Járás alapú személyazonosítás és cselekvésfelismerés LiDAR szenzorokkal

Gálai Bence, Benedek Csaba

Gépi Érzékelés Kutatólaboratórium, Magyar Tudományos Akadémia,
Számítástechnikai és Automatizálási Kutatóintézet
{vezetéknév.keresztnév}@sztaki.mta.hu

Absztrakt. Cikkünkben új algoritmusokat mutatunk be személyek járás alapú biometrikus azonosítására és különböző cselekvések felismerésére forgó többszenzoros LiDAR rendszerek méréseit felhasználva. Eljárásunk képes a videofelügyeleti alkalmazásokhoz illeszkedő valószerű jelenetekből kinyerni a felismeréséhez szükséges jellemzőket, feltételezve, hogy több személy egyidejűleg mozog a helyszínen egymást gyakran keresztező útpályákat követve, és kölcsönös kitakarások valamint zajként jelentkező háttérmozgások is hatással vannak a megfigyelésre. Mivel a szakirodalomban elérhető nyilvános tesz-adathalmazok nem alkalmasak a módszerünk kiértékelésére, létrehoztunk és publikáltunk egy új LiDAR alapú járás és cselekvés adatbázist, ami 30 percnyi kézzel kiértékelt LiDAR-os méréssorozatot tartalmaz összesen 28 különböző tesztszemély szerepeltetésével. Eredményeink bizonyítják, hogy módszerünk képes elvégezni a megfigyelt területet időlegesen elhagyó, majd később visszatérő személyek újrafelismerését, valamint öt jellegzetes cselekvés (lehajolás, karórára nézés, telefonálás valamint egy- és kétkezes integetés) hatékony megkülönböztetését.

1. Bevezető

[1]Az automatizált videofelügyeleti alkalmazások fontos szerepet játszanak napjainkban a közbiztonság megteremtésében. Felhasználásuk széles skálán mozog egyszerű behatolásellenőrzést végző rendszerektől kezdve a terrorizmus elleni harc komplex feladatainak megvalósításáig. A megfigyelési folyamat során az egyik központi feladat a személyazonosság megállapítása. Számos elterjedt biometrikus jellemző, például ujjlenyomat vagy írisz-kód vizsgálata nem jöhet szóba ezekben az esetekben, mivel elemzésük csak a megfigyelt személy hozzájárulásával, a mérőeszközzel való fizikai kontaktus útján történhet. Bár az arcfelismeréshez már nem feltétlenül szükséges a célszemély közreműködése, a legtöbb esetben csak közelről lehetséges elegendően jó minőségű arcfelvételeket készíteni, valamint problémákat jelentenek az előforduló takarások, elmosódások, és a rossz szögből készített fotók. A fenti jellemzőkkel szemben a járás vizsgálata hatékony

[1] Ezzel a cikkel Gálai Bence pályázik a Kuba Attila díjra. A cikkben közölt eredmények eredetileg angol nyelven, az IEEE Transactions on Circuits and Systems for Video Technology folyóiratban jelentek meg [1].

alternatívát nyújthat, hiszen a járás alapú felismeréshez elegendő távolról megfigyelni a személyeket, így valós eseményeket monitorozó kültéri megfigyelő rendszerekben is lehetőség nyílik az alkalmazására. A járás, mint biometrikus jellemző felhasználhatóságát biztonságtechnikai rendszerekben már évtizedek óta vizsgálják [2]. A szakirodalomban számos hatékony vizuális úton megfigyelhető járásjellemzőt ismertetnek [3–5], amelyek segítségével különböző nyilvános teszthalmazokon sikeresnek bizonyult az azonos személyekhez tartozó járásminták automatikus összerendelése. A gyakorlati felhasználás lehetséges forgatókönyve az úgynevezett *gyenge* biometrika, ahol az egyéni jellemzőket nem próbálják meg például nagy körözési adatbázisokból történő személyazonosításra felhasználni, *csupán* azt várják el, hogy a helyszínen jelenlévő korlátozott számú ember *újrafelismerését* tegyék lehetővé, ha távozás után rövid időn belül visszatérnek a kamerák által megfigyelt területre, vagy egy másik kamera területére sétálnak át. Ezek a feladatok kritikus elemét képezik a hosszútávú cselekvéselemzésnek, valamint növelik a rövidtávú alakzatkövetés megbízhatóságát is [6].

A járásvizsgálat szakirodalmi módszereinek többsége hagyományos *optikai* kamerák felvételein végzi a felismerést. Az egykamerás rendszerek hátránya, hogy korlátos a látóterük, valamint a kinyert jellemzők erősen függenek a nézeti iránytól. Ezekben az esetekben tipikusan szükséges minden személy esetén különböző haladási irányú tanító-felvételek készítése, ami nem realiszikus elvárás egy videofelügyeleti környezetben. A többkamerás rendszerek előnye, hogy nagy területet képesek belátni, valamint átfedő kameranézetek esetén származtatott (sztereo) 3D információkon keresztül nézetfüggetlen jellemzők szintetizálására is alkalmasak. Hátrányuk azonban, hogy a kamerákat gyakran előzetesen kalibrálni kell, és a kalibrációt kis elmozdulások esetén újból meg kell ismételni, így a kitelepítésük és működtetésük drága és bonyolult folyamat. Optikai kameráknál szintén problémát okoznak a változó fényviszonyok, melyek a mérések minőségére és a képfelismerő algoritmusokra is közvetlen hatással vannak. Erre a problémára egy lehetséges alternatív megoldást kínálnak az aktív fénnyel működő time-of-flight (ToF) kamerák és különböző mélységszenzorok (Microsoft Kinect), ezek azonban fizikai korlátaik miatt egyelőre szinte kizárólag viszonylag kis (néhány méter átmérőjű) területet megfigyelő beltéri alkalmazásokban használhatók fel. Hátrányuk azonban, hogy a kamerákat gyakran előzetesen kalibrálni kell, és a kalibrációt kis elmozdulások esetén újból meg kell ismételni, így a kitelepítésük és működtetésük drága és bonyolult folyamat. Optikai kameráknál szintén problémát okoznak a változó fényviszonyok, melyek a mérések minőségére és a képfelismerő algoritmusokra is közvetlen hatással vannak. Erre a problémára egy lehetséges alternatív megoldást kínálnak az aktív fénnyel működő time-of-flight (ToF) kamerák és különböző mélységszenzorok (Microsoft Kinect), ezek azonban fizikai korlátaik miatt egyelőre szinte kizárólag viszonylag kis (néhány méter átmérőjű) területet megfigyelő beltéri alkalmazásokban használhatók fel.

Nagyobb dinamikus helyszínek megfigyeléséhez lehetőséget nyújtanak a közelmúltban elterjedt forgó többszenzoros LiDAR (FT-LiDAR, Light Detection and Ranging) lézerszkennerek. FT-LiDAR fő előnye, hogy képes nagy pontosságú 2.5D méréssorozatokat rögzíteni kültéri helyszíneken függetlenül a külső meg-

világítási körülményektől, ugyanakkor a mért adatok térbeli sűrűsége ritka és egyenetlen, az időbeli felbontás pedig szintén alacsonyabb az optikai és ToF szenzoroknál elérhető értékeknél. Az FT-LiDAR-ok felhasználása biometrikus azonosításra így nem nyilvánvaló feladat, amely tudomásunk szerint az itt ismertetett munkánk [1, 7] előtt nem rendelkezett szakirodalmi referenciákkal.

A videofelügyeleti rendszerek fontos feladata a személyazonosság meghatározásán túl bizonyos cselekvések automatikus jelzése. A szakirodalomban számos különböző módszert ajánlottak meghatározott cselekvések felismerésére, ám ezek gyakran csak egy adott specifikus feladat különböző eseményeinek azonosítását oldják meg, például a [8] Kinect alapú eljárásban egy táncjáték egyes lépéseit választják el egymástól. Mélységszenzorok használata felmerül további alkalmazásokban is (például MoCap, vagy MSR-Action3D [9]), azonban a szenzor korlátai kültéri működés során ezekben az esetekben is jelentkeznek.

A módszerek kiértékelésénél fontos tényező a teszthalmaz megválasztása. A korábbi járásfelismerést végző módszereket általában egyszerű tesztkörnyezetben készített adathalmazokon validálták: például a MoBo [10] adatbázis tesztalanyai jól megvilágított szobában egy futógépen sétáltak, míg a mozgást több nagy felbontású kamera rögzítette párhuzamosan. Ilyen ideális mérési körülményekre nem számíthatunk a valós életben, ami kérdéseket vet fel az eredmények alkalmazhatóságával kapcsolatban valódi rendszerekben. Munkánk során ezért hangsúlyt fektettünk a realisztikus tesztkörnyezetet biztosítására: különböző jeleneteket rögzítettünk egy Velodyne HDL-64E FT-LiDAR szenzorral, melyekben egyszerre 3-8 személy sétált szabadon egy udvaron, akik gyakran takarták egymást a LiDAR szemszögéből. A méréseinken végzett felismerési eredmények kiértékelésére létrehoztunk és publikáltunk egy új LiDAR alapú járás- és cselekvésminta adatbázist (SZTAKI Gait-and-Activity, SZTAKI-LGA)[2].

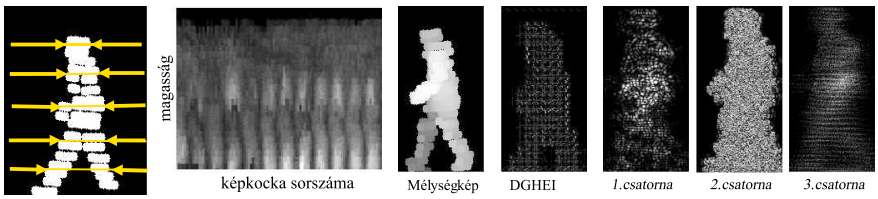
A cikk további részének tartalmi kivonata a következő: a 2. fejezet a járás- és cselekvésfelismerés szakirodalmában fellelhető eddigi eredményekről ad áttekintést. A 3. fejezetben a LiDAR alapú járásfelismerő módszerünket ismertetjük és értékeljük ki, összehasonlítva az új eljárást korábbi szakirodalmi módszerekkel. A cselekvésfelismerést megvalósító eljárásunkról a 4. fejezetben adunk részletes leírást. Végül az 5. fejezetben összefoglaljuk munkánk eredményét.

2. Szakirodalmi előzmények

A szakirodalomban közölt járás- és cselekvésfelismerő eljárások többnyire a *modell alapú* vagy a *megjelenés alapú* megközelítést követik. A *modell alapú* módszerek csontvázakat, vagy egyéb egyszerűsített geometriai modelleket illesztenek az emberekre, majd a modellek konfigurációs paramétereit felhasználva végzik el a felismerési feladatot. Ilyen paraméterek lehetnek például az egyes testrészek (fej, lábak, stb.) méretei, vagy a végtagok által bezárt szögek változásának dinamikája. Ezeknek a módszereknek azonban részletgazdag és jó minőségű bemeneti adatokra (képekre vagy pontfelhőkre) van szükségünk, melyekből a

[2] Az adatbázisunk elérhetősége: `http://web.eee.sztaki.hu/i4d/SZTAKI-LGA-DB`.

kívánt paraméterek biztonsággal kinyerhetőek [11]. A fentiekkel ellentétben a *megjelenés alapú* módszerek nem az egyes testrészeket, hanem a teljes testről származó mérést használják, mint jellemzőt, és abból nyernek ki tipikusan szemantikai tartalommal nem rendelkező, de az automatikus osztályozáshoz felhasználható leírókat. Mivel esetünkben a FT-LiDAR mérések ritka pontfelhőket szolgáltatnak, melyeken a takarások következtében akár teljes testrészek (pl. fej vagy kar) is hiányozhatnak, a modell alapú leírók használata bizonytalan. A *megjelenés alapú* eljárások közül gyakori kiindulási adat a személyek sziluettképe, ezért a következőkben bemutatunk néhány kapcsolódó módszert.



(a) Sziluett lenyomat (b) DGHEI jell. (c) CGCI jellemző térkép

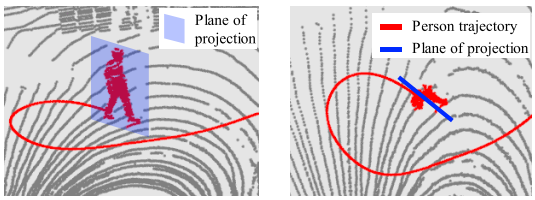
1. ábra: A sziluett lenyomat [5] módszer jellemző képei (a), valamint a DGHEI [4] (b) és CGCI [12] (c) módszerek leírói LiDAR adatokon.

2.1. A sziluett lenyomat módszer

Az eredetileg optikai képeken definiált sziluett lenyomat módszer [5] alapja az az észrevétel, hogy a sziluett szélességének változása eltérő magasságokban eltérő dinamikát mutat. A módszer működése közben minden mintavételi ciklus során eltároljuk az árnykép szélességét különböző magasságokban. Az időbeli követést kihasználva a járás lenyomata egy vektorfolyam lesz, amely szemléletesen ábrázolható grafikusan is (1(a). ábra). A lenyomatok összehasonlítását a dinamikus idővetemítés (Dynamic Time Warping, DTW) algoritmussal végzi a módszer, majd a döntés a vetemített minta- és tesztjelek összevetése alapján történik.

2.2. Mélység-gradiens hisztogram energiakép

A mélység-gradiens hisztogram energiaképén alapuló (Depth Gradient Histogram Energy Image, DGHEI) módszert a Kinect szenzor mélységképeire fejlesztették ki [4]. Az egyes képkockákon mélységi gradiensek kinyerése történik, ezekből lokális hisztogramokat számolnak, melyeket egy-egy teljes járáscikluson belül átlagolnak (1(b). ábra). Az energiaképeken ezután dimenziócsökkentést hajtanak végre főkomponens analízissel (Principal Component Analysis, PCA), a döntést pedig egy hagyományos *legközelebbi szomszéd* (nearest neighbor) osztályozó végzi.



(a) Madártávlati nézet (b) Felülnézet

2. ábra: Sziluett vetítés: (a) egy megfigyelt személy és a vetítési síkja madártávlati nézetből (b) vetítési sík felülnézetből, a simított trajektóriára érintőlegesen.

2.3. Színezett járás-görbületi kép

A színezett járás-görbületi kép módszerét (Color Gait Curvature Image, CGCI) a DGHEI-hez hasonlóan Kinect szenzorokhoz vezették be [12]. Az eljárás Gaussi elmosást, átlagolt görbületi és pontfelhősűrűséggel kapcsolatos jellemzőket nyer ki, amiből egy háromcsatornás CGCI nevű képet származtat. A dimenziócsökkentést 2D koszinusz transzformáció és 2D-PCA alkalmazásával végzik el az egyes csatornákon külön-külön, végül a jellemzővektorok különbségét képezik az egyes komponenseket eltérően súlyozva. A CGCI kép három csatornáját az 1(c). ábra szemlélteti.

2.4. Véletlen foglaltsági minták

A véletlen foglaltsági minta (Random Occupancy Pattern, ROP) jellemzőt cselekvések pontfelhőkön történő osztályozására vezették be [13]. A módszer a pontfelhősorozatból egy bináris négydimenziós reprezentációt készít, melyben az egyes térrács-elemek *1* értéket vesznek fel, amennyiben az adott voxelen belül található legalább egy alakzatpont, *0* értéket ellenkező esetben. A cselekvések felismerését egy szupport vektor gép (SVM) végzi, melynek tanító adatai a 4D-s tömb különböző pozícióin különböző voxelmérettel kinyert foglaltsági jellemzők.

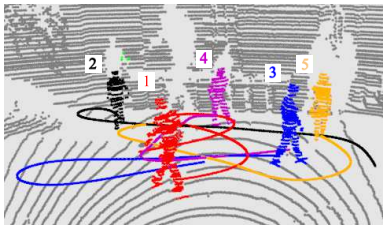
2.5. 3D mozgásfolyam

A Kinect szenzor által kinyert színezett pontfelhőből lehetőség nyílik háromdimenziós mozgásfolyam számítására, amit [14] módszere egy 64 voxeles 3D rácson belül értékel ki. Az eljárás az egyes voxeleken belül összegzi minden pont elmozdulását, és normálja azokat. 30 időegységen keresztül kinyerve az elmozdulásokat készít egy jellemzővektort, melyen a cselekvések felismerését a legközelebbi szomszédok módszerrel végzi.

3. Járásfelismerés a LiDAR méréssorozatokon

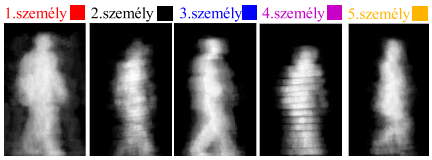
Ebben a fejezetben bemutatjuk az általunk javasolt járásfelismerő módszert, melynek kiindulási alapja az eredetileg optikai képekre definiált *járás energiakép*

(Gait Energy Image, GEI). Az egyes járókelőkhöz tartozó pontfelhőrészletek kinyerésére a [6]-ban bemutatott többcélpontos követő eljárást használtuk. A pontfelhőrészleteket a trajektória aktuális érintője által meghatározott függőleges síkra vetítettük (2. ábra), majd morfológiai műveletekkel összefüggő sziluetteket hoztunk létre. Így elértük, hogy a további lépésekben többnyire a járásfelismeréshez előnyös oldalnézeti sziluettképekkel tudunk dolgozni [3, 5].



(a) Tesztalanyok a LiDAR felvételen

(b) Vetített sziluettek



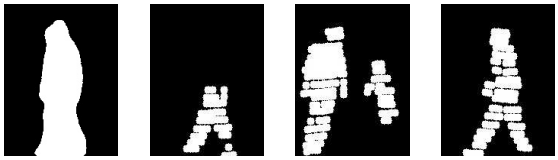
(c) Generált LGEI-k

3. ábra: Pillanatfelvétel az egyik többszereplős tesztjelentből. (a) Követés eredménye, (b) szereplők levetített sziluettjei, (c) az egyes szereplőkhöz tartozó aktuális LGEI-k.

3.1. A LiDAR alapú járás energiakép módszere

A Han és Bhanu által bevezetett járás energiakép [3] a járásciklusokon belül átlagolt sziluettekből számított képi jellemző. Dimenzióját a szerzők főkomponens analízissel (PCA), majd pedig többszörös diszkrimináns analízissel (multiple discriminant analysis, MDA) csökkentették, kompakt jellemzővektorokhoz jutva, amelyeknek a tanító és teszt minták között számolt abszolút távolságát használták az osztályozáshoz. A szerzők a valós mintákon kívül bevezettek úgynevezett szintetikus sablonokat, amelyekkel növelték a tanító mintahalmazt. A felismerést külön-külön végezték a valós és szintetikus képeken, majd ezek eredményeit fuzionálták a közös döntéshozatalhoz. Az általunk bevezetett módszer, a LiDAR alapú járás energiakép (LGEI) a GEI átlagoló ötletén alapszik, de több helyen fontos módosításokat kellett alkalmaznunk a tesztkörnyezetünkhöz való adaptálásához.

Az első módosítás, hogy elhagytuk az előzetes járásciklus-detekciós lépést, mivel ez az alacsony térbeli és időbeli felbontású adatsorozaton nem bizonyult



(a) Maszk (b) Rossz1 (c) Rossz2 (d) Jó

4. ábra: A sziluettszelekciós maszk és a szűrés eredményei

stabilnak, valamint egy-egy járásciklus adatai gyakran túl kevés információt szolgáltattak az azonosításhoz. A ciklusok helyett fix számú ($l = 60$) egymást követő képkockán történt az átlagolás, amely a gyakorlatban nagyjából 3-4 járás ciklusnak felelt meg LGEI mintánként.

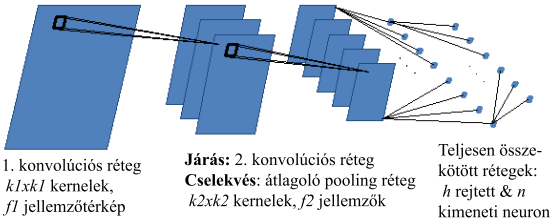
Másrészt mivel számos kitakarás és jelentős háttér zaj jelent meg a felvételeinken, készítettünk egy előfeldolgozó algoritmust a hiányos vagy gyenge minőségű sziluettek kiszűrésére. A detektált sziluettek globális átlagolásával generáltunk egy sziluett burkoló maszkot (4(a) ábra), majd a felismerés során a vizsgálatokból kizártuk azokat az észlelt sziluetteket akiknek az átfedése a burkolóval adott arányértéknél alacsonyabb volt. A 4. ábrán látható további példák a sziluett kiválasztó eljárás kimenetét szemléltetik. A későbbi tesztek során bebizonyosodott, hogy a rossz minőségű sziluettek (a teljes adathalmaz 10-12%-a) kiszűrése átlagosan 5%-kal növelte az eljárásunk teljesítményét.

Osztályozáshoz egy konvolúciós neurális háló (CNN) és egy hagyományos többrétegű perceptron (MLP) együttesét használatuk, hasonlóan [15] módszeréhez. A felhasznált konvolúciós háló struktúrája és paramétereinek a leírása az 5. ábrán látható. Míg a CNN bemenete egy LGEI kép, az MLP-é annak egy PCA-MDA előfeldolgozáson keresztül kinyert jellemző vektora. Mindkét esetben a bemeneti sziluettképek 20×15-ös méretűvé történő leskálázása után összegzett LGEI térképeket alkalmaztuk.

A neurális hálózataink *tanh* aktivációs függvényt használtak, melynek kimenete a [-1,1] tartomány eleme, így az i-dik személyhez tartozó háló elvárt kimeneti értékének a tanítás során 1-est adtunk, ha hozzá tartozó minta került a bemenetre, ellenkező esetben pedig -1-et. A felismerési fázisban a betanított hálók kimenetei o_{cnn} és o_{mlp} szintén a [-1,1] tartományon vettek fel az értékeiket, míg a CNN-MPP együttesének a kimenetét a két komponens kimeneteinek maximumaként határoztuk meg: $o = max(o_{cnn}, o_{mlp})$. Egy adott G teszt LGEI osztályozásához kiszámoltuk legjobban illeszkedő $i_{max} = argmax_i(o)$ indexet, és G-t azonosítottuk az i_{max}-edik személyként, amennyiben $o^{i_{max}} > 0$. Ha nem történt sikeres azonosítás, G-t mint új személyt jeleztük a rendszernek.

3.2. A járásfelismerés módszerünk kiértékelése

A járásfelismerési modult a SZTAKI-LGA adatbázis 10 tesztsorozatán értékeltük ki. Az egyes felvételeken 3-8 személy szerepel, amint az udvaron egymás útját



Paraméterek	k1	f1	k2	f2	h	n
Járásfelismerés	3	5	7	9	98	NP
Cselekvésfelismerés	7	5	2	-	20	1

5. ábra: A konvolúciós neurális háló (CNN) struktúrája. NP a tanító halmazban szereplő alanyok számát jelöli.

gyakran keresztezve szabadon sétálnak. A felvétel közepén valamennyien elhagyják a helyszínt, majd tetszőleges sorrendben visszaérkeznek a látótérre és folytatják a sétát. Kiértékelésünkben minden felvételen az első fázist használtuk tanítóminták gyűjtésére, majd a felismerést a második fázisban kinyert jellemzők alapján végeztük. Az eredményesség méréséhez a személyek sikeres *újrafelismerésének* arányszámát határoztuk meg.

A tesztek során a javasolt LiDAR alapú járás energiakép (LGEI) módszert a 2. fejezetben ismertetett a *sziluett lenyomat* (SP+DTW), DGHEI, és CGCI eljárásokkal vetettük össze. A módszerek kiértékeléséhez 100 tanító és 200 ettől különböző tesztmintát generáltunk minden egyes személyhez, majd a felismerést az egyes tesztminták külön-külön történő felhasználásával végeztük. Így egy adott N személyel felvett szekvencián $200 \cdot N$ független teszteset keletkezett. A helyes újrafelismerések arányai a 1. táblázatban hasonlíthatók össze.

Annak ellenére, hogy a CGCI [12] eljárás jó eredményeket ért el Kinect felvételeken, a módszer jellegéből adódó előnyeit a Velodyne ritkább pontfelhőire nem lehetett kihasználni. Jól látható a 1. táblázatban, hogy az összes módszer közül a legrosszabb eredményeket itt értük el az alacsony sűrűségű LiDAR adatokon.

A szélesség vektor alapú SP+DTW teszteknél a jó minőségű a sziluettkontúrok meglétének szükségességét tapasztaltuk. Az eljárás egyedül az első szekvencián (`Winter0`) teljesített jól, melyen közel teljesen összefüggő és jó minőségű alakzatokat láthattunk, azonban a sziluettek minőségének romlásával és a kitakarások számának növekedésével az SP+DTW teljesítménye gyorsan romlott.

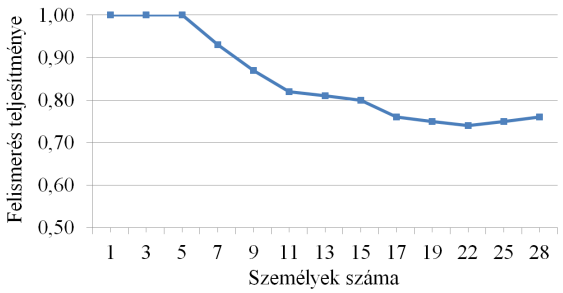
A DGHEI [4] bizonyult a második legjobb járásleírónak, amely csak az LGEI módszertől maradt el 5% százalékkal. A megfigyelt eredmény egyik közvetlen oka, hogy a DGHEI eljárás a GEI mélység-gradiensekkel és hisztogram átlagolással kiterjesztett változatának tekinthető, azonban ellentétben a két nagyságrenddel nagyobb sűrűségű Kinect pontfelhőkön végzett sikeres kísérletekkel [4], az alacsony felbontású LiDAR adatokon nem tudjuk jól kihasználni az eljárás potenciális előnyeit.

1. táblázat: Az összehasonlított módszerek újrafelismerési arányai. N a személyek számát jelöli.

Scene	N	SP+DTW	DGHEI	CGCI	LGEI		
					CNN	MLP	Mix
`Winter0`	4	0.96	0.97	0.36	0.94	0.98	**0.99**
`Winter1`	6	0.33	0.89	0.27	0.85	0.90	**0.95**
`Spring0`	6	0.64	0.81	0.32	0.91	0.95	**0.98**
`Spring1`	8	0.33	0.59	0.20	0.63	0.66	**0.70**
`Summer0`	5	0.39	0.97	0.40	0.99	0.95	**1.00**
`Summer1`	6	0.33	0.83	0.29	0.77	0.95	**0.95**
`Summer2`	3	0.33	0.98	0.53	0.96	0.99	**0.99**
`Summer3`	4	0.50	**0.94**	0.32	0.94	0.93	**0.94**
`Summer4`	4	0.25	**0.95**	0.27	0.91	0.90	0.91
`Summer5`	4	0.50	**0.80**	0.32	0.77	0.74	**0.80**
Átlag	5	0.46	0.87	0.33	0.87	0.90	**0.92**

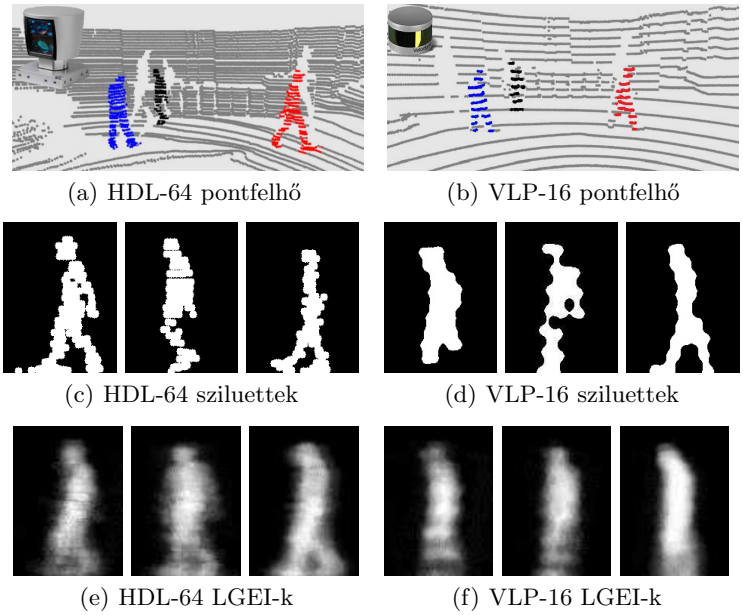
Az LGEI módszernél először külön teszteltük az MLP és a CNN hálózatok kimeneteit, majd ezek után a kettő együttesét. A 1. táblázat utolsó három oszlopában láthatjuk a kapcsolódó eredményeket. Az MLP és a CNN külön-külön egymással versengve jól teljesítettek az egyes szekvenciákon, együttes használatuk pedig további javulást eredményezett. Amint azt [7]-ben részleteztük, az LGEI eljárás MLP-CNN osztályozóval szintén jobban teljesített, mint a [3]-ban javasolt egyszerű vektordifferenca alapú döntés.

Az 1. táblázatból szintén kiolvasható, hogy a nagyobb létszámú teszt szekvenciákon valamelyest romlik a felismerés teljesítménye, hiszen a járókelők számának növekedésével több kitakarás keletkezik, ami rontja a LGEI képzéshez használt adatok minőségét. Szintén érdekes kérdés az LGEI leíró információs kapacitásának a meghatározása, azaz közel ép sziluettek kinyerését feltételezve a felismerési arány alakulásának vizsgálata a mintahalmaz növelésének függvényében. Mivel kísérleteink során összesen 28 személy járását rögzítettük, a különböző szekvenci-



6. ábra: Felismerési eredmények alakulás az adatbázis fokozatos bővítésekor.

ákról kigyűjtött adatokat felhasználva elvégeztünk egy tesztet, ahol a résztvevők számát folyamatosan növelve (2, 3, ..., 28) mértük a felismerési eredményeket. A 6. ábrán láthatjuk az LGEI módszer felismerési teljesítményét a személyek számának függvényében. A kezdetben csökkenő görbén 17-28 személy esetén egy 75% körüli stagnálást vehetünk észre.



(a) HDL-64 pontfelhő (b) VLP-16 pontfelhő

(c) HDL-64 sziluettek (d) VLP-16 sziluettek

(e) HDL-64 LGEI-k (f) VLP-16 LGEI-k

7. ábra: HDL-64 és VLP szenzorral készített pontfelhők (1. sor), vetített sziluettek (2. sor), valamint a hozzájuk tartozó LGEI-k (3. sor) összehasonlítása.

3.3. Kísérletek kisebb felbontású szenzorral

Az eddigi kísérleteinkben használt nagy méretű Velodyne HDL-64E szenzor mellett teszteltük a módszer teljesítményét a Velodyne kompakt 16 sugaras (VLP-16) szenzorával is. A VLP-16 pontfelhői jelentősen ritkábbak, így a kinyert sziluettek és LGEI képek minősége is lényegesen alacsonyabb, amit a 7. ábra is szemléltet.

Az két szenzort összehasonlító kísérletek során az előzőekkel azonos körülmények között történtek a mérések. A VLP-16 eredményességének felméréséhez összesen 5 mérést végeztünk, ahol a felvételeket mindkét szenzorral párhuzamosan rögzítettük. A helyes újrafelismerések arányai a 2. táblázatban találhatók. Az N3/1, N3/2 és N3/3 szekvenciákon három személy volt jelen, és a VLP-16 szenzor közel volt a sétálókhoz, míg az F4 és F5 szekvenciákon négy, illetve

2. táblázat: A helyes újrafelismerések aránya a HDL-64E és a VLP-16 szenzorok méréseit összehasonlító kísérletben.

Szekvencia	HDL-64	VLP-16
N3/1	0.96	0.81
N3/2	0.85	0.84
N3/3	0.93	0.81
F4	0.79	0.68
F5	0.93	0.54

öt személy sétált a VLP-16 szenzort nagyobb távolságra helyezve (a HDL-64 szenzort nem mozdítottuk el a kísérletek során). Az N3 mérés egyes szegmensein a tesztelés keresztkiértékeléssel történt, tehát az N3/2 teszteléséhez az N3/1 tanító mintáit használtuk fel amit a 2. tábla 1. sora mutat, stb. Az F4 illetve F5 szekvenciákat az eddigiekhez hasonlóan tanító és teszt szegmensekre osztottuk a kiértékeléshez. Jól látható, hogy bár a HDL-64E-t használó megoldás minden szekvencián felülmúlja a VLP-16 eredményeit, a kisebb LiDAR is jól teljesített a szenzorhoz közeli méréseknél (80% fölötti eredmények). Az alakzatoktól való távolság növelésével viszont jelentősen romlott a kisebb felbontású szenzor adataira támaszkodó döntés teljesítménye.

4. Cselekvések felismerése

(a) Videokép referencia (b) LiDAR felhő

8. ábra: Választot képkocka négy különböző cselekvéssel az eseményfelismeréshez használt tesztfelvételekről

A személyek biometrikus (újra)felismerése mellett központi feladat különböző események megkülönböztetése a videofelügyeleti rendszerekben. Ebben a fejezetben eljárást mutatunk be különböző ritkán előforduló események felismerésére, a személyek cselekvéseit vizsgálva.

A szakirodalomban ismertetett cselekvésfelismerő eljárások között több pontfelhő alapú módszert is találhatunk, amelyek például a 2. fejezetben is részletezett foglaltsági mintákat [13], irányított főkomponensek hisztogramjait [16], vagy

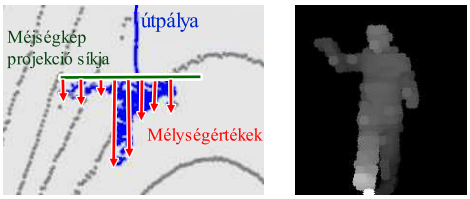
3D mozgásfolyamot [14] nyernek ki a ponthalmazokból. Annak ellenére, hogy sűrű pontfelhőkön (pl. Kinect) jól működnek ezek az eljárások, a LiDAR által szolgáltatott ritka pontfelhőkön a jellemzők kinyerése szűk keresztmetszetet jelent a használhatósághoz. Különböző jellemzőkkel történt kísérleteink alapján úgy döntöttünk, hogy az eseményanalízis során ismét a képátlagoló megközelítéssel fogunk élni. Az egyszerű sétáláson kívül öt felismerésre váró cselekvést választottunk ki: *lehajolás*, karórára pillantás (*karóra*), *telefonálás*, *integetés*, és kétkezes integetés (*integetés2*). Kísérletünk egy pillanatképe látható a 8 ábrán.

4.1. A cselekvésfelismerés megvalósítása

Eljárásunkat az LGEI alapú járásanalízisnél is használt módszerek inspirálták, azonban ismét több kulcsfontosságú módosítással kellett élnünk. Míg a járást oldalnézetből tudjuk a lehető legjobban megfigyelni, a fent említett cselekvések esetén elölnézetből jobban tudjuk a jellemző mozgást vizsgálni. Ebből a megfontolásból a pontok vetítési síkját a cselekvés felismeréshez az aktuális trajektóriára merőleges síknak választottuk, ahogyan ez a 9 ábrán is látható (ez a sík tehát merőleges a járásanalízisnél alkalmazott vetítési síkra).

Megfigyeltük továbbá, hogy egyes események, mint például a telefonálás vagy az integetés esetén a mozgás egy mélységképen jobban értelmezhető, mint a bináris sziluetteken (pl. a kéz a test elé kerül). Ennek érdekében a sziluettek levetítésénél nem bináris képet generáltunk, hanem a 9(a) ábrán szemléltetett vetítéssel egy mélységképeket hoztunk létre. Az így kinyert jellemzőképeket időben átlagoljuk $k = 40$ képkockán (a tesztjeink során mért átlagos cselekvési idő), létrehozva az úgynevezett Átlagolt Mélységkép jellemzőt (Averaged Depth Map, ADM). Az öt esemény ADM-jei láthatók a 10 a)-f) ábrákon.

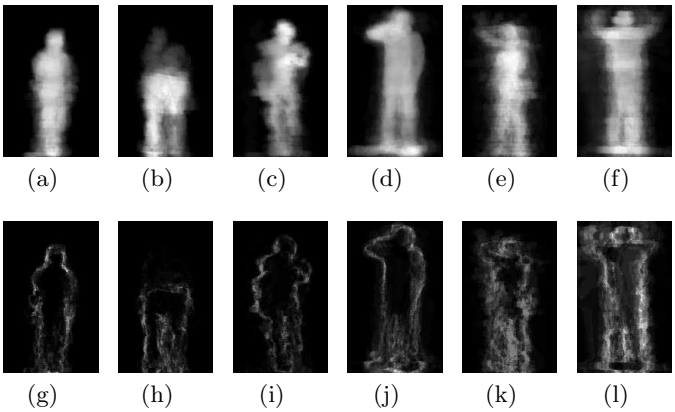
Amíg az ADM-ek az egyes cselekvések során a jellemző testtartást rögzítik, a mozgások során érdemes lehet kinyerni annak dinamikáját is. Az integetés például gyors rövidebb mozdulatok sorozata, amely a felsőtesten nagy változásokat eredményez az egyes képkockák között. Ennek a jelenségnek a kihasználására vezettünk be egy újabb jellemzőt, az Átlagolt XOR képet (Averaged XOR, AXOR). Egy AXOR kép az időben egymást követő frontális bináris sziluetteken végzett XOR műveletek négyzetes átlagolásából keletkezik, tehát az AXOR kép



(a) Vetítés (b) Mélységkp

9. ábra: Elölnézeti projekció, valamint az így készített mélységkép megjelenítése. A vetítési sík merőleges a személy trajektóriájára.

a hirtelen mozgásokat emeli ki az egyes képrégiókban. A séta és az öt esemény AXOR térképei láthatók a 10 g)-l) ábrákon.



10. ábra: *ADM* és *AXOR* képek a (a, g) sétálás, (b, h) lehajlás, (c, i) karórára pillantás, (d, j) telefonálás, (e, k) integetés és (f, l) kétkezes integetés (wave2) cselekvésekről.

Az ADM és AXOR jellemzőtérképek értelmezése hasonlóan történt a járáselemzés során bemutatott megoldásokhoz. A *lehajlás*, *karóra*, *telefonálás*, *integetés* és *integetés2* események mindegyikére két konvolúciós neurális hálót (CNN) tanítottunk, egyet az ADM, egyet pedig az AXOR jellemzőre. Hasonlóan a korábbiakhoz, itt is egy kis, 4 rétegű hálót terveztünk, melynek bemenetei a leskálázott 20×16 pixeles ADM és AXOR képek voltak. A tanítás során a pozitív találathoz 1.0, a negatívhoz -1.0 értékeket adtunk meg a háló elvárt kimenetének. Szintén felvettünk a tanításhoz *negatív* (semmilyen különleges cselekvéshez sem tartozó) mintákat a *sétálást* rögzítő videorészekről. Felismeréskor a CNN-ek kimenetei a -1.0 és 1.0 értékek között helyezkednek el, és egy teszt mintát akkor fogadunk el az adott eseményként, ha a megfelelő ADM alapú és AXOR alapú hálók kimenetei mindketten egy ν küszöb feletti értéket adnak ($\nu = 0.6$-ot használtunk). Amennyiben az öt közül egy cselekvést sem ismerünk fel, nem küldünk jelzést (pédául *sétáló* embereknél).

Megjegyezzük, a fenti osztályozási módszer többszörös találatokat is engedélyez (pl. az egy- és kétkezes integetést egyszerre). Itt feltételeztük, hogy a megfigyelési rendszerben a fő cél a hibásan figyelmen kívül hagyott találatok minimalizálása, míg az esetleges hamis riasztásokat az operátorok ellenőrizhetik.

4.2. A cselekvésfelismerés kiértékelése

A cselekvésfelismerést végző eljárás teszteléséhez összesen 10 cselekvési szekvenciát használtunk keresztkiértékeléses megközelítéssel. Az egyes szekvenciákon az

események felismeréséhez a többi kilenc szekvencia kézzel címkézett cselekvés mintáival tanítottuk be a két konvolúciós hálót. Mind a tanítás során, mind a felismerésnél figyelembe vettünk egyszerű sétálásról vett (negatív) mintákat is. Az így beiktatott *sétálásos* ADM/AXOR minták száma arányos volt a többi esemény mintáinak előfordulási gyakoriságával.

3. táblázat: A cselekvésfelismerés igazságmátrixa.

Észlelt→ / Valódi↓	*Lehajlás*	*Karóra*	*Telefon*	*Integetés*	*Integetés2*	*FN*	*FP*
Lehajlás	85					3	
Karóra		37	1		4	11	3
Telefon		5	36	2	2	5	6
Integetés			4	44	5	5	3
Integetés2			5	9	31	1	2

Az eseményfelismeréses kísérletünk igazságmátrixa (*confusion matrix*) látható a 3. táblázatban. A mátrix i-edik sora és j-edik oszlopa az i-edik cselekvés azon mintáinak számát jelöli, melyeket j-edik cselekvésként ismertünk fel. Az utolsó két oszlopban a hamis negatív (FN) és hamis pozitív (FP) detekciós értékeket láthatjuk az i-edik cselekvéssel összefüggésben a következő definíciókat alkalmazva:

- FN: i-edik sorszámú cselekvés típus olyan előfordulásainak száma, ahol az eseményt egyik cselekvésként sem ismertük fel
- FP: azoknak az esetekben a száma, melyeket hamisan az i-edik cselekvésként jeleztünk, miközben valójában egyik figyelt cselekvés sem történt meg

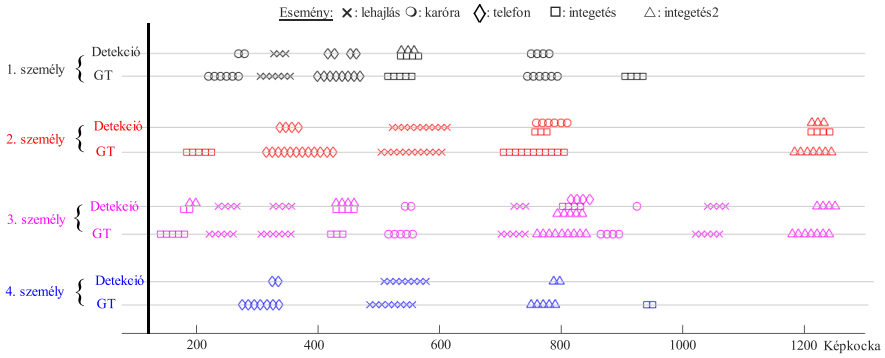
Látható, hogy a *lehajlás*, *telefonálás*, *integetés* és kétkezes integetés (*integetés2*) cselekvéseket szinte mindig eltalálta a rendszerünk (FN$\leq$ 5). A karórára pillantáshoz (*karóra*) összesen 11 hamis negatív minta tartozik, mivel a kitakarások és a háttérzajok miatt a kis karmozgásokat több esetben nem sikerült észlelni. A *lehajlás* eseménynek a felismerése bizonyult a legkönnyebbnek, amit soha nem tévesztett össze más cselekvéssel az eljárásunk. Ugyanakkor az *integetés* és *integetés2* eseményeket viszonylag gyakran összekeverte a rendszer. A 4. táblázat az egyes cselekvések összesített precizitás (precision) és felidézés (recall) értékekeit mutatja.

Érdemes még kiemelni, hogy összességében alacsony volt a hamis pozitív találatok száma (Σ_iFP kevesebb mint 5%-a az összes tényleges cselekvésnek), tehát a rendszer ritkán küldött hamis riasztásokat sétáló emberektől. Ez az előnyös tulajdonság jól nyomon követhető a 11. ábra idővonalán, ahol egy kültéri szekvencián mutatjuk be a különböző érzékelt eseményeket. A vízszintes tengelyen az idő, míg a függőleges tengelyen a négy teszt személy különböző cselekvései láthatók ikonokkal megjelenítve (az ábra tetején látható magyarázattal).

4. táblázat: Az egyes cselekvések precizitás/felidézés értékei.

	Lehajlás	Karóra	Telefon	Integetés	Integetés2
Minták száma	88	53	50	58	46
Pecizitás	1.00	0.82	0.69	0.76	0.70
Felidézés	0.97	0.77	0.88	0.90	0.97

Az egyes személyeknél a *Detekció* sor mutatja a cselekvésfelismerő rendszer által jelzett különböző eseményeket, míg a *GT* sor a kézzel címkézett valós események előfordulásait jelzi. Látható, hogy szinte minden cselekvést sikerült felismerni egy kis időkésleltetéssel, ami az ADM és AXOR jellemzők generálása miatt elkerülhetetlen. A cselekvések között eltelt időben a tesztalanyok a korábbi kísérletekhez hasonlóan, egymást keresztező útvonalakon szabadon sétáltak.



11. ábra: Cselekvésfelismerés eredményei az egyik tesztszekvencián (4 személy). *Detekció*: azok a képkockák melyeken módszerünk adott cselekvést észlelt, *GT* (Ground Truth): a valódi cselekvésekhez tartozó kézzel címkézett képkockák.

5. Konklúzió

Cikkünkben a forgó többszenzoros LiDAR lézerszkennerek felhasználhatóságát vizsgálatuk videofelügyeleti rendszerekben. Új módszereket vezettünk be személyek járás alapú azonosítására és különböző cselekvések felismerésére, amelyek megbízhatóan működnek valódi körülményeket szimuláló kültéri felvételeken is, felkészülve több személy együttes jelenlétére, gyakori kitakarásokra és különféle zajhatásokra. A járásanalízist végző módszer hatékonyságát bemutattuk egy kompakt LiDAR szkennert felhasználva is. A kutatáshoz kapcsolódó további

demonstrációk a szerzők laboratóriumának honlapján találhatók[3]. A projektet a Magyar Tudományos Akadémia Bolyai János Kutatási Ösztöndíja, és a Nemzeti Kutatási, Fejlesztési és Innovációs Alap (NKFIA #K_120233) támogatta.

Irodalom

1. C. Benedek, B. Gálai, B. Nagy, and Z. Jankó, "Lidar-based gait analysis and activity recognition in a 4D surveillance system," *IEEE Trans. on Circuits and Systems for Video Technology*, 2016, In Press, DOI: 10.1109/TCSVT.2016.2595331.
2. M. P. Murray, "Gait as a total pattern of movement," *American Journal of Physical Medicine*, vol. 46, no. 1, pp. 290–333, 1967.
3. J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. and Mach. Intel.*, vol. 28, no. 2, pp. 316–322, Feb 2006.
4. M. Hofmann, S. Bachmann, and G. Rigoll, "2.5D gait biometrics using the depth gradient histogram energy image," in *Int'l Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, Sept 2012, pp. 399–403.
5. A. Kale, N. Cuntoor, B. Yegnanarayana, A.N. Rajagopalan, and R. Chellappa, "Gait analysis for human identification," in *Audio- and Video-Based Biometric Person Authentication*, vol. 2688 of *LNCS*, pp. 706–714. 2003.
6. C. Benedek, "3D people surveillance on range data sequences of a rotating Lidar," *Pattern Recognition Letters*, vol. 50, pp. 149–158, 2014, Special Issue on Depth Image Analysis.
7. B. Gálai and C. Benedek, "Feature selection for Lidar-based gait recognition," in *Int'l Workshop on Computational Intelligence for Multimedia Understanding*, Prague, Czech Republic, October 2015, pp. 1–5, IEEE.
8. Michalis Raptis, Darko Kirovski, and Hugues Hoppe, "Real-time classification of dance gestures from skeleton animation," in *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, New York, NY, USA, 2011, SCA '11, pp. 147–156, ACM.
9. W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, June 2010, pp. 9–14.
10. R. Gross and J. Shi, "The CMU Motion of Body (MoBo) Database," Tech. Rep. CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA, June 2001.
11. Chew-Yean Yam and Mark S. Nixon, *Gait Recognition, Model-Based*, pp. 633–639, Springer US, Boston, MA, 2009.
12. J. Tang, J. Luo, T. Tjahjadi, and Y. Gao, "2.5D multi-view gait recognition based on point cloud registration," *Sensors*, vol. 14, no. 4, pp. 6124–6143, 2014.
13. *Robust 3D Action Recognition with Random Occupancy Patterns*. Springer, October 2012.
14. M. Munaro, G. Ballin, S. Michieletto, and E. Menegatti, "3D flow estimation for human action recognition from colored point clouds," *Biologically Inspired Cognitive Architectures*, vol. 5, pp. 42 – 51, 2013.
15. D. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in *International Joint Conference on Neural Networks (IJCNN)*, July 2011, pp. 1918–1921.

[3] `http://web.eee.sztaki.hu/i4d/demo_surveillance.html`

16. H. Rahmani, A. Mahmood, D. Q Huynh, and A. Mian, "HOPC: Histogram of oriented principal components of 3d pointclouds for action recognition," in *European Conf. Computer Vision*, vol. 8690 of *LNCS*, pp. 742–757. 2014.