

Automatische gedragsanalyse voor effectiever cameratoezicht in de openbare ruimte

Henri Bouma, Jeroen van Rest, Gertjan Burghouts, Klamer Schutte, Jan Baan

TNO, Oude Waalsdorperweg 63, 2597 AK The Hague, The Netherlands
E-mail: henri.bouma@tno.nl

ABSTRACT

To improve security in crowded environments, such as airports, shopping malls and railway stations, the number of surveillance cameras (CCTV) is rapidly increasing. However, the number of human operators remains limited and only a selection of the video streams can be observed. This makes it hard for an operator to be proactive. This paper gives an overview of novel developments that may lead to more efficient camera surveillance and a more proactive role for camera operators. It focuses on three main steps in this process of video content analysis: pedestrian tracking, action recognition and behavior analysis. Tracking and re-identification (i.e. recognizing a person in another camera) was initially only evaluated on off-line benchmark datasets, though recently it has gained in maturity with live demonstrations in realistic crowded environments and measured improved operator efficiency. For action recognition and automatic behavior recognition, we observe that the simple patterns, such as loiter detection, are emerging in many applications. Human action recognition obtains very high performance values in controlled environments and it is progressing towards more realistic environments. More advanced approaches, such as pickpocket recognition in a shopping mall and the detection of threats to trucks on a parking lot have been developed and the first systems have been presented in live demonstrations. Our main contribution is that we structure the recent advances and the emerging applications of video analysis for security applications, explain and interpret the results, and identify opportunities for the near future.

Keywords: *Behavior analysis, Threat detection, Action recognition, Tracking, Re-identification.*

Title: Automatische gedragsanalyse voor effectiever cameratoezicht in de openbare ruimte
Authors: H. Bouma, J.H.C. van Rest, G.J. Burghouts, K. Schutte, J. Baan
Journal: Tijdschrift voor Veiligheid
Year: 2014
Volume: 13
Issue: 4
Pages: 20 – 34
Publisher: Boom Lemma Uitgevers (The Hague, The Netherlands)
DOI: <http://dx.doi.org/10.5553/TvV/187279482014013004002>

Automatische gedragsanalyse voor effectiever cameratoezicht in de openbare ruimte

Henri Bouma, Jeroen H.C. van Rest, Gertjan J. Burghouts, Klamer Schutte, Jan Baan

TNO, Oude Waalsdorperweg 63, 2597 AK Den Haag
E-mail: henri.bouma@tno.nl

Om de veiligheid te verhogen in drukke omgevingen worden steeds meer beveiligingscamera's gebruikt. Het aantal cameraoperators blijft echter beperkt en slechts enkele camera's kunnen worden bekeken. Hierdoor zit de operator vaak in een reactieve rol. Dit artikel geeft een overzicht van de laatste ontwikkelingen die kunnen leiden tot effectiever cameratoezicht en een meer proactieve rol van cameraoperators. Hiervoor wordt gefocust op video analyse technieken waarmee mensen kunnen worden gevolgd of worden teruggevonden, en op het automatisch detecteren van verdacht gedrag. Voor beide wordt de recente vooruitgang en de opkomende toepassingen getoond.

1. INTRODUCTIE

De praktijk van cameratoezicht in de openbare ruimte is meestal reactief. Een incident wordt door burgers of straat-surveillance gemeld, met een verzoek aan de cameraoperator om "even mee te kijken". Met de cameraoperator heb je een extra paar ogen, en het is voor de bewijskracht achteraf prettig als (de rest van) het incident opgeslagen wordt. Ook kan de cameraoperator efficiënter naar een groter of ander gebied overschakelen als dit nodig mocht zijn.

Hoewel dit een praktisch nuttige bijdrage is van cameratoezicht, is het niet de route om meer veiligheid tegen lagere kosten te krijgen. De rol van de cameraoperator verschuift daarom van ondersteuner op afroep naar proactieve aangever (Flight, 2013). Daarom worden toezichthouders op straat tegenwoordig bijvoorbeeld getraind in *predictive behaviour profiling*¹, een methode die bedoeld is om al vóór een incident effect te bereiken. Voorkomen is immers beter dan genezen. Dit levert namelijk zowel meer veiligheid, als minder kosten op. En doordat een incident in de openbare ruimte in potentie ook door veel mensen wordt ervaren, mag er ook een positief effect op de subjectieve veiligheid verwacht worden.

Er zijn een aantal ontwikkelingen op het gebied van video analyse technieken (VCA) die de veiligheid kunnen vergroten. VCA is het automatisch interpreteren van video beelden. In de context van toezicht van de openbare ruimte, is video content analyse het automatisch interpreteren van beelden van toezichtcamera's om daarmee relevante onderdelen automatisch te selecteren. Al meer dan tien jaar geleden bestond de verwachting dat VCA meerwaarde zou hebben bij het surveilleren van de openbare ruimte. Ondanks allerlei projecten en initiatieven is die verwachting tot op heden nog niet waargemaakt. De *Gartner Hype Cycle* van 2013 stelt dat *content analytics* ongeveer op de piek van de hype is (Fenn, 2013). Eenvoudige vormen van VCA hebben zich echter ruim bewezen. Functies die op grote schaal worden toegepast zijn indringerdetectie, het tellen van mensen in wachtrijen, het detecteren van rondhangen, het meten

¹ <https://www.politieacademie.nl/onderwijs/onderwijsaanbod/Pages/opleiding.aspx?code=3200978>

van de drukte in winkels en het identificeren van mensen en voertuigen in gecontroleerde omstandigheden. VCA in de openbare ruimte is echter uitdagend, want het gaat hierbij om relatief grote gebieden, waarbij het leven van alledag zo weinig mogelijk gestoord mag worden, en waarbij grote variaties zijn te vinden in gedrag, (weers)omstandigheden en dreigingsniveau.

Recent onderzoek op het gebied van VCA maakt het nu echter mogelijk om ook te midden van normaal gedrag de dreiging van een overval of zakkenrollerij vroegtijdig te signaleren en om daarna de daders live door de openbare ruimte automatisch te volgen. Hierdoor krijgt de cameraoperator een meer proactieve rol, en kan hij helpen zorgen voor een effect vóór het incident. Minder kosten dus door meer veiligheid. De centrale probleemstelling van dit artikel is dus welke technologische ondersteuning er mogelijk is om deze rolverschuiving te ondersteunen.

De structuur van dit artikel is als volgt. In het vervolg van Sectie 1 wordt ingegaan op privacy en factoren die bepalen of VCA bijdraagt aan toezicht. Sectie 2 introduceert een aantal voorbeeld scenario's om de behoeftes van cameratoezichthouders mee te illustreren. Sectie 3 gaat gedetailleerder in op de technische ontwikkelingen van video analyse. In Sectie 4 wordt beschreven hoe recente ontwikkelingen op het gebied van VCA de behoeftes van cameraoperators kunnen helpen vervullen. Tenslotte bevat Sectie 5 de conclusies.

1.1 Ethiek en privacy van intelligente camera's

De vrijheid van de openbare ruimte – en van de mensen daarin – is het waard om te worden beschermd. Soms is deze vrijheid echter kwetsbaar, en helaas zijn er regelmatig concrete dreigingen. Het is dus ethisch om actief aan veiligheid te werken, en in Nederland hebben we daar op allerlei niveaus verantwoordelijkheden voor belegd. Echter, veiligheidsmaatregelen kunnen ook negatieve effecten hebben, ook op de (beleving van) vrijheid, bijvoorbeeld door een inperking van de privacy. Intelligente camera's kunnen een extra impact hebben op de privacy ten opzichte van reguliere camera's, zowel positief als negatief. Een negatief voorbeeld is dat bestaande *biases* bij toezichthouders een veel grotere impact hebben als ze mensen kunnen vinden of selecteren op huidskleur, kleding of gedrag (Bellanova et al., 2011). Een voorbeeld van een positieve impact is dat beelden automatisch afgeschermd of weggegooid kunnen worden als ze niet relevant blijken te zijn (Roelofsen, 2003).

De concrete negatieve impact van de inzet van intelligente camera's op de privacy hangt veel af van of en hoe privacy wordt meegenomen in de ontwikkeling en uitrol van concrete systemen en diensten. Ontwerpen betekent keuzes maken, maar om keuzes te maken moeten er wel opties zijn: zowel procedureel, organisatorisch als technisch. *Privacy-by-design* is een integrale methodiek voor het meenemen van privacy vanaf de vroegste fases van de ontwikkeling van een systeem (Rest, Boonstra et al., 2014). In de uitwerking van de voorbeelden zullen we beschrijven hoe deze methodiek helpt om de impact van intelligente camera's op de privacy te verkleinen.

1.2 Factoren die bepalen of VCA bijdraagt aan toezicht in de openbare ruimte

De openbare ruimte is strikt genomen een ruimte die voor iedereen toegankelijk is. Voorbeelden zijn een winkelcentrum, treinstation, parkeerplaats, winkelstraat, stadsplein of het winkelgebied van een luchthaven. Voor intelligente camera's maakt het weinig uit of er toegangscontrole is. De bevindingen zijn dus relevant voor alle omgevingen *waarbinnen* mensen zich vrij kunnen bewegen, bijvoorbeeld *airside* van een luchthaven, of het terrein in een gevangenis, en dan zowel binnen als buiten.

Er zijn diverse factoren die bepalen of VCA aan toezicht in de openbare ruimte een bijdrage kan leveren. Deze factoren kunnen worden gegroepeerd in context, omgeving en systeemconcept.

Onder de noemer context scharen we ten eerste het juridisch kader (Flight, 2013). De wet Bescherming Persoonsgegevens regelt bijvoorbeeld hoe cameratoezicht mag worden gebruikt, dat het kenbaar gemaakt moet worden en wat de bewaartermijn is. Ten tweede is er de mate van privacy- en veiligheidsbewustzijn van de mensen die het toezicht ondergaan, gekoppeld aan het dreigingsniveau. Het draagvlak voor meer invasieve beveiligingsmaatregelen zal hiervan afhangen. Het risico op incidenten is ook belangrijk omdat veiligheidsmaatregelen dienen om risico's te verkleinen. Tenslotte is ook de modus operandi van belang, omdat soms voorbereidende handelingen kunnen worden gedetecteerd (Rest, Nunen & Roelofs, 2014).

Onder de noemer omgeving scharen we bijvoorbeeld de mensdichtheid. Hoe hoger deze wordt, hoe moeilijker het is om goed onderscheid te maken tussen individuen en hoe minder functionaliteit VCA kan leveren. Door waargenomen dominante looprichtingen in een druk gebied kunnen individuen beter worden gevolgd in de menigte (Hu et al., 2012; Thida et al., 2013). Ook de invloed van het weer en de tijd van de dag zijn van belang als de ruimte buiten is, of althans veel zicht naar buiten heeft.

Onder de noemer systeemconcept horen, naast de technologie, ook de werkprocessen. In dit artikel wordt gefocust op de openbare ruimte, maar daarbinnen zijn nog allerlei varianten te maken, met name in het samenspel tussen toezichthouders op straat en cameratoezichthouders op afstand. De mens – met zijn competenties, opleiding en training – is hierin een relevante factor. Naast zijn sterktes, zijn ook zwaktes (zoals vooroordelen) relevant voor technologie die wordt ingezet. Ook het opnameplatform is van belang. Naast de bewegende pan-tilt-zoom (PTZ) camera's, zien we een groei van echt mobiele apparaten, zoals bodycams, smartphones en onbemande luchtvaartuigen (UAV's). Beelden die daarmee gemaakt worden zijn moeilijker te ontsluiten doordat ze geen constant gezichtspunt hebben. Tenslotte is de mate van interoperabiliteit een factor van belang bij samengestelde systemen. Als een toezichtstelsel bestaat uit verschillende subsystemen dan moet extra aandacht worden besteed aan de koppelvlakken om de beelden geschikt te houden voor VCA (Little et al., 2013).

In de volgende sectie wordt aan de hand van scenario's beschreven welke concrete behoeftes er zijn en de daaropvolgende sectie beschrijft hoe VCA daaraan kan bijdragen.

2. SCENARIO'S IN DE OPENBARE RUIMTE

Er zijn diverse scenario's die de behoefte van (camera)toezichthouders in de openbare ruimte kunnen helpen verduidelijken. Deze scenario's zijn gekozen op basis van onderzoek naar behoeftes zoals in algemene zin geformuleerd in de Roadmap Beeldtechnologie van het Ministerie van BZK (Flight, 2011), de aansturing van de vraaggestuurde onderzoeksprogramma's door het Ministerie van Veiligheid en Justitie en andere OOV organisaties (Don, 2012), en in diverse Europese Kaderprogramma's. Bijlage G van (Rest, Nunen & Roelofs, 2014) bevat een overzicht van (Europese) projecten en bijbehorende (KP7) call referenties waarin om technologie op gebied van afwijkend gedrag wordt gevraagd. We behandelen achtereenvolgens het volgen van een overvaller direct na de overval, het vroegtijdig signaleren van zakkenrollerij in een winkelcentrum en het detecteren van een overval op een vrachtwagen op een parkeerplaats langs de snelweg.

2.1 Volgen van een overvaller direct na de overval

Een overval is een incident waarbij bedreiging of geweld wordt gebruikt. Wanneer er een dergelijk incident heeft plaatsgevonden, wordt een melding gemaakt. Als de melding vlak na het incident binnenkomt, dan kan een cameraoperator de overvaller op twee manieren vinden. Hij kan in verschillende camera's rondkijken of daar een verdachte aanwezig is of hij kan (wanneer hij daartoe is bevoegd) terugkijken in opgenomen materiaal op de plaats van het incident tot hij

de verdachte ziet. Vanaf dat moment kan hij versneld de posities van de verdachte voorwaarts in de tijd volgen tot de positie is bereikt waar de verdachte zich nu bevindt. Het vinden van de overvaller kan tijdrovend zijn, zeker als moet worden gekeken in een groot gebied met veel camera's en als er blinde vlekken bestaan tussen de camera's. Vanaf het moment dat de huidige locatie van de dader is gevonden zal de verdachte worden gevolgd en zullen agenten of beveiligingspersoneel worden begeleid tot ze ter plaatse zijn. Het is hierbij belangrijk dat een operator in staat is om de verdachte snel te vinden en vervolgens kan blijven volgen over meerdere camera's voordat de gedachte het gebied met beveiligingscamera's verlaat. De behoefte om sneller personen terug te vinden ontstaat niet alleen in de heterdaad situatie maar ook voor reconstructie achteraf. Daarbij gaat het niet alleen om de vraag wat er gebeurde na het incident ('waar ging een verdachte naar toe?'), maar ook wat er gebeurde voor het incident ('waar kwam een verdachte vandaan?'). Bij het terugvinden van een verdachte in een groot aantal camera's zou VCA de operator kunnen ondersteunen. Vanuit o.a. juridische overwegingen is het echter niet vanzelfsprekend dat technische koppelingen bestaan tussen verschillende cameratoezichtsystemen als de vluchtende overvaller zich verplaatst. Dit kan vertraging opleveren die de pakkans verkleint.

2.2 Vroegtijdig signaleren van zakkenrollerij in een winkelcentrum

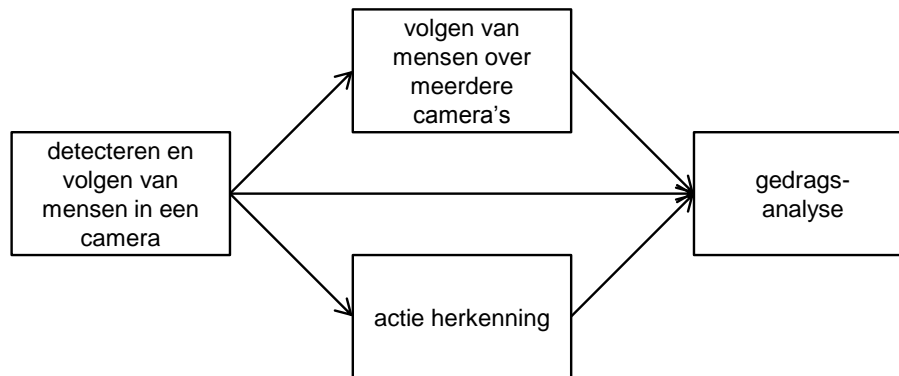
Incidenten die veel voorkomen leiden tot een grote belasting van beveiligingspersoneel. Eén voorbeeld van *high-volume crime* (HVC) is het zakkenrollen. Agenten ter plaatse en cameraoperators kijken vaak in het gebied of er afwijkende of verdachte activiteiten plaatsvinden. Hiervoor wordt het gedrag van de personen bestudeerd. Vaak wordt geconstateerd dat er niets verdachts plaatsvindt, maar als een cameraoperator wel iets verdachts waarneemt wordt assistentie gevraagd van beveiligers of agenten ter plaatse. Cameraoperators hebben nu vaak de verantwoordelijkheid voor een groot gebied met een groot aantal camera's. Het aantal is zo groot dat de camera's beurtelings moeten worden bekeken. In één uur wordt per camera typisch minder dan vijf minuten gekeken. Verreweg het grootste gedeelte van de tijd wordt de informatie die een camera zou kunnen leveren dus genegeerd. VCA zou de operator kunnen ondersteunen in de cameraselectie waardoor de juiste camera's worden bekeken op het moment dat er iets verdachts gebeurt. Het risico bestaat echter dat intelligente camera's gaan letten op gedrag of uiterlijke kenmerken die het optreden van biases versterken. Bijvoorbeeld door automatisch te leren welke kleding mensen aanhebben die vaak bij zakkenrollerij betrokken zijn.

2.3 Detecteren van dreigingen op een vrachtwagen op een parkeerplaats langs de snelweg

Op parkeerplaatsen langs snelwegen wordt veel diefstal van lading van vrachtwagens gepleegd. Het is ondoenlijk om alle parkeerplaatsen in Europa – die worden gebruikt voor internationaal transport – uit te rusten met beveiligingssystemen of om intensief toezicht te realiseren op alle parkeerplaatsen. Daarom is er behoefte aan een VCA systeem dat, bijvoorbeeld vanaf een vrachtwagen zelf, een stuk veiligheid garandeert. Gezien het aantal vrachtauto's, en de noodzaak van rust voor de chauffeur moet een dergelijk systeem automatisch eventuele dreigingen detecteren. De grote uitdaging hierbij is het onderscheid maken tussen het gedrag van een dief en het normale achtergrond gedrag op een parkeerplaats.

3. VIDEO ANALYSE TECHNIEKEN (VCA)

In ieder van de voorgaande scenario's is er al een rol voor cameratoezicht. In dit hoofdstuk beschrijven we relevante nieuwe ontwikkelingen op het gebied van VCA. Er zijn verschillende functies die met VCA verkregen kunnen worden. Basisfuncties zijn het detecteren en volgen van mensen en het herkennen van eenvoudige acties (zoals lopen, rennen en vallen). Op basis van deze basisfuncties kan meer complex gedrag worden herkend, zoals dat van zakkenrollers (Figuur 1). In deze sectie zal worden ingegaan op de technische ontwikkelingen van VCA. In de sectie daarna wordt getoond hoe deze kunnen bijdragen in de geschetste scenario's.



Figuur 1: Complexe functies worden opgebouwd uit basisfuncties.

3.1 Detecteren en volgen van mensen in één camera

Het automatisch volgen van mensen in één camera in rustige omgevingen is een volwassen technologie geworden die wordt gebruikt in diverse toepassingen². De automatische mensdetectie systemen die zijn geïntegreerd in de camera zijn vaak gebaseerd op beweging in het camerabeeld. Er wordt in de loop van de tijd een model opgebouwd van de achtergrond en wanneer een bewegend object (b.v. een persoon) verschijnt wordt een verschil met het achtergrond model waargenomen. Deze benadering is voldoende voor rustige omgevingen, zoals de beveiliging van een bedrijventerrein dat 's nachts verlaten hoort te zijn, maar het is onbruikbaar om mensen te volgen in drukke omgevingen. In drukke omgevingen zijn zoveel bewegingen dat alle personen met elkaar zouden versmelten. Daarom wordt in moderne oplossingen een mens-model gebruikt (Dollar et al., 2010; Felzenszwalb et al., 2010). Zo'n moderne detector is in staat om een persoon waar te nemen in een enkel beeld op basis van de vorm van een mens. Daarnaast zijn er nog een aantal recente aanpassingen om te zorgen dat ook goed kan worden omgegaan met oclusies (i.e. blokkade in het zicht waardoor iets of iemand niet – of niet compleet – zichtbaar is). Door deze aanpassingen is het waarnemen van alleen het bovenlijf voldoende om een mens te detecteren. Na de detectie worden mensen gevolgd in het camerabeeld op basis van verplaatsing en kleurkenmerken (Hu et al., 2012).

3.2 Volgen van mensen over meerdere camera's

Het terugzoeken en volgen is een zelfstandige toepassing, omdat een cameraoperator vaak wil weten: “Waar kwam iemand vandaan?” of “Waar ging hij naartoe na een incident?”. Aan de andere kant is het ook een ‘enabling’ technologie omdat de opgebouwde *tracks* informatie bevatten over het looppatroon en omdat deze technologie het mogelijk maakt om zwakke

² URL's: vinotion.nl, ipsotek.com, vdgsecurity.com, nice.com.

signalen van verschillende locaties te combineren tot sterkere signalen voor gedrags- en dreigingsanalyse (Bouma, Vogels et al., 2013).

Er zijn verschillende systemen beschikbaar die claimen dat ze het terugzoeken of volgen ondersteunen, waaronder de systemen van Aralia, AgentVI, 3VR, AxxonSoft, Ipsotek, iOmniscient, DVTel, ObjectVideo en Bosch³. In het Engels wordt deze functionaliteit ook wel 'tracing', 'multi-camera tracking', 'forensic search' of 'person re-identification' (Gong, 2014) genoemd. De functionaliteit is relatief nieuw en het wordt nog niet breed gebruikt voor toezicht in de openbare ruimte. Sommige van deze producten zijn niet geoptimaliseerd voor het vinden van personen, maar eerder voor auto's of andere objecten. Het is gebruikelijk om mensen terug te vinden op basis van een voorbeeldplaatje, b.v. een uitsnede uit een camerabeeld. Recent begint het ook mogelijk te worden om een semantische beschrijving te geven van b.v. kledingkleur (Satta et al., 2012). Dit is vooral nuttig wanneer er geen voorbeeldplaatjes beschikbaar zijn om de zoekopdracht mee te starten.

Het volledig automatisch volgen van meerdere personen over meerdere camera's die geen overlappend gezichtsveld hebben in een drukke omgeving is op dit moment niet beschikbaar als betrouwbaar product voor eindgebruikers. Het volgen wordt betrouwbaarder als er meer camera's zijn, bijvoorbeeld wanneer personen altijd vanuit meerdere standpunten worden waargenomen (Rest et al., 2009). Als in een rustige omgeving camera's dichtbij elkaar staan kunnen mensen worden gevolgd door gebruik te maken van de geschatte positie van die persoon. Om personen in een drukke omgeving over meerdere camera's te kunnen volgen zonder overlap in het blikveld is het belangrijk om te kunnen vergelijken op basis van uiterlijke kenmerken. Gezichtsherkenning werkt goed in een omgeving waar medewerking kan worden afgedwongen (b.v. bij een sportstadion of vliegveld waar poortjes pas opengaan als iemand recht in een hoge-resolutie camera heeft gekeken), en recente publicaties laten zien dat in vrije omgevingen met gewone beveiligingscamera's gezichtsherkenning ook kan worden toegepast (Vera-Rodriguez et al., 2014; Phillips & O'Toole, 2014). Het blijft lastiger om gezichten te herkennen wanneer mensen hun gezicht afwenden of afdekken met een hoofddekseel, en juist de doelgroep zal proberen oogcontact met de camera's te vermijden. Maar ook gedeeltelijke gezichtsbedekking is niet meer onoverkoombaar (Alyuz et al., 2012). Hoewel het volgen van mensen lagere eisen stelt aan de resolutie dan die voor identificatie, blijken de gezichtsherkenningssystemen nog altijd een relatief hoge resolutie nodig te hebben (b.v. 40 pixels van top-tot-teen volstaat niet) en het aantal momenten dat informatie op het gezicht bruikbaar is, is lager dan het aantal momenten dat b.v. een onderdeel van de jas zichtbaar is.

Een VCA oplossing voor het volgen van mensen is zinloos in een kleine omgeving. Daarom is het noodzakelijk voor dit onderwerp om de schaalbaarheid naar een grote omgeving te beschouwen. De beeldverwerking kan geïntegreerd zijn in de camera, waardoor het gedistribueerd wordt uitgevoerd, of het kan centraal worden uitgevoerd op een cluster van computers. Wanneer de centrale rekenkracht onvoldoende is om real-time alle camera's te verwerken zal alleen een zoekopdracht in een klein gebied en een kleine tijdsperiode kunnen worden ondersteund en snelle interactie is niet mogelijk. Om een uitbreiding naar een groter gebied mogelijk te maken, moet de toewijzing van rekenkracht aan camera's meebewegen met de zoektocht. Door berekeningen uit te voeren *in* de camera kan een schaalbare oplossing worden gecreëerd en kunnen centrale berekeningen worden geminimaliseerd. Veel moderne camera's claimen intelligente videoanalyse te ondersteunen. Het is gangbaar om in de camera bewegende objecten te kunnen detecteren en een compacte beschrijving (metadata) van het object te genereren. Als de kwaliteit van metadata hoog zou zijn, dan kan een grootschalige zoekactie worden uitgevoerd met snelle mens-machine interactie. Helaas is de kwaliteit van de

³ URL's: araliasystems.com, agentvi.com, 3vr.com, axxonsoft.com, boschsecurity.com, ipsotek.com, iOmniscient.com, dvtel.com, objectvideo.com

detectie en metadata in de huidige camera's vaak onvoldoende voor het volgen van mensen in een druk gebied.

Mensen kunnen worden gevolgd door te kijken naar uiterlijke kenmerken, zoals de dominante kleur of door een eenvoudig kleuren histogram te berekenen (Bosch, 2012). Er zijn diverse publieke *benchmark* datasets (Gray et al., 2008; Baltieri et al., 2011; Bialkowski et al., 2012) om te beoordelen hoe goed systemen in staat zijn om personen aan elkaar te verbinden die zijn waargenomen vanuit verschillende standpunten, belichtingen en camera's. De VIPeR dataset (Gray et al., 2008) is één van de datasets waar veel verschillende algoritmes op zijn getest (Farenzena et al., 2010; Hamdoun et al., 2008; Bouma et al., 2012), en waar ook actuele ontwikkelingen op worden bijgehouden (An et al., 2013; Ma et al., 2014). Uit de benchmark is gebleken dat een gespecialiseerd zoekalgoritme voor mensen veel beter is dan een eenvoudig kleuren histogram en vijf keer sneller dan het handmatig zoeken in een database (Bouma et al., 2012).

Enkele van deze systemen zijn niet alleen geëvalueerd op offline datasets met uitsnedes van personen, maar recent is ook gebleken dat hiermee live demonstraties konden worden gegeven in een realistische omgeving, wat wijst op een *technological readiness level* (TRL) van tenminste 6. Een zoekalgoritme moet daarvoor worden geïntegreerd in een verwerkingsketen waarbij ook mensen worden gedetecteerd. Een efficiënte architectuur is om in elk van de camera's real-time alle personen te detecteren en te volgen, zodat per persoon in een camera een compacte beschrijving kan worden gemaakt. Deze compacte beschrijving kan vervolgens worden gebruikt om een persoon die in één camera is waargenomen terug te vinden in andere camera's. Het gebruik van een dergelijk systeem leidt tot minder gemiste personen (Bouma, Baan et al., 2013).

3.3 Actieherkenning

Acties zijn enkelvoudige, kortstondige gedragingen, zoals lopen, rennen, vallen, rondhangen, of het achterlaten van een object. Herkenning hiervan is interessant voor detectie van specifieke situaties, zoals iemand die een terrein betreedt, een persoon die wegrent, iemand die onwel wordt, hangjongeren, of achtergelaten bagage (Ferryman et al., 2013). Diverse systemen zijn op de markt beschikbaar, die eenvoudige acties herkennen op basis van looppatronen. Er zijn hoofdzakelijk drie categorieën van systemen te onderkennen: detectie van aanwezigheid of richting van een persoon, zoals wanneer een persoon een bepaalde lijn passeert (voor perimeter beveiliging), detectie van afwijkende looppatronen (voor verdachte situaties), en detectie van specifieke looppatronen (voor b.v. rennen of vallen)⁴.

In de wetenschappelijke arena worden momenteel grote stappen gezet om complexere acties te herkennen (Poppe, 2010), zoals interpretatie van een collectie YouTube filmpjes van sport acties, acties in een realistische surveillance omgeving (Oh et al., 2011), muziek maken en filmpjes die door mensen thuis zijn opgenomen⁵. Er zijn veel verschillende kenmerken voorgesteld om zulke acties te herkennen. Een belangrijk kenmerk is beweging – gemodelleerd door *Spatio-Temporal Interest Points* (STIP) – waarmee de veranderingen van de positie van een arm of been kunnen worden gekarakteriseerd (Laptev, 2005). Dit is geschikt voor acties zoals gooien. Beweging blijkt als een algemeen kenmerk van acties erg onderscheidend op grote collecties videobeelden (Wang & Schmidt, 2013). Andere kenmerken van acties zijn de houding van het gehele lichaam (Felzenszwalb et al., 2010), wat geschikt is voor acties zoals een golf swing. Er zijn ook complexere acties, die een interactie tussen personen, met een object, of met de omgeving betreffen. Voorbeelden zijn respectievelijk vechten, een object op de grond zetten,

⁴ www.agentvi.com, brslabs.com, www.iomniscient.com

⁵ crcv.ucf.edu/ICCV13-Action-Workshop/

en een object begraven. Voor deze acties is een combinatie van kenmerken nodig, zoals de interactie zelf, en het betrokken object of de locatie in de omgeving, die samen kunnen worden geïnterpreteerd met patroonherkenning methoden (Burghouts, Schutte, Bouma et al., 2014).

Recente onderzoeken richten zich op het toepasbaar maken van actie herkenning, als een eerste beproeving richting operationele systemen. Het is interessant dat deze toepassingen zich baseren op verschillende typen camerasystemen. De waarnemingsystemen komen oorspronkelijk uit de *gaming*, waarvoor ook specifieke sensoren zoals *Kinect* en *ASUS Xtion Pro* zijn ontwikkeld. Deze sensoren produceren naast een kleurenbeeld ook een dieptebeeld, wat extra informatie geeft over hoe iemand zich beweegt in drie dimensies. De acties die nodig zijn om te kunnen interacteren met de game, zijn goed onderscheidbaar met dit type sensoren. Ook complexere acties met meerdere personen kunnen worden herkend (Koppula et al., 2013) en surveillance toepassingen, zoals indringerdetectie (Minar et al., 2013). Een ander specifiek toepassingsgebied zijn de camera's die op het lichaam gedragen worden, zogenaamde *body of wearable cams*. Dit heeft een vlucht genomen door de rage om zelf acties te kunnen opnemen, zoals tijdens skiën (bv. *GoPro*). Eerste proeven zijn gedaan met agenten, om extra bewijslast (of ontlastend bewijs) te kunnen verzamelen tijdens een optreden. Voor de lichaamsgedragen camera's worden specifieke methoden ontwikkeld die kunnen omgaan met de beweging die de camera zelf maakt, en tegelijk de beweging van de waargenomen persoon kunnen interpreteren (Ryoo & Matthies, 2013). Voor toezicht in de openbare ruimte met conventionele camera's is aangetoond op een publieke dataset (Weinland et al., 2007) dat de prestaties van actie herkenning significant toenemen indien meerdere camera's kunnen worden gebruikt (Burghouts et al., 2013).

3.4 Gedragsanalyse

Er zijn diverse producten beschikbaar voor het detecteren van de eerder genoemde eenvoudige gedragingen, acties en events. De meer complexere of subtiele gedragingen die zijn samengesteld uit meerdere acties over een langere tijd zijn moeilijker te detecteren (Gowsikhaa et al., 2012) en de technologie om dit uit te voeren is minder volwassen. Recent zijn een aantal methodes voorgesteld om automatisch afwijkend of verdacht gedrag te detecteren (Elhamod et al., 2013; Ouivirach et al., 2013; Bouma et al., 2014; Burghouts et al., 2014).

Het automatisch detecteren van zakkenrollers is bijvoorbeeld een uitdagend probleem voor VCA omdat zakkenrollers een drukke omgeving uitzoeken, omdat mensen rondlopen in een gebied dat niet kan worden afgedekt met één camera, omdat het aantal voorbijgangers veel groter is dan het aantal zakkenrollers, omdat de eigenlijke diefstal een zeer subtiele actie is en omdat bij zakkenrollers complexe interacties tussen personen plaatsvinden. Het zakkenrollen is een definitie op een hoog abstractieniveau. Dat niveau kan worden vertaald naar concretere fasen of kenmerken. Typische fasen bij het zakkenrollen zijn bijvoorbeeld: de omgeving observeren, communiceren met een handlanger, het slachtoffer omsingelen, de eigenlijke diefstal, overdracht van de buit aan een handlanger, en het verlaten van de scene (Rest et al., 2013). Om zakkenrollers te kunnen detecteren met automatische technieken wordt gebruik gemaakt van een combinatie van technologieën: detectie van mensen, het volgen van de mensen (dat is een beschrijving op laag abstractieniveau), berekening van kenmerken en analyse van looppatronen (beschrijving op midden niveau; bijvoorbeeld splitsen en samenkomen) en uiteindelijk zakkenrollerclassificatie (hoog abstractieniveau) (Bouma et al., 2014). Hierbij bleek dat o.a. de loopsnelheid en richtingsverandering indicatoren zijn die zakkenrollers onderscheiden van normaal winkelend publiek.

Voor het detecteren van dreigingen naar vrachtauto's is het van belang om naar een hoog niveau beschrijving van een dreiging te gaan. Een typische dreiging bestaat uit een observatie fase, een naderingsfase, een aanval en daarna weggaan. Hierbij is het de combinatie van de verschillende fasen die het tot een dreiging maakt. Om dit te realiseren is op laag niveau gekeken naar het

bereken van tracks en het bepalen van (STIP) bewegingskenmerken, op het midden niveau is dit vertaald naar acties en relaties tussen personen zoals nadering, en ten slotte wordt dit getoetst aan een complete beschrijving op hoog niveau (Burghouts, Schutte, Hove et al., 2014).

4. BIJDRAGE VAN VCA AAN TOEZICHT OP DE OPENBARE RUIMTE

In deze sectie wordt getoond hoe VCA technieken kunnen bijdragen aan toezicht in de openbare ruimte. Hiervoor worden de drie concrete cases uit Sectie 2 behandeld die elk zijn onderbouwd met recente resultaten. Het eerste gaat over het automatisch volgen van een verdachte na een incident (Sec. 4.1), het tweede over zakkenroller detectie (Sec. 4.2) en het derde over dreigingsdetectie op een parkeerplaats (Sec. 4.3).

4.1 Automatisch volgen van een overvaller na een overval

In het scenario waarbij de cameraoperator mag terugkijken in opgenomen video, kan na een melding eerst worden teruggedaan naar het moment dat de overvaller het plaats delict verlaat. Zonder VCA zal de operator een stukje video bekijken, en wanneer de overvaller het zicht van de ene camera verlaat, proberen over te schakelen naar een volgende camera. Als de camera's dichtbij elkaar staan is dit eenvoudig, maar als er blinde vlekken tussen camera's bestaan is dit lastig en tijdrovend. Er ontstaat onzekerheid over de looprichting en daardoor moeten meerdere camera's worden bekeken. De aanwezigheid van de verdachte in de videobeelden kan gemakkelijk worden gemist, zodat hij niet (op tijd) wordt gevonden, waardoor het geen heterdaad meer is, maar moet worden overgedragen aan opsporing. Als wel VCA wordt gebruikt kan een afbeelding van de verdachte in de eerste camera worden gebruikt om het systeem te laten zoeken in camera's in de omgeving. Hierdoor wordt het grote aantal aanwezige personen teruggebracht tot een veel kleiner aantal, en alleen de mensen die veel lijken op de verdachte worden gepresenteerd. Hierdoor kan een operator sneller de verdachte terugvinden en gemakkelijker blijven volgen tot de veiligheidsdiensten ter plaatse zijn. Uit een human-in-the-loop operator experiment in een druk winkelcentrum is gebleken dat een dergelijk systeem leidt tot 37% minder gemiste personen, wat een significante verbetering is (Bouma, Baan et al., 2013). De architectuur blijkt volwassen genoeg te zijn om live demonstraties mee uit te voeren, zoals bijvoorbeeld in Polen in het EU-FP7 project PROTECTRAIL (Marck et al., 2014), waarbij een persoon is teruggevonden na een (in scene gezet) incident met achtergelaten bagage.

4.2 Zakkenroller detectie in een winkelcentrum

Recent zijn experimenten uitgevoerd in het automatisch detecteren van zakkenrollers in een druk winkelcentrum (Bouma et al., 2014). Er zijn meer dan 20 zakkenroller incidenten opgenomen om de prestaties van een VCA systeem te schatten. Het systeem blijkt in staat om in het grote aantal mensen in het winkelend publiek een selectie van personen te maken die rijk is aan zakkenrollers. In meer dan 250.000 tracks van winkelend publiek kon het systeem vijf detecties per zakkenroller genereren met een sensitiviteit van 66% en slechts 0.5% loze alarmen (met één detectie per zakkenroller werd dit zelfs 95.6% sensitiviteit en 0.5% loze alarmen). Toch is het op dit moment nog onduidelijk of dit zal leiden tot een verhoging van de efficiëntie van cameraoperators. Aan de ene kant lijkt 0.5% weinig, maar in hetzelfde experiment bleek dit overeen te komen met 3 loze alarmen per minuut. De toepasbaarheid hangt onder andere af van de prioriteit van het vinden van zakkenrollers en de manier van gebruik van het systeem. Als alarmeringssysteem is het onbruikbaar door het aantal loze alarmen, maar als cameraselectiesysteem of als zoekstelsel zou het toegevoegde waarde kunnen hebben. Om duidelijkheid te krijgen over de toepasbaarheid zijn nieuwe experimenten noodzakelijk waarbij

wordt gemeten hoe efficiënt een operator kan werken met en zonder het systeem. Doordat dit concept niet werkt op basis van uiterlijkheden is de impact op privacy minimaal.

4.3 Dreigingsdetectie op een parkeerplaats

In het EU-FP7 project ARENA (Waern et al., 2014) is gewerkt aan het vroegtijdig detecteren van een dreiging voor een geparkeerde vrachtauto of zijn chauffeur. Uitgangspunt was dat het surveillance systeem op de vrachtauto gemonteerd is, zodat de dreigingsdetectie mogelijk is op elke parkeerplaats. De gebruikte techniek gaat uit van het detecteren en volgen van personen rond een vrachtauto, en het beschrijven van hun activiteiten (Andersson et al., 2013) (Burghouts, Schutte, Hove et al., 2014).

Essentiële technische componenten om een dergelijk dreiging scenario te kunnen herkennen is het volgen van personen over langere tijd en het beschrijven van activiteit, zoals rondhangen, lopen en openbreken van de laaddeur (Sanromà et al., 2014). Om een systeem bruikbaar te maken is het belangrijk om de dreigingen goed te kunnen onderscheiden van normaal achtergrond gedrag. Hierbij helpt onder meer een context service, die voor elke parkeerplaats kennis heeft over de locatie van b.v. toiletten en een wegrestaurant, zodat op basis daarvan normale loop patronen ingeschat kunnen worden. Onderscheid tussen de bestuurder van een vrachtauto en anderen is belangrijk, en kan bijvoorbeeld worden gedetecteerd door de chauffeur een *app* op zijn smartphone te geven die steeds zijn locatie aan het systeem doorgeeft.

Het systeem is getest op offline videodata (uit Reading, UK) die publiek beschikbaar is gesteld in de PETS benchmark⁶. Daarnaast is recent een live demonstratie gegeven op een andere locatie (in Parijs, FR). Tijdens deze demonstratie zijn enkel camera's op de vrachtauto gebruikt. Indien op een parkeerplaats vaste CCTV infrastructuur aanwezig is kan dit de systeemprestaties verbeteren. Ook kunnen meerdere vrachtauto's met een dergelijk systeem hun informatie delen om een beter beeld van de situatie te krijgen.

5. CONCLUSIE

In dit artikel zijn drie scenario's behandeld: het volgen van een overvaller direct na de overval, het vroegtijdig signaleren van zakkenrollerij in een winkelcentrum en het detecteren van een overval op een vrachtwagen op een parkeerplaats. De behoefte van cameraoperator bestaat om in elk van deze situaties efficiënter en pro-actiever te kunnen omgaan met de camerabeelden. Daarnaast hebben wij laten zien hoe recente technologische ontwikkelingen op het gebied van VCA die behoeftes kunnen helpen vervullen. VCA kan helpen door personen na een incident sneller terug te vinden of door dreigingen waar te nemen voordat het incident plaatsvindt. Hierdoor kan een cameraoperator vroeg in het proces een sturende rol nemen door collega's te attenderen op (aanstaande) ongewenste situaties of door hen te begeleiden naar de huidige locatie van een verdachte. Eenvoudige VCA systemen – zoals de beveiliging van een leeg bedrijventerrein – worden breed ingezet, maar de meer complexe dreigingsdetectie op basis van lange-termijn gedrag van personen wordt nu nog niet operationeel gebruikt. De eerste resultaten van deze VCA systemen in realistische omgevingen zijn bemoedigend, en geven aanleiding tot vervolgstappen in de vorm van beproeving in een grootschalige operationele omgeving, zoals bijvoorbeeld het complete centrum van een stad of een groot vliegveld. Een dergelijke vervolgstap naar een operationele omgeving vereist niet alleen een schaalbare technische implementatie, maar ook een integratie in werkomgeving, aanpassing van de mens-machine interface en mogelijk een aanpassing van de werkprocessen.

⁶ <http://www.cvg.rdg.ac.uk/PETS2014/a.html>

ACKNOWLEDGEMENT

Dit artikel werd mede mogelijk gemaakt door de projecten ‘Passieve sensoren’ (top sector HTSM), EU FP7 PROTECTRAIL, EU FP7 ARENA, ‘Watching people security services’ (WPSS) en CORTEX (DARPA Mind’s Eye). De opnames voor de zakkenroller dataset was een gezamenlijke inspanning met het TNO onderzoeksprogramma ‘Verdacht gedrag’ (MinV&J). De auteurs danken het “Centrum voor Innovatie en Veiligheid” (CIV) en het “Diensten Centrum Beveiliging” (DCB) in Utrecht voor de fieldlab faciliteiten en ondersteuning.

LITERATUURLIJST

- [1] Alyuz, N., Gokberk, B., Spreuwers, L., Veldhuis, R., Akarun, L. (2012). Robust 3D face recognition in the presence of realistic occlusions. *IAPR Int. Conf. Biometrics*, 111-118.
- [2] An, L., Kafai, M., Yang, S., Bhanu, B., (2013) Reference-based person re-identification, *IEEE AVSS*, 244-249.
- [3] Andersson, M., Patino, L., Burghouts, G., e.a., (2013) Activity recognition and localization on a truck parking lot, *IEEE AVSS*, 263-269.
- [4] Baltieri, D., Vezzani, R., Cucchiara, R., (2011). 3DPES: 3D people dataset for surveillance and forensics. *Proc. ACM Human Gesture and Behavior Understanding*, 59-64.
- [5] Bellanova, R., Vermeulen, M., Gutwirth, S., et al. (2011). Deliverable 1.1 - Smart Surveillance - State of the Art, <http://www.sapientproject.eu>
- [6] Bialkowski, A., Denman, S., Lucey, P., et al., (2012) A database for person re-identification in multi-camera surveillance networks. *Proc. Int. Conf. Digital Image Comp. Techn. Appl.*, 1-8.
- [7] Bosch, (2012) Intelligent video analysis IVA 5.5; Operation manual, *Grasbrunn (Germany): Bosch Security Systems*.
- [8] Bouma, H., Borsboom, S., Hollander, R. den, Landsmeer, S., Worring, M., (2012) Re-identification of persons in multi-camera surveillance under varying viewpoints and illumination, *Proc. SPIE 8359*.
- [9] Bouma, H., Burghouts, G., Penning, L., e.a., (2013) Recognition and localization of relevant human behavior in videos, *Proc. SPIE 8711*.
- [10] Bouma, H., Baan, J., Landsmeer, S., Kruszynski, C., Antwerpen, G., Dijk, J., (2013) Real-time tracking and fast retrieval of persons in multiple surveillance cameras of a shopping mall, *Proc. SPIE 8756*.
- [11] Bouma, H., Vogels, J., Aarts, O., e.a., (2013) Behavioral profiling in CCTV cameras by combining multiple subtle suspicious observations of different surveillance operators, *Proc. SPIE 8745*.
- [12] Bouma, H., Baan, J., Burghouts, G., e.a., (2014) Automatic detection of suspicious behavior of pickpockets with track-based features in a shopping mall, *Proc. SPIE 9253*.
- [13] Burghouts, G., Eendebak, P., Bouma, H., Hove, J., (2013) Improved action recognition by combining multiple 2D views in the bag-of-words model, *IEEE AVSS*, 250-255.
- [14] Burghouts, G., Schutte, K., Hove, R. et al. Instantaneous threat detection based on a semantic representation of activities, zones and trajectories, *Signal Image and Video Processing SIVP*, (2014).

- [15]Burghouts, G., Schutte, K., Bouma, H., Hollander, R., (2014) Selection of negative samples and two-stage combination of multiple features for action detection in thousands of videos, *Machine Vision and Applications MVA* 25(1), 85-98.
- [16]Dollar, P., Belongie, S., Perona, P., (2010) The fastest pedestrian detector in the west, *BMVC* 2(3).
- [17]Don, B., (2012). *Thema Integrale Veiligheid Vraaggestuurd Programma 2012-2014 VP Security Bijstelling 2013*. Rapport TNO-2012-R10591, Den Haag: TNO.
- [18]Elhamod, M., Levine, M. (2013) Automated real-time detection of potentially suspicious behavior in public transport areas. *IEEE Trans. Intelligent Transportation Systems* 14(2), 688-699.
- [19]Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M., (2010) Person re-identification by symmetry-driven accumulation of local features, *IEEE CVPR*, 2360-2367.
- [20]Felzenszwalb, P., Girshick, R., McAllester, D., Ramanan, D., (2010) Object detection with discriminatively trained part-based models, *IEEE Trans. PAMI* 32(9), 1627-1645.
- [21]Fenn, J. (2013) Hype cycle for emerging technologies 2013, Gartner Research, <http://www.gartner.com/newsroom/id/2575515>
- [22]Ferryman, J., Hogg, D., Sochman, J., e.a., (2013) Robust abandoned object detection integrating wide area visual surveillance and social context, *PRL* 34(7), 789-798.
- [23]Flight, S. (2013) *Cameratoezicht in Nederland*, Den Haag: WODC, Min. V&J.
- [24]Gong, S., Cristani, M., Yan, S., Loy, C., (2014) *Person re-identification*. London: Springer.
- [25]Gowsikhaa, D., Abirami, S., Baskaran, R., (2012) Automated human behavior analysis from surveillance videos: a survey, *AI Review*.
- [26]Gray, D., Tao, H., (2008). Viewpoint invariant pedestrian recognition with an ensemble of localized features, *Europ. Conf. Computer Vision* (1), 262-275.
- [27]Hamdoun, O., Moutarde, F., Stanculescu, B., Steux, B., (2008) Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences, *IEEE Distr. Smart Cameras*.
- [28]Hu, N., Bouma, H., Worring, M., (2012) Tracking individuals in surveillance video of a high-density crowd, *Proc. SPIE* 8399.
- [29]Koppula, H.S., Gupta, R., Saxena, A. (2013) Learning Human Activities and Object Affordances from RGB-D Videos, *Int. Journal Robotics Research*.
- [30]Laptev, I. (2005) On space-time interest points, *Int. Journal of Computer Vision*, 64 (2/3).
- [31]Little, S., Clawson, K., Mereu, A., et al. (2013) Identifying and addressing challenges for search and analysis of disparate surveillance video archives, *Imaging for Crime Detection and Prevention*.
- [32]Ma, B., Su, Y., Jurie, F., (2014) Covariance descriptor based on bio-inspired features for person re-identification and face verification, *Image and Vision Comp.* 32(6-7), 379-390.
- [33]Marck, J., Bouma, H., Baan, J., Oliveira-Filho, J., Brink, M., (2014) Finding suspects in multiple cameras for improved railway protection, *Proc. SPIE* 9253.
- [34]Minar, J., Riha, K., Tong, H. (2013). Intruder detection for automated access control systems with Kinect device. *IEEE Int. Conf. Telecomm. Signal Processing*.
- [35]Oh, S., Hoogs, A., Perera, A., (2011). A large-scale benchmark dataset for event recognition in surveillance video. *Computer Vision and Pattern Recognition*, 3153-3160.
- [36]Ouvirach, K., Gharti, S., Dailey, M., (2013). Incremental behavior modeling and suspicious activity detection. *Pattern recognition* 46(3), 671-680.
- [37]Phillips, P., O'Toole, A. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*, 32(1), 74-85.

- [38]Poppe, R. (2010). A survey on vision-based human action recognition. *Image and vision computing*, 28(6), 976-990.
- [39]Rest, J. van, Bovenkamp, E., Eendebak, P. Baan, J., Munster, R., (2009) Sensors and tracking crossing borders, *Proc. Safety Security Systems Europe*.
- [40]Rest, J. van, Grootjen, F., Grootjen, M., e.a., (2013) Requirements for multimedia metadata schemes in surveillance applications for security, *Multimedia Tools and Appl.*, (2013).
- [41]Rest, J. van, Boonstra, D., Everts, M., e.a., (2014) Designing Privacy-by-Design, *Privacy Technologies and Policy LNCS 8319*, 55-72.
- [42]Rest, J. van, Nunen, A. van, Roelofs, M. (2014) *Afwijkend Gedrag*, Rapport TNO-2014-R10987, Den Haag: TNO.
- [43]Roelofsen, G., (2003) Method and System and Data Source for Processing of Image Data, Patent WO03/010728/A1
- [44]Ryoo, M.S., Matthies, L. (2013) First-person activity recognition: what are they doing to me? *IEEE CVPR*.
- [45]Satta, R., Fumera, G., Roli, F., (2012) A general method for appearance-based people search based on textual queries, *ECCV LNCS 7583*, 453-461.
- [46]Sanromà, G., Patino, L., Burghouts, G., e.a., (2014) A unified approach to the recognition of complex actions from sequences of zone-crossings, *Image Vision Comp.* 32 (5), 363-378.
- [47]Thida, M., Yong, Y., Climent-Perez, P., et al. (2013). A literature review on video analytics of crowded scenes. *Intelligent Multimedia Surveillance*, 17-36.
- [48]Vera-Rodriguez, R., Tome, P., Fierrez, J, Ortega-Garica, J. (2014). Comparative analysis of the variability of facial landmarks for the forensics using CCTV images. *Image and Video Techn.*, 453-461.
- [49]Waern, A., Andersson, M., Petersson, H., (2014) Multi-sensory surveillance for moving vehicles, SPIE newsroom.
- [50]Wang, H., Schmid, C. (2013) Action Recognition with Improved Trajectories, *IEEE ICCV*.
- [51]Weinland, D., Boyer, E., Ronfard, R.. (2007). Action recognition from arbitrary views using 3D exemplars. *ICCV*.