende Mercedes remt op tijd af dankzij de ingebouwde automatische voetgangerdetectie. FOTO UVA

DE COMPUTER LEERT NAAR MENSEN KIJKEN

Hoe kan een computer mensen en hun bewegingen herkennen? Beeldherkenningssoftware vergelijkt videobeeldjes met elkaar. Video bevat 24 of 30 beeldjes per seconde, die elk honderdduizenden pixels bevatten (van VGA-kwaliteit: 640x480 of high definition (HD): 1920x1080 pixels – maar HD-camera's zijn nog zeldzaam in de beveiligingsbranche). Ieder van die pixels heeft in het computergeheugen een waarde. Dat getal bevat informatie over de kleuren en de lichtintensiteit.

Een vaste bewakingscamera filmt bewegende objecten, zoals een wandelaar, ten opzichte van een relatief vaste achtergrond. Met de zogeheten *background substraction*-techniek wordt dan de waarde van de pixels van het beeld met alleen de achtergrond afgetrokken van het beeld met de wandelaar erop. Zo blijft alleen de wandelaar over, als een bewegend silhouet in een yerder zwart plaatje. Maar in de praktijk beweegt ook vaak de achtergrond, bijvoorbeeld door waaiende bomen of langsrijdende auto's. Een geavanceerder manier is om de camerasoftware de wandelaar zelf te laten vinden en analyseren. En hem niet 'over te houden' door het achtergrondbeeld weg te rekenen.

Dat kan door de computer te laten rekenen aan de kleurwaarden van de pixels, hun samenhang met buurpixels en de plaats in het beeld. Clusters van pixels worden dan geselecteerd op huidskleur en op gezichtskenmerken. Zo zijn menselijke figuren ook te herkennen tegen een drukke achtergrond.

Als een persoon eenmaal 'in beeld' is, kan de camerasoftware een twee- of zelfs driedimensionaal skelet over het bestaande videobeeld plakken, waardoor mensen beter te volgen zijn.

Met een 'woordenboek' van veelvoorkomende bewegingen (bijvoorbeeld lopen, rennen, draaien, zitten, zwaaien en wijzen) kan de computer volgen welke acties de personen uitvoeren. En alarm slaan als het gedrag 'afwijkt'. De scenario's worden gecompliceerder als er ook interactie tussen objecten (dragen, gooien, schudden) of tussen mensen (hand geven, omhelzen, slaan) aan toegevoegd wordt.