

Képi alapú többszintű környezetelemzés  
(Image based multi-level environment analysis)

**Válasz Prof. Hajdu Andrásnak**

az MTA doktori tézis opponensének

Benedek Csaba

2020. június 9.

Mindenekelőtt köszönöm Prof. Hajdu András szakértő, gondos és alapos bírálatát. A bírálat megjegyzéseire és kérdéseire az alábbiakban adok részletes választ.

- 1. A 3.2. szakaszban a szerző ismerteti, hogy egy képpár egyszerű regisztrációjához Fourier eltolási tételen alapuló technikát [118] tekintettek. Kísérleteztek más módszerekkel is erre a célra, összességében mennyire befolyásolhatta az eredményt a választott regisztrációs technika?*

Igen, az említett Fourier eltolási tételen alapuló módszer (FCS<sup>1</sup>) mellett kipróbáltunk egy klasszikus pixelmegfeleltetéseken alapuló megoldást (PCH<sup>2</sup>) is a képeinkre, és az összehasonlítás eredményeit az [54] kutatási riportunkban bemutattuk.

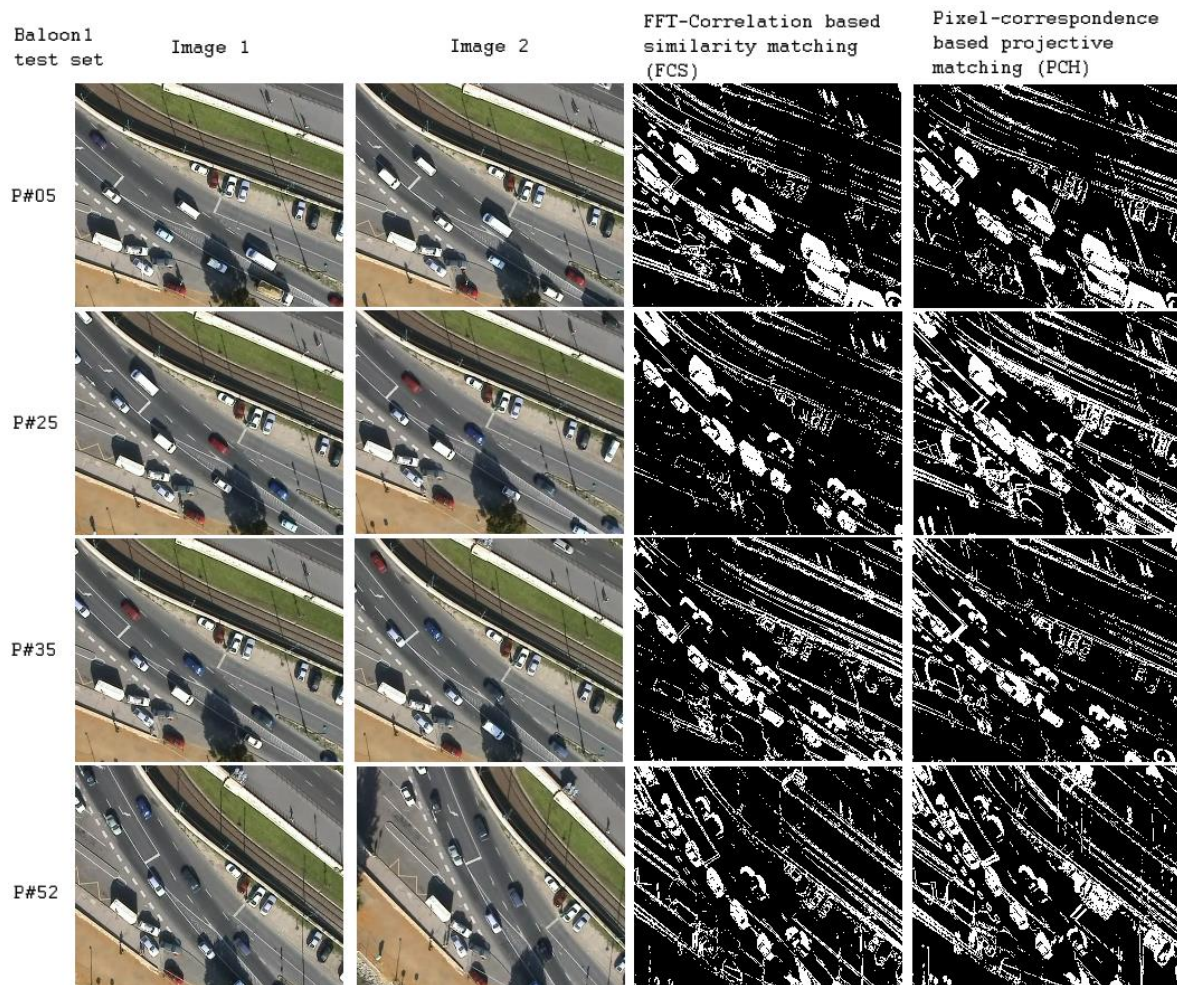
Röviden összefoglalva, a PCH megközelítésben első lépésként sarokpontok között kerestünk megfeleltetéseket a piramidális Lucas-Kanade eljárással. A hibás (outlier) megfeleltetéseket a RANSAC módszerrel szűrtük, majd egy optimális homográfia transzformációt illesztettünk a megmaradt pontpárookra a visszavetítés hibájának minimalizálásával. Az FCS és a PCH hasonló minőségű eredményeket adott, de az FCS-t valamennyivel hatékonyabbnak találtuk, ugyanis PCH alkalmazásakor túl sok mozgó objektum esetén a RANSAC nem mindig tudta hatékonyan kiszűrni az autókon megjelenő sarokpontokat, és a homográfiabecslés hibás eredményt adott.

Az 1. ábra demonstrálja az FCS valamint a PCH módszerekkel történő regisztráció eredményeit. Látható, hogy egyik eredmény sem okoz tökéletes fedést, ezt egy homográfia transzformációtól nem is várhatjuk el. Azonban míg a PCH-nál nagyobb területű összefüggő hibás régiók is megjelennek, amit az L<sup>3</sup>MRF modell sem tud eltüntetni, az FCS hibái főleg a statikus régiók éleinek környékére korlátozódnak, amit a javasolt módszerünk hatékonyan kompenzál a korrelációs jellemző segítségével.

FCS<sup>1</sup>: FFT-Correlation based similarity transform

PCH<sup>2</sup>: Pixel-correspondence based homography matching

[54] Cs. Benedek, T. Szirányi, Z. Kato and J. Zerubia, "A three-layer MRF model for object motion detection in airborne images", Research Report 6208, INRIA Sophia Antipolis, France, June 2007  
Elérhetőség: <https://eprints.sztaki.hu/4897/> Releváns rész: 5-7 oldalak



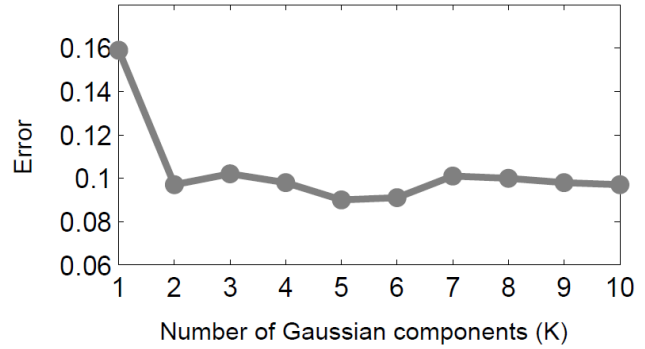
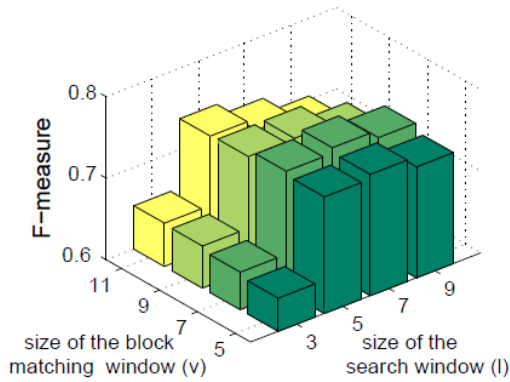
1. ábra az FCS és a PCH alapú előregisztráció eredményeinek kvalitatív összehasonlítása

2. Az L3MRF módszer kvantitatív kiértékeléséhez F-score került felhasználásra. Más mérőszámok felmerültek (ROC, AUC alapon)? Esetleg a ROC (jellegű) görbéknél megfigyelhető valamilyen erősebb vagy gyengébb viselkedési rész?

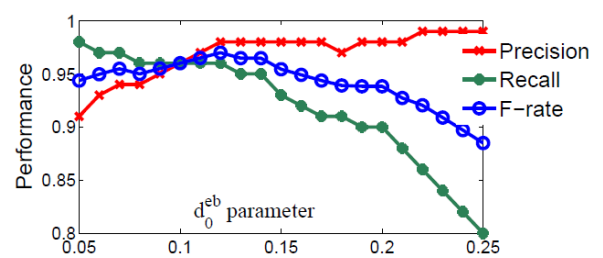
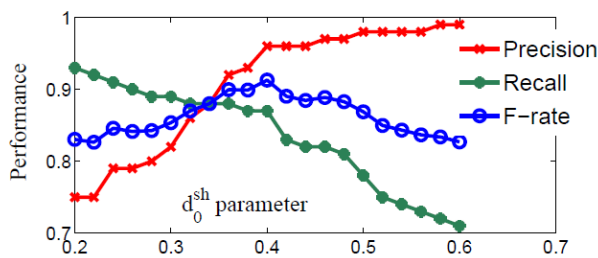
ROC és AUC görbéket éppen nem számoltunk, de a dolgozatban bemutatott több modell esetén is vizsgáltuk a teljesítmény/hibamértékek alakulását különböző paraméterek változtatásának hatásra.

A  $L^3MRF$  modell esetén a hiperparaméterek (korrelációs és mozgó ablak mérete) változásának függvényében ábrázoltuk az F-score-t (2. ábra, balra), megállapítva, hogy egy széles tartományban robusztusan viselkedik. A CXM modellnél az eljárás intenzitásmodelljének hibáját rajzoltuk ki a Gaussi keverékeloszlás különböző komponensszámai függvényében (2. ábra, jobbra). Hasonló kísérleteket a többkamerás emberdetekció (4.1 altézis) hiperparamétereire is elvégeztünk, amit később a 11. kérdésnél részletezek. A mobil lézerszkenneléses méréseket osztályozó eljárásnál szintén végeztünk a hiperparaméter-térben egy átfogó keresést (részletek a 18. kérdésnél). Az EMPP modell tesztelése során több alkalmazás több különböző paraméteréhez ábrázoltuk a precizitás (precision, Pr), fedés (recall, Rc) és F-score görbéket a releváns dinamikatarományaikon (3. ábra).

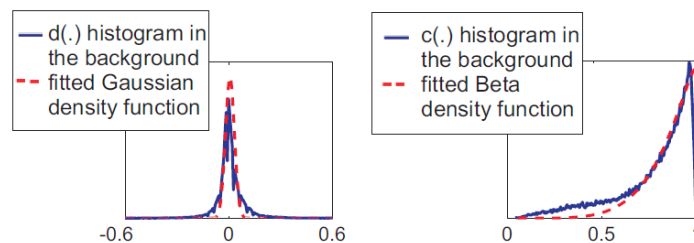
Visszatérve az  $L^3MRF$  modellt érintő eredeti kérdésre, itt az adatfüggő jellemzők  $f_d(\cdot)$ , és  $f_c(\cdot)$  eloszlásait meghatározó paramétereket a tanítómintákból Expectation-Maximization eljárással egzakt módon határoztuk meg (4. ábra), ezért itt nem merült fel a Pr-Rc görbék vizsgálata, természetesen ez elvégezhető lenne a kapott eloszlásparaméterek perturbációt megvizsgálva.



2. ábra: Balra: az L<sup>3</sup>MRF model teljesítményértékei a hiperparaméterei különböző beállításai mellett (v korrelációs ablak és l keresési ablak oldalhosszai) [12] Jobbra: a CXM eljárás intenzitásmodelljének hibája a Gaussi keverékeloszlás különböző komponensszámai függvényében [13]



3. ábra Az EMPP [2] modell teljesítménye két különböző alkalmazás, két választott paramétere változtatása során: balra az épületdetekció, árnyékleírójának paramétere, jobbra: Lidar alapú járműdetekció külső evidencia jellemző paramétere



4. ábra Az L<sup>3</sup>MRF model adatfüggő energiatajaiban használt eloszlások meghatározása a tanító mintákból származtatott hisztogramok alapján Expectation Maximization eljárásokat alkalmazva

3. A B.1 algoritmus a korábbi Modified Metropolis Dynamics (MMD) algoritmus egy kiterjesztése. Az MMD egy rögzített küszöböt (a kódban Tau) használ az elfogadási feltételben, ami azt eredményezi, hogy egy bizonyos hőmérséklet alatt az algoritmus lényegében sztochasztikusból determinisztikussá válik. (Pl. szimulált hűtésnél egy véletlen érték generálódik minden iterációban, így nem válik determinisztikussá teljesen soha.) Felmerül a kérdés, hogy miután az algoritmus determinisztikus lesz, szükséges-e továbbra is minden egyes csomópontra kiszámítani a célfüggvény értéket, vagy lehetne esetleg gyorsítani az optimalizációt azzal, hogy azon csomópontok címkéi, amelyek az előző valamennyi iterációban nem változtak rögzítésre kerülnek? (Esetleg egy nagyobb szomszédság figyelembe vételével.)

A tapasztalataim azt mutatták, hogy az MMD eljárás a vizsgált feladatokon (és képeken) a determinisztikus fázisba érve néhány iteráción belül konvergál (hasonlóan a klasszikus Iterated

conditional modes – ICM – algoritmusához), ezért a determinisztikus optimalizációs rész további gyorsításának igénye nem merült fel a munkám során. Az MMD igazi erőssége az ICM típusú teljesen determinisztikus algoritmusokhoz képest az, hogy nagy zajszint mellett is hatékony eredményt szolgáltat, azonban ez épp a sztochasztikus fázis időbeli elhúzódsát igényli, amit a következő demonstráció is bemutat:

<https://users.itk.ppke.hu/kep/materials.html#mrfdemo>

Mindemellett a Bíráló által javasolt megoldás valóban kézenfekvő, és könnyen is implementálható, nagyobb képeken/nagyobb szomszédság figyelembe vételekor előnyös lehet.

#### *4. Felmerült a 3.3. szakaszban (Task 2) bemutatott bővített modell első feladatára (Task 1) alkalmazhatósága? Összehasonlítható a két modell egy feladaton?*

Felmerült és proof-of-concept szinten kipróbálásra is került a CXM modell alkalmazása az objektummozgások detekciójára, azonban ez a megoldás az egyszerűbb (így kisebb számítási igényű)  $L^3$ MRF modell eredményeit nem szárnyalta túl, ezért nem éreztem indokoltnak a további alkalmazását.

Az  $L^3$ MRF modell sikerét az első feladatban az okozta, hogy ott két olyan alkalmazásspecifikus jellemzőt találtunk, melyek közül

- egyik sem okoz számottevő mennyiségű hamis negatív találatot (a mozgó alakzatokat mindketten nagy biztonsággal jelzik)
- egyenként gyakran, de egyszerre ritkán okoznak hamis pozitív találatot

A fenti megfigyelés miatt a fúziós szabály a két jellemző által javasolt mozgásmaszk ÉS kapcsolatban történő figyelembevétele, amit a simaság fenntartáshoz Potts-féle szomszédsági tagokkal ötvöztünk.

A hosszútávú változásdetekciós feladatnál a fenti egyszerű fúzió nem működött, ezért vezettem be az egy réteggel összetettebb CXM modellt: itt három jellemzőt fuzionálunk, melyek közül kettő az eredményül várt pixel címkére szavaz (itt változás vagy nem változás), míg a harmadik a jellemzők lokális megbízhatóságának becsléséhez járul hozzá.

Végző soron tehát az  $L^3$ MRF és a CXM modellek közötti választás alkalmazásspecifikus tulajdonságokon alapulhat.

#### *5. A 3.3.3. szakasz kvantitatív összehasonlításánál a szerző említi, hogy valamivel gyengébben teljesített a modell rosszabb minőségű képeknél. Felmerült, hogy ezzel kapcsolatban akár a modellben, akár más eszközök bevonásával (előfeldolgozás, stb.) javítás történjen?*

A hosszútávú változásdetekcióhoz használt légi képeket ortofotó formájában szereztük be a Földmérési és Távérzékelési Intézettől (FÖMI, ma a Lechner Tudásközpont része), így ezek az ipari standardok szerint professzionális geometriai és fotogrammetriai korrekción estek át, mielőtt eljutottak hozzánk, ezért mi a további javításukkal külön nem foglalkoztunk a kutatómunka során. A dolgozatban említett rosszabb minőség az 1984-es „Archive” felvételen a kezdetlegesebb technológiai háttérből adódott, amit az 5. ábrán szemléltetnek (kisebb élesség, csak szürkeárnyaltos rögzítés stb.), a 2005-ös ortofotóval való összehasonlításban.





5. ábra A Szada környékéről készült „ARCIEVE” 1984-es (fenn) és a frissebb 2005-ös (lenn) ortofotó a Földmérési és Távérzékelési Intézet (FÖMI, ma a Lechner Tudásközpont része) adatbázisából, melyeket a munka során használtunk

A *Bíráló* felvetése ugyanakkor teljes mértékben releváns és számunkra is időszerű, mivel egy tervezett projekt keretein belül szeretnénk többek között a CXM módszert a Lechner központ rendszerébe integrálni. Itt mindenképpen foglalkoznunk kell majd az előfeldolgozási és képjavítási kérdésekkel is akár az ipari partner eszköztárát alapul véve, akár azoknak alternatívát mutatva.

*6. Nagyon pozitív, hogy a szerző a CXM modell esetében összehasonlításokat végzett azóta megjelent eljárásokkal. Történt ugyanilyen összehasonlítás az L<sup>3</sup>MRF modell esetében, illetve ez a megoldás vajon hogyan viszonyul az azóta megjelent módszerek pontosságához?*

A CXM modell utóélete valóban lényegesen gazdagabb, mint L<sup>3</sup>MRF modellváltozaté.

A két módszerből egyértelműen a CXM-nek volt jelentősebb szakmai visszhangja, például a közlő IEEE Trans. folyóiratcikkek ([13] és [12], mindkettő 2009 októberi megjelenésű) Google Scholar hivatkozásainak aránya a válasz írásakor 105:35, a 2019-es hivatkozásokat tekintve 18:2. A CXM cikkhez publikált adatbázisunkhoz e mellett nagyságrendekkel több letöltési kérés is érkezett, mint a jelszó birtokában szintén elérhetővé tett, objektum-elmozdulásokat tartalmazó tesztalmanachunkhoz. Ennek az oka a CXM kevert Markovi modelljének nagyobb fokú eredetisége mellett az is lehetett, hogy a megcélzott hosszútávú változásdetekciós feladat sokkal inkább típusfeladatnak számított a távérzékeléses adatelemzés nemzetközi szakirodalmában, így maga a probléma iránt is nagyobb volt az érdeklődés. A 2015-ös összehasonlító folyóiratcikkünk [4] is ennek az érdeklődésnek a jegyében született meg.

Az L<sup>3</sup>MRF ezzel szemben a mozgó kameráról történő objektumdetekció (szintén népszerű) feladatát viszonylag speciális peremfeltételeket támasztva közelítette meg: a szokásos légi videók feldolgozása helyett két, pár másodperc különbséggel készített bemeneti képet vizsgáltunk, és az objektumfelismerés vagy -követés helyett kizárólag az elmozdulások maszkjának a kinyerésére fókuszáltunk. Alkalmazási javaslatként sem a valós idejű célpontkövetést jelöltük meg, hanem például a sztereo rekonstrukció során a mozgás miatt használhatatlan képrégiók szűrését. Az L<sup>3</sup>MRF egyelőre nem jutott el ezen a területen alkalmazási szakaszba, így eddig az összehasonlítása sem merült fel újabb módszerekkel.

*7. A 3. fejezetben tárgyalt mindkét módszert illetően felmerül, hogy a futási időkkal kapcsolatban mi mondható el róluk (összehasonlítva például a többi tárgyalt külső módszerrel)?*

Az L<sup>3</sup>MRF modellt kameramozgástól független objektum-elmozdulások régióinak kinyerésére 320×240 méretű képrészleteken teszteltük, körülbelül 10 másodpercet számolt egy képpáron, melynek fele a regisztráció és a jellemzőkinyerés, másik fele háromrétegű MRF optimalizáció volt. Számszerű futási-idő-összehasonlítás itt nem készült a referenciamódszerekkel, de a publikáció idején a sebesség és teljesítmény együttes figyelembevételével versenyképes volt a megoldás a referenciaeljárásokhoz képest. A publikációban jeleztük, hogy a módszer ebben a formában és implementációban nem alkalmas valós idejű objektumdetekcióra és -követésre, azonban használható forgalomsűrűség becslésére, vagy sztereo rekonstrukció előtt a dinamikus, „outlier”-ként megjelenő régiók kiszűrésére akár interaktív akár offline alkalmazásokban.

A Markov mezős hosszútávú változásdetekciós módszerek 2015-ös összehasonlító kiértékelése során [4] cikkben kitértünk a három összehasonlított módszer, az általunk javasolt CXM, a Multicue MRF [135] valamint a Fúziós MRF [117] futási idejeinek az áttekintésére is. A kiértékelés az általunk készített SZTAI AirChange adatbázis 952×640 pixel méretű képpárjain történt:

- A CXM módszer C++ implementációban 25-30 másodpercig számolt Intel(R) Core(TM) i7 3.20 GHz CPU processzoron. Ebből a jellemzőkinyerés körülbelül 5 másodpercet vett igénybe, a fennmaradó időt a négyrétegű MRF energiafüggvény optimalizációja vitte el.
- A Multicue MRF [135] szintén C++ implementációban valamivel gyorsabbnak bizonyult: szerzői mérése szerint 10-15 másodpercig számolt közepes teljesítményű laptop-on, hasonló összesített hibaértékkel, de gyengébb F-score-al mint a CXM [4].
- A Fúziós MRF [117] modell Matlab implementációban körülbelül 2 percig számolt asztali számítógépen, aminek mindegy 80%-át a *Cluster Reward Algorithm* alapú jellemzőszámítás emésztette fel.

A [4] cikkben hangsúlyoztuk, hogy mivel tipikusan offline feladat a hosszútávú változásdetekció, a mért futási időparaméterek nem jelentenek szűk keresztmetszetet az alkalmazásukhoz.

Hivatkozások:

[117] T. Szirányi and M. Shadaydeh, "Segmentation of remote sensing images using similarity measure-based fusion-MRF model," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, pp. 1544-1548, Sept 2014.

[135] P. Singh, Z. Kato, and J. Zerubia, "A Multilayer Markovian Model for Change Detection in Aerial Image Pairs with Large Time Differences," in *International Conference on Pattern Recognition*, (Stockholm, Sweden), May 2014.

A bíráló kérdései, megjegyzései a 4. fejezethez (2.1. altézis és 2.2. altézis) kötődően:

8. *A 4.2. szakaszban tárgyalt alkalmazás (épületek detektálása) további pontosíthatóságával nem merült fel más objektumok detektálása és modellbe emelése? Az épületekhez kapcsolódóan különösen az úthálózat lehet például ilyen, amit egyrészt térképészeti alkalmazásokkal is potenciálisan lehet hangolni, másrészt az utakhoz való geometriai viszonya a házaknak (élek merőlegessége/párhuzamosság) javíthatja a becslést.*

Eddig még nem volt ilyen jellegű kísérletünk, de köszönöm a felvetést, amit a jövőbeli munka szempontjából hasznosnak és aktuálisnak is érzek. Korábban születtek már jelölt pontfolyamat modell (JPM) alapú megoldások úthálózatok kinyerésére is légi képekről [Lacoste et. al. 2005], és épp a közelmúltban jelent meg egy új megoldás, ami mélytanulást és JPM módszerek fúzióját alkalmazza ugyanerre a feladatra [Li et.al 2019]. Az úthálózat és az épületpopulációk együttes modellezése ezért érdekes perspektívát vetít elő városmodellezés területén, ezzel együtt persze az egyes modellkomponenseket erősebben alkalmazásspecifikusabbá teheti.

[Lacoste et. al. 2005] C. Lacoste, X. Descombes and J. Zerubia, "Point processes for unsupervised line network extraction in remote sensing," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1568-1579, Oct. 2005, doi: 10.1109/TPAMI.2005.206.

[Li et.al 2019] Tianyu Li, Mary Comer, Josiane Zerubia. Feature extraction and tracking of CNN segmentations for improved road detection from satellite imagery. *ICIP 2019 – IEEE International Conference on Image Processing*, Sep 2019, Taipei, Taiwan

9. *A 4.3. szakaszban ISAR felvételeken való objektumkövetéshez használt FmMPP modell implementációjával kapcsolatban felmerül, hogy ennek elosztott implementációja (különösen GPU/FPGA platformon) mennyire támogatott? Azaz a futási idő várhatóan javítható-e ilyen*



*módon? Picit általánosabban fogalmazva, a disszertációban tárgyalt többi modellnél lehetnek-e lényeges különbségek párhuzamosíthatósági szempontokból?*

Magam nem dolgoztam elosztott implementációs feladatokon (eddig), azonban a Markov mezős, illetve jelölt pontfolyamat modelles megoldásoknak számottevő szakirodalma van erre vonatkozólag, hiszen a gráf alapú modellstruktúrák lehetővé teszik, hogy az energiaminimalizációs eljárás során az egymástól független csomópontokhoz tartozó energiakomponenseket egyidőben vizsgáljuk, és szükség esetén a csomópont paramétereit egymástól függetlenül módosítsuk.

Az  $F^m$ MPP modellt esetén az időszekvencia alapú reprezentáció miatt az egyes időkereteken detektált objektumok csak a megelőző és a következő képkockákon lévő találatokkal vannak közvetlen interakcióban, a sztochasztikus optimalizáció kiindulópontjaként használt determinisztikus elő-detekciós lépések (*preliminary target extracion, target structure refinement*) pedig minden képkockán külön-külön futnak le, így ennek a modellnek a párhuzamosítása - kis módosítások után- valóban hatékonyan elvégezhető lehet.

Az általánosabb kérdésre válaszolva lehetőség van szinte az összes, a dolgozatban említett modell valamilyen fokú párhuzamosítására:

A Markov véletlen mezős képszegmentálási módszerekhez általam is használt módosított Metropolis optimalizációs algoritmus egy párhuzamos implementációját már 20 éve publikálták (Szirányi et. al. 2000), az alapötlet az ott használt cellurális neurális hálózat helyett más párhuzamos architektúrákra is könnyen átültethető. A (Mottishaw e.t al. 2010) cikk szerzői egy *Belief Propagation* (BP) alapú Markov mezős energiaminimalizációs algoritmusnak készítették el a párhuzamosított változatát modern többmagú processzoros architektúrákra (például 8 maggal rendelkező Quad-Core Intel Xeon E5405 processzorokra, illetve egy négyprocesszoros Six-Core AMD Opteron 2435, összesen 24 maggal dolgozó architektúrára). Átlagosan 8-szoros gyorsítást értek el a soros implementációhoz képest, 600x480-as képeken 12.7 s-ról 1.6 s-re csökkentve az optimalizáció átlagos idejét.

A jelölt pontfolyamat modell (JPM) alapú objektumdetekciós módszerek legelterjedtebb optimalizációs technikája a Reversible Jump Makrov Chain Monte Carlo (RJCMCMC) alapú iteratív eljárás. Mivel az RJCMCMC egy lépésében legfeljebb egyetlen objektumjelöltet adhat hozzá az aktuális populációhoz, az eljárás gyakran rendkívül lassan konvergál. Az általam előnyben részesített többszörös születés és halál algoritmus (*Multiple Birth and Death*, MBD, [87], 2009) ezzel szemben egy lépésben számos új objektumot ad hozzá a konfigurációhoz, ezért *kizárólag egyszálú, soros implementációkat használva* a [87]-ben bemutatott tesztek alapján az MBD eljárással akár egy nagyságrenddel rövidebb számítási időn belül érhetők el hasonló minőségű eredmények, mint RJCMCMC-t alkalmazva. Ugyanakkor érdemes megemlíteni, hogy később megjelent az RJCMCMC alapú JPM optimalizációs algoritmusnak egy párhuzamosított változata (Verdie, et.al. 2014), amely a cikkben közölt kísérletek szerint fölülmúlta az ugyanilyen elvek alapján teljesen nem párhuzamosítható MBD sebességét. Azzal kapcsolatban, hogy pontosan milyen típusú energiafüggvényeknél hatékonyabb a párhuzamos RJCMCMC az MBD-nél, egyelőre nem születtek részletes összehasonlító elemzések, mindenesetre a disszertációban bemutatott JPM modellek energiafüggvényei optimalizálhatók RJCMCMC-vel is, tehát a párhuzamos implementációnak nincs elvi akadály.

A negyedik téziscsoportozhoz tartozó (főként a 4.3 altézisben szereplő) pontfelhőfeldolgozó eljárásainkhoz doktorandusz hallgatóimnak köszönhetően a több módszernek is elkészült párhuzamos implementációja többmagos CPU vagy GPU platformokra. A pontfelhőfolyam („stream”) beolvasása, az úttest detekciója és az objektumok elkülönítése és követése valós időben működik többszálú, CPU-n futó implementációval, ezek az eljárások később ipari rendszerbe is beépítésre kerültek egy projekt során (a párhuzamosítás saját méréseink szerint itt 3-4-szeres gyorsulást eredményezett). A mobil



lézerszenveléssel nyert 3D pontfelhő szemantikus osztályozásához a tanítást Nvidia GeForce GTX 1080 GPU-n végeztük.

(Szirányi et. al. 2000) T. Szirányi, J. Zerubia, L. Czúni, D. Geldreich, Z. Kato, Image Segmentation Using Markov Random Field Model in Fully Parallel Cellular Network Architectures, Real-Time Imaging, Volume 6, Issue 3, 2000,

(Mottishaw et al. 2010) S. Mottishaw, S. Zhuravlev, L. Tang, A. Fedorova, and G. Hamarneh An Evaluation of Parallelization Techniques for MRF Image Segmentation, International Workshop on High-Performance Medical Image Computing for Image-Assisted Clinical Intervention and Decision-Making (MICCAI HP) 2010.

(Verdie, et.al. 2014) Verdié, Y., Lafarge, F. Detecting parametric objects in large scenes by Monte Carlo sampling. Int J Comput Vis 106, 57–75 (2014)

[87] X. Descombes, R. Minlos, and E. Zhizhina, "Object extraction using a stochastic birth-and-death dynamics in continuum," Journal of Mathematical Imaging and Vision, vol. 33, pp. 347–359, 2009.

A bíráló kérdései, megjegyzései az 5. fejezethez (3.1. altézis és 3.2. altézis) kötődően:

*10. Van-e mélységi limit a hierarchiára nézve? Mennyire növeli meg a műveletigényt, ha növeljük a mélységi szintek számát?*

A hierarchiaszintek bővítésének nincs elvi korlátja, elsősorban a modellezés bonyolultságát, és az optimalizáció végrehajtási idejét növeli a mélyebb struktúra.

A futási idő szempontjából a szintek száma mellett az egyes szintek komplexitása is lényeges tényező. Az EMPP modell három bemutatott példaalkalmazásában közös volt, hogy egy tipikus populáción belül a (szülő) objektumok számánál lényegesen kevesebb volt az objektumcsoportok száma, illetve egy szülőhöz csak egy vagy néhány gyermekobjektum tartozott. Ezek az alkalmazási környezetből adódó megkötések lényegesen felgyorsították a D. függelékben részletezett optimalizációs algoritmus tényleges lefutását: például [2]-ben közölt kiértékelésünk alapján az épületdetekciós és a forgalomfelügyeleti alkalmazásokban csupán 20-30%-os többletszámítást mértünk, mint az ugyanahhoz a bemeneti adathoz tartozó egyszintű (az MBD eljárással csak a szülő objektumokat detektáló) JPM modellek esetén [2]. Más alkalmazási körülmények, például sok szülőt sok gyerekkel tartalmazó konfigurációk – esetleg mélyebb hierarchiaszinteket is bevonva - jelentősebb számítási többletet eredményezhetnek, így ezeknek a hatékony kivitelezése további kutatást igényelne.

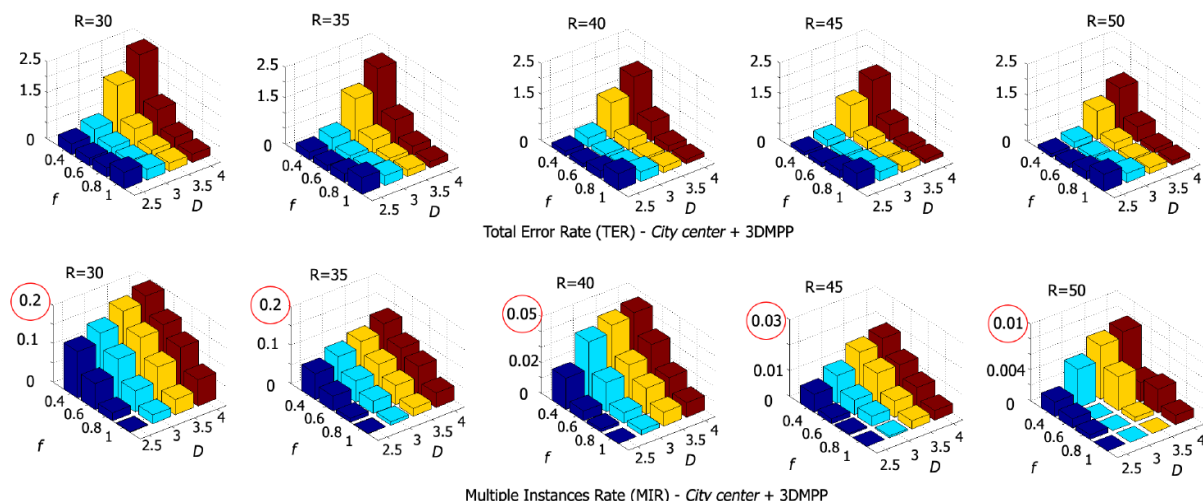
A bíráló kérdései, megjegyzései a 6. fejezethez (4.1. altézis, 4.2. altézis és 4.3. altézis) kötődően:

*11. A 6.2. szakaszban (többkamerás felvétel alapján történő feldolgozás) merül fel az egyszerű kérdés, hogy mi az oka az emberi test hengerrel való reprezentációjának? Miért nem befoglaló téglatestet tekintünk, melyik a természetesebb modell?*

A kezek és lábak mozgás során megfigyelhető szabad mozgása, esetleg tárgyak, hátizsák cipelése miatt azzal az egyszerűsítő feltevessel éltünk, hogy (szinte) minden irányban egyforma valószínűséggel elérhet egy ember valamely része a függőleges középvonalához képest. A hengermodell a jelölt pontfolyamatba ágyazva egyfajta távolságtartási kényszerként definiálható, azaz preferálja, hogy az egyes emberek középvonalai ne legyenek túlzottan közel egymáshoz, ami a mindennapi életben egy alapvetően áthaladásra használt közterületen (az érintkezési pillanatokat leszámítva) jó közelítés. A

hengerek  $R$  sugarát hiperparaméternek tekintettük, és a [9] publikációban részletes kiértékelést mutattunk be a hibamérték értékeiről különböző  $R$  választások mellett az adatfüggő energiatag  $f$  és  $D$  paramétereinek szimultán változtatásaival (6. ábra). Minimális hibát (TER = 0.123) az  $f = 0.8$ ,  $D = 3$  és  $R = 40$  cm-es hiperparaméter-beállításokkal értünk el a nyilvános Benchmark-ként használt City Center (PETS 2009) adathalmazon.

(PETS 2009). Dataset: Performance Evaluation of Tracking and Surveillance [Online]. Available: <http://www.cvg.reading.ac.uk/PETS2009/a.html>

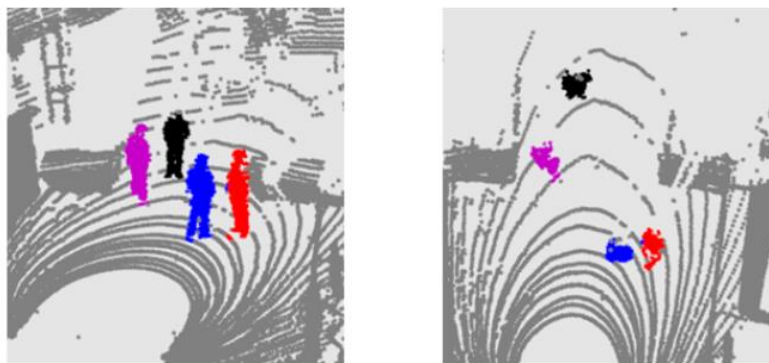


6. ábra a többkamerás felvételeken alapuló személylokalizációt és magasságbecslést megvalósító modellünk teljesítményének kiértékelése különböző paraméterállások mellett.  $R$  az emberek modellezéséhez használt henger sugara, amit az eljárásunk hiperparaméterként használt

## 12. A 6.3. szakaszban két kisebb értelmezési hiba tűnt fel:

- „The centre of each extracted blob is considered as a candidate for foot position ont he ground.” Nem inkább a „blob” alsó pontja lenne ez?

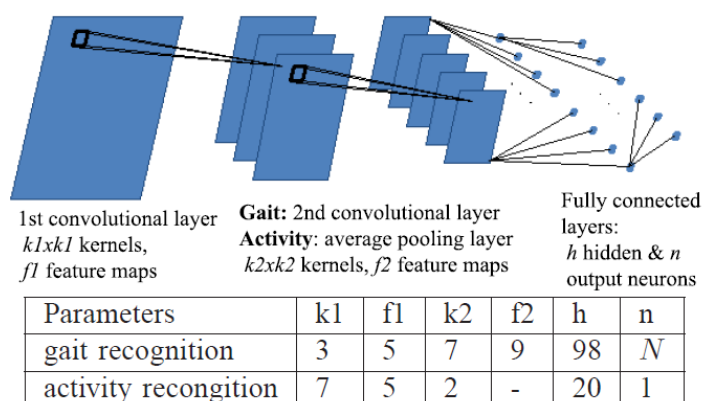
Az idézett műveletet megelőző lépésben a detektált 3D előtérpontokat a (lokálisan) becsült talajsíkra vetítjük, így már a síkon 2D-ben kell értelmezni a „foltokat”, aminek a (síkon vett) középpontját kiszámítjuk. A [6] publikációban a könnyebb érthetőség kedvéért egy illusztráló kép is szerepelt, amit itt a 7. ábra bemutatok, sajnos terjedelmi okokból ez az ábra a disszertációba nem fért bele.



7. ábra Gyalogosok elkülönítésének eredménye a Lidar felvételeken. Balra és középen: a 3D pontfelhő oldalnézeti valamint felülnézeti megjelenítése; Jobbra: a 2D talajsíkra vetített előtér régiók, az elkülönítés valójában ezen a 2D képen történt, és a régiók középpontjait feleltettem me a gyalogosok talppontjainak a talajsíkon

b. a 6.17. ábra táblázatának Task oszlopában szereplő rövidítéseket nehéz értelmezni.

Valóban, itt szerkesztési hibát vétettem a dolgozat készítése során. A módszert közlő [1] publikációmban (itt: 8. ábra) még konzisztensen szerepeltek a jelölések az ábrához tartozó táblázatban:  $(k1, f1)$ , illetve  $(k2, f2)$  az első illetve a második konvolúciós réteg kernelméretét, valamint a használt filterszámot írja elő (cselevésfelismerésnél a második réteg pooling volt, így ehhez nem tartozott filterszám). A  $h$  paraméter a teljesen összekötött réteg rejtett neuronjainak a száma,  $n$  pedig a kimeneti réteg neuronszáma, ami járásfelismerésnél az adatbázisban szereplő emberek  $N$  számával volt egyenlő.



8. ábra A Lidar alappú járás- és cselevésfelismerési feladatokhoz használt CNN hálózat sémája és paraméterezése

13. A 6.3.6. szakaszban a  $k=100$  és  $l=60$  paraméterbeállítás mi alapján történt?

A két paraméter a járásenergiakép számításának módját határozza meg:  $k$  az egy adott személyhez tartozó, véletlenszerűen választott referenciakeretek száma, amelyek az egyes minták gyűjtéséhez szolgáló szekvenciárészletek kiindulópontjai (tehát  $k$  egyben az egy személyről gyűjtött minták száma),  $l$  pedig az egymást követő Lidar pontfelhőkeretek száma amire az átlagolást elvégezzük egy adott minta generálásához. A paramétereket kísérleti úton, több érték kipróbálása útján állítottuk be, azonban a választásban szerepet játszottak intuitív megfontolások is. A kísérlet alapjául szolgáló hagyományos videókat feldolgozó megoldásban [173] járásciklusonként generáltak egy mintát a szerzők, ez normál sétatempót feltételezve körülbelül 1 másodpercnyi felvétel, ami körülbelül 30 képkockányi információt tartalmazott a videókból. A Lidaros kísérletünkben a képrögzítési sebesség 10 fps volt, ami harmada a hagyományos videokénak. Ezen kívül a pontok térbeli sűrűsége is lényegesen alacsonyabb volt a Lidar mérésen, mint a videokamera képein, így javasoltunk 6 másodperces adatgyűjtési ciklusokat, amiből az  $l = 60$  választás következett. Ez a személyenként 6 másodperces megfigyelési folyamat az általunk rögzített Lidar méréseken többnyire kivitelezhető is volt, ugyanis a sziluetteket mindig a járásgörbe pillanatnyi érintőjével párhuzamos síkra vetítettük, tehát a tesztalanyok ciklus közbeni irányváltása nem okozott gondot. Viszonylag nagyszámú minta kiválasztását ( $k = 100$ ) az tette lehetővé, hogy hosszú, 2:00-3:30 perc közötti felvételeket rögzítettünk, és nem írtuk elő, hogy az egyes mintákhoz tartozó Lidar keretek feltétlenül diszjunktak legyenek. Bár a fenti kísérleti feltételek erős megkötéseknek tűnnek valóságos biztonságtechnikai alkalmazásokban előforduló, általános körülményekhez képest, megjegyezzük, hogy így is jelentős lépést tettünk a realisztikus körülmények között működő járásfelismerés megoldása felé, hiszen a képi alapú szakirodalmi módszerek jelentős része csupán műtermi (például futószalagos) mozgások elemzésére alkalmazható.



[173] J. Han and B. Bhanu, "Individual recognition using gait energy image," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, pp. 316–322, Feb 2006.

*14. A 6.3.4. szakaszban említett CNN esetében történ hiperparaméter-optimalizáció (pl. a struktúrára vagy a szabad paraméterekre nézve)?*

A szóbanforgó kutatás fő célja a forgó többsugaras lézerszkennerek egy új lehetséges funkciójának a feltérképezése és proof-of-concept szintű bizonyítása volt. A gépi tanulás részt nem jártuk részletesen körbe, mivel a mintaosztályozási feladat csak egy korlátozott része – és nem kifejezetten az újdonságrésze - volt az egész folyamatnak, és a kísérletek során azt tapasztaltuk, hogy maga a mintaosztályozás a figyelembe vett szcenárióknál viszonylag könnyen tanulható volt. Ahogy a dolgozat 6.3 táblázata mutatja, a feladat jól megoldható volt egy standard Multi-Layer Perceptron (MLP) hálózattal is, amihez a CNN-es kiegészítés viszonylag kis javulást tett még hozzá. Utóbbi struktúráját próbálkozás útján véglegesítettük, például a rejtett neuronok számának változtatásával, így jutottunk el a dolgozat 6.17 ábráján (itt a 8. ábrán) bemutatott nagyon egyszerű CNN hálózathoz.

*15. A 6.3.4. szakaszban miért a megjelölt 5 tevékenység (bend, watch, phone, wave, wave2) került kiválasztásra? Gyakorlati szempontból logikusabbnak tűnne pl. biztonsági megfigyelőrendszereknél várhatóan fontosabb tevékenység (pl. ütés, kúszás).*

A kutatás előzménye, hogy 2012-2013 környékén a SZTAKI-ban rendelkezésre álló Velodyne HDL-64E valósidejű Lidar szkennerre építve egy olyan technikai demonstráció kidolgozását céloztuk meg, ami nyilvános eseményeken, például *Kutatók éjszakáján*, vagy nemzetközi kiállításokon valós időben bemutatható a szenzor kihelyezésével. Ezért feladatként tűztük ki a rögzített pozícióba helyezett szenzor előtt elhaladó járókelők észlelését, követését, bizonyos fokú biometrikus (újra)felismerhetőségét, és egyszerű cselekvések beazonosítását. Utóbbi feladatnál a megjelölt öt tevékenység életszerűnek, és emellett könnyen kivitelezhetőnek és tesztelhetőnek is bizonyult; ellentétben például egy verekedés szimulációjával, vagy a kúszással, ami egy kiállításon nem mutatott volna elegánsan. A Bíráló által felvetett javaslatok természetesen relevánsak lehetnek valódi piaci alkalmazásokban gondolkodva.

A forgó Lidar-os technológia elterjedésének korai éveiben még nagy újdonságereje volt olyan új alkalmazások demonstrációjának, amelyek túlmutattak a szenzor alapvető funkciójaként ismert járművek nagy biztonságú akadályészlelésén. Jelentősebb sikert értünk el a 2015-ös Salzburg-i *Capturing Reality Forum* címet viselő közös ipari-akadémiai konferencián és kiállításon, ahol a Velodyne európai részlegének vezetősége érdeklődését is felkeltettük a bemutatónkkal. Bár a Lidar alapú biometrikai megoldásunk ár-érték arányban a nagy hardverkötség miatt még korántsem volt versenyképes (nehéz elképzelni egy 60.000 USD árú szenzort videomegfigyelő kameraként), a gyártó már akkor tervezte a hasonló elvű kisebb és olcsóbb szenzorok létrehozását, amihez tervei szerint jól illeszkedett a feladat. Akkori ígéretükhöz híven 2017 nyarán megkerestek minket az új VLP-32C szenzoruk megjelenése után, amihez lehetőségünk nyílt egy interaktív demonstrációs programot készíteni, ami két hétig futott a 2017 szeptemberi [Frankfurti Autószalon](#) (Frankfurt Motorshow) kiállításon.

**Hivatkozás:**

<https://www.sztaki.hu/innovacio/hirek/gepi-erzekeles-kutolaboratorium-elo-lidaros-bemutatott-tart-frankfurti>

*16. A tevékenységcsoportok esetleg valamilyen klaszterezéssel is kialakíthatóak lehetnek, történt ilyen jellegű vizsgálat?*

A tevékenységcsoportok automatikus klaszterezéssel történő kialakítására nem volt még kísérletünk. Elvi akadályát nem látom, bár a kapcsolódó tesztszenáriókat itt is gondosan meg kell tervezni, és általánosságban a felhasználási módot sem feltétlenül könnyű definiálni. Bár a személykövetés hatékonyan működik a bemutatott sokszereplős, de nem túlzottan zsúfolt helyszíneken, tömegjelenetek esetén az egyes szereplők elkülönítése és követése már korlátokba ütközhet. Nagyobb tömegek esetén kivitelezhető megoldásnak érzem felülnézetből rögzített felvételek készítését, amin könnyebben megoldható az emberek követése. Az így nyert 2D trajektóriákat felügyelet nélküli (*unsupervised*) tanulással klaszterezhetjük, ami a jellegzetes útpályat-típusokat automatikusan kialakíthatja (például a Fourier leírók terében).

*17. Továbbá – inkább csak intuitív érdeklődéssel –, a Szerző elképzelhetőnek tartja egy alacsonyabb dimenziójú látens tér kialakítását sziluettek számára, követve napjaink népszerű beágyazási technikáit (pl. Variational Auto Encoder mintájára)?*

A személy- vagy cselekvésfelismeréshez felhasznált sziluett-energiaképek (GEI) alacsonydimenziós reprezentációja mindenképpen hasznos javaslat az osztályozásuk elősegítésére, tulajdonképpen az ötletet adó [173] cikkben is ez történik, csak itt még mélytanuló hálók helyett főkomponens analízis (Principal Component Analysis, PCA) és többszörös diszkrimináns analízis (Multiple Discriminant Analysis, MDA) módszerekkel. Azt tesztek nélkül nem tudom megelőlegezni, hogy a javasolt módszerhez képest így lényegesen jobb eredményeket lehetne-e elérni, nálunk mindenesetre a gépi tanulás alkalmazásainál jellemző szükséges adatmennyiségekhez képest viszonylag kis adatbázis (28 különböző tesztalany) állt rendelkezésre, ami vélhetően nem lenne elég egy hatékony látens-tér kialakításához. A módszerünk nagyobb adathalmazra történő kiterjesztése megkövetelheti hasonló technikák alkalmazását, de itt figyelembe kell azt is vennünk, hogy valós alkalmazásokban sem mindig feltételezhetjük, hogy az összes tesztalany elegendő mintát szolgáltat.

A GEI alapú megközelítés újszerűbb alternatívája a Long Short Term Memory (LSTM) típusú, mérésszekvenciák közvetlen feldolgozására alkalmas mély neurális hálózatok alkalmazása. [Feng, 2016] LSTM alapú autoencoder megoldást javasolt járásjellemzők kinyerésére, ami több nézetből párhuzamosan rögzített optikai videófelvételeken működik. A módszer ebben a formában nem adaptálható a Lidar-os környezetünkbe, de az LSTM használata – megfelelő tanító adatok rendelkezésre állása esetén nálunk is szóba jöhet.

[173] J. Han and B. Bhanu, "Individual recognition using gait energy image," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, pp. 316–322, Feb 2006.

[Feng, 2016] Yang Feng, Yuncheng Li and Jiebo Luo, "Learning effective Gait features using LSTM," 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, 2016, pp. 325-330

*18. A 6.4. szakaszban a  $K \times K \times K$ -s voxel környezethez a  $K=23$  beállítás hogyan került meghatározásra?*

A dolgozatban nem említettem, de itt egy előzetes hiperparaméter-optimalizációt végeztünk. A disszertáció benyújtása után megjelent [Nagy & Benedek, 2019] folyóiratcikkünkben bemutatunk kiértékelési eredményeket a  $\lambda$  voxelméret  $K$  3D kernel méretét leíró hiperparaméterek különböző beállításával, a végső  $\lambda = 0.1$  méter, és  $K = 23$  értékeket ezek alapján véglegesítettük. A vonatkozó eredménytáblázatokat a 9. ábra tartalmazza.

| Parameters       | Voxel size $\lambda$ [m], using a fixed $K = 23$ kernel size |          |         |        |        |        |       |
|------------------|--------------------------------------------------------------|----------|---------|--------|--------|--------|-------|
|                  | 0.02                                                         | 0.05     | 0.1     | 0.2    | 0.3    | 0.4    | 0.5   |
| Number of voxels | 812500000                                                    | 52000000 | 6500000 | 812500 | 240596 | 101563 | 52000 |
| Precision        | 34.7                                                         | 77.8     | 90.4    | 83.6   | 76.3   | 64.2   | 44.7  |
| Recall           | 29.8                                                         | 69.7     | 90.2    | 85.9   | 77.8   | 61.7   | 48.5  |
| F measure        | 32.1                                                         | 73.5     | 90.3    | 84.7   | 77.0   | 62.9   | 46.5  |

| Parameters | Data cube's side length ( $K$ ), using a fixed 0.1 m voxel size |      |      |      |      |      |      |      |      |      |      |
|------------|-----------------------------------------------------------------|------|------|------|------|------|------|------|------|------|------|
|            | 7                                                               | 11   | 17   | 21   | 23   | 25   | 27   | 29   | 31   | 37   | 41   |
| Precision  | 58.4                                                            | 72.5 | 81.1 | 87.6 | 90.4 | 88.5 | 86.4 | 83.2 | 78.9 | 72.8 | 69.4 |
| Recall     | 55.7                                                            | 69.7 | 82.6 | 87.1 | 90.2 | 89.1 | 87.2 | 85.6 | 82.2 | 71.2 | 69.6 |
| F measure  | 57.0                                                            | 71.1 | 81.8 | 87.4 | 90.3 | 88.8 | 86.8 | 84.4 | 80.5 | 72.0 | 69.5 |

9. ábra a Mobil lézerszkenneléssel nyert pontfelhő szemantikus szegmentációjára javasolt 3D CNN modell hiperparamétereinek optimalizációja során mért teljesítményértékek  $\lambda$  voxelméret (fenn), és  $K$  3D kernel méretét (lenn) beállításokra.

[Nagy & Benedek, 2019] B. Nagy and C. Benedek, "3D CNN-Based Semantic Labeling Approach for Mobile Laser Scanning Data," in *IEEE Sensors Journal*, vol. 19, no. 21, pp. 10034-10045, 1 Nov.1, 2019

19. A 6.4.3. szakaszban bemutatott CNN esetében említésre kerül, hogy különféle architektúrák kipróbálása után véglegesedett a háló szerkezete. Nem lett volna egyszerűbb itt is hiperparaméter-optimalizációval meghatározni az architektúrát?

Itt nem történt szisztematikus (például *grid search* alapú) hiperparaméter-optimalizáció, mivel próbálkozással (*trial-and-error*) módszerrel viszonylag gyorsan sikerült egy hatékony és egyszerű struktúrát megtalálnunk, ami négy-négy konvolúciós és pooling rétegből, és egy teljesen összekötött rétegből állt (a próbálkozások során a rétegek számát, típusát és a tanítási rátát – *learning rate* – variáltuk). Ez a szakterület a kapcsolódó [3] cikkünk benyújtása óta (2016 nyara) nagyon gyorsan fejlődött, ma már valóban gyakori egy-egy gépi tanulási feladat megoldása során a széleskörű hiperparaméter-optimalizáció lefuttatása, vagy különböző architektúrák numerikus összehasonlítása.

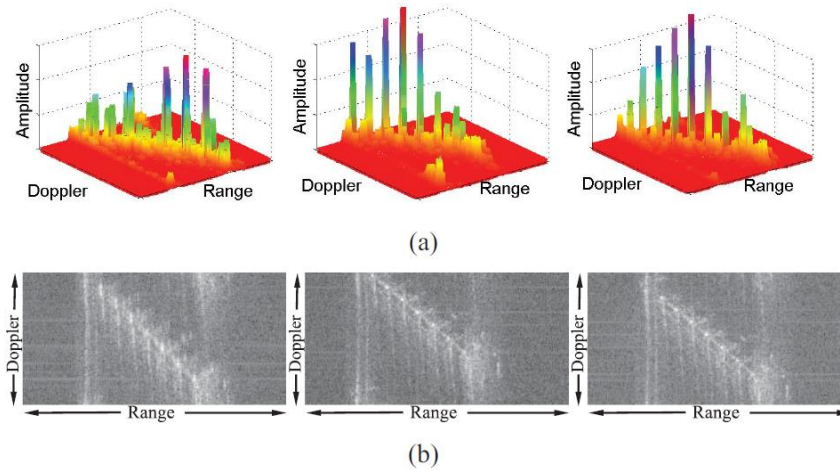
[3] A. Börcs, B. Nagy, and C. Benedek, "Instant object detection in Lidar point clouds," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 992 – 996, 2017

A bíráló kérdései, megjegyzései az egyes tézispontokhoz kötődően túl, a dolgozat egészét érintően:

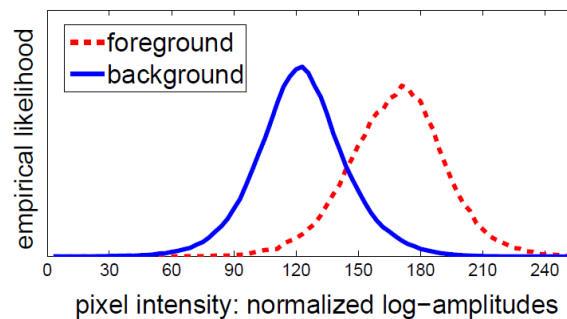
20. Általánosságban felmerül a kérdés, hogy a bemutatott modelleknél megfogalmazhatnánk-e elméleti eredményeket, ha pl. tér/időbeli statisztikai eloszlásokat ismerünk a feldolgozandó adattal kapcsolatban? Be lehetne-e építeni ezeket az ismereteket a modellekbe, vagy akár az optimalizációs eljárásokba?

Ennél a kérdésnél külön beszélhetünk az adatoknak az adatforrások természetéből adódó eloszlásáról, valamint a zaj statisztikai modellezéséről. A Markovi megközelítések (Markov véletlen mezők, jelölt pontfolyamat modellek) alapvetően Gaussi fehérzajjal számolnak, ez a feltételezés a dolgozatban bemutatott kísérleti eljárások nagy részénél helytálló közelítésnek is bizonyult. Például a 3.2(c) ábrán (itt 4. ábra, balra) megfigyelhetjük, hogy a két egymásután készített légi kép közötti  $f_d(\cdot)$  intenzitáskülönbség-jellemző empirikus hisztogramja valóban Gaussi eloszlást követ: itt feltételezhetjük, hogy egy kis átlagintenzitás-különbség fennáll a két kép között, valamint mindkét kép Gaussi zajjal terhelt.

A 4.3 fejezetben (2.2 altézis) ismertetett radarképek analízise során a szegmentáláshoz felhasználtuk az ISAR képfeldolgozás szakirodalomban elterjedt log-normal megközelítését [White & Williams, 1999], azaz a mért amplitúdóértékeket logaritmikus skálán vizsgáltuk (10. ábra). A C.2.1. függelékben bemutatott Markov mezős előtér-háttér szegmentációs algoritmusban így a két osztály által generált amplitúdóeloszlásokat a logaritmus tartományban normális eloszlással közelítettük, ami saját mérésein szerint is hatékonynak bizonyult (11. ábra). A 10. ábra szintén megfigyelhető, hogy a karakterisztikus jellemzőpontokat okozó úgynevezett domináns visszaverő felületek („*dominant scatterers*”) a logaritmikus tartományban is jól kivehetők [Giusti et. al., 2011].



10. ábra ISAR képek szemléltetése: (a) egymást követő képkockák amplitúdóinak a megjelenítése a range-doppler tartományban (b) ugyanezen képek normalizált logaritmikus megjelenítése képként (intenzitás a log-amplitúdókat kódozva), forrás: [7]



11. ábra A logaritmus tartományban ábrázolt amplitúdóértékek hisztogramja egy 18 képkockából álló képsorozaton. Az előtér (foreground) és háttér (background) régiókat manuálisan különítettük el demonstrációs célból.

Amennyiben Gaussitól eltérő zajmodellel számolnánk, a jellemzőstatisztikákat leíró eloszlás is más alakú lenne, amit a megfelelő parametrikus függvény választásában mi is kifejezhetünk (például béta-eloszlás használata a  $f_c(\cdot)$  korrelációs jellemzőnél), vagy alternatívaként használhatnánk Gaussi keverékeloszlásokat, mint például a CXM modell  $f_g(\cdot)$  jellemzőjénél (a dolgozat 3.8(a) ábrája).

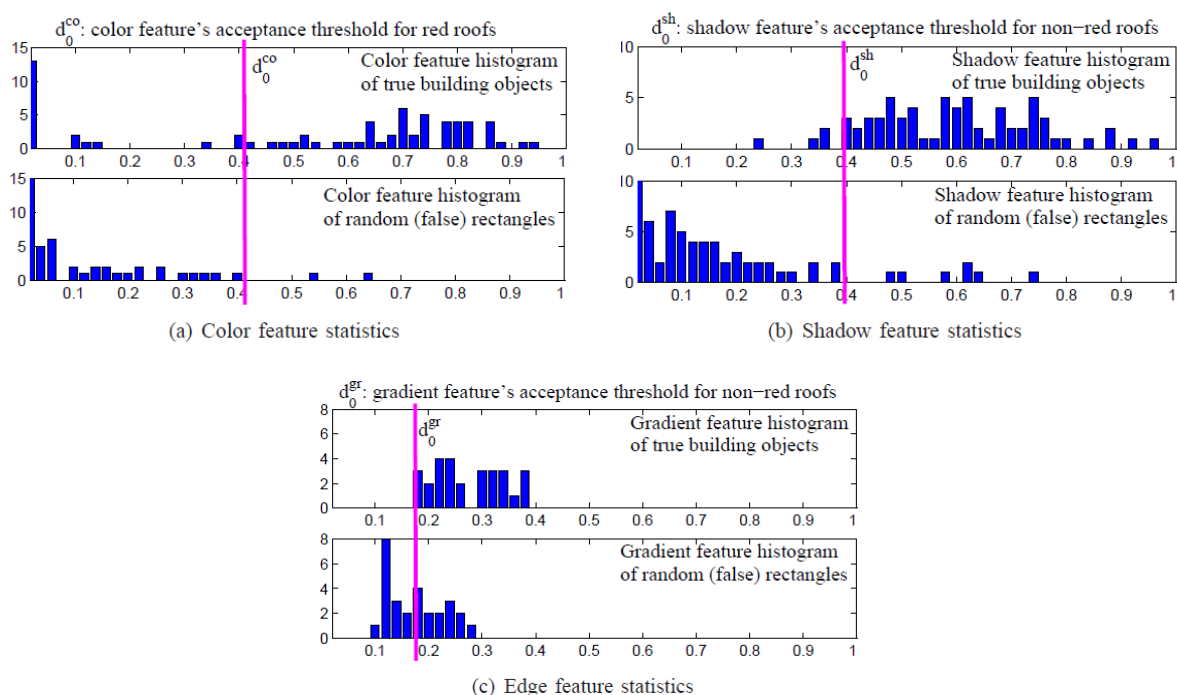
A fentiekén túl felmerülhet a nemparametrikus kernel alapú sűrűségfüggvényekkel történő modellezés (*non-parametric kernel density estimation*), amit a dolgozatban a 6.3.1.3 szakaszban a Lidar alapú mélységképen végzett háttérlevonás DMRF modelljében az előtér osztály leírására használtam. Végül megemlítem, hogy implicit módon a neurális hálózatos megoldások a tanítás során az egyes osztályok



karakterisztikus jellemzői mellett szintén megtanulják az adatokban rejlő különböző tér-időbeli statisztikai összefüggéseket és zajmodelleket.

Markov mezős modellezés során az alapvető Gaussi fehérzajmodell mellett léteznek megoldások például só és bors zaj figyelembevételére [Baghaie, 2016]. A diszkriminatív véletlen mezők fő előnye, hogy nem írja elő a pixelintenzitások feltételes statisztikai függetlenségét, így a valóságban is gyakran előforduló térben korrelált képi, vagy zajjelenségeket hatékonyan le lehet írni velük [Kumar&Herbert, 2006].

A jelölt pontfolyamat alapú modelljeink esetén a használt jellemzőeloszlások  $d_0$  paraméterét kéit empirikus úton, a vonatkozó jellemzőhisztogramok vizsgálata útján állítottuk be (az 12. ábra például az épületdetekció jellemzőstatisztikáit mutatja), itt szintén használhatók elméleti modellek is, amennyiben rendelkezésre állnak.



12. ábra Épületdetekciós alkalmazás: szín-, árnyék- és gradiensjellemzők hisztogramjai, és a megfelelő  $d_0$  eloszlásparaméterek forrás: [2]

Hivatkozások:

[2] C. Benedek, "An embedded marked point process framework for three-level object population analysis," IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4430–4445, 2017

[7] C. Benedek and M. Martorella, "Moving target analysis in ISAR image sequences with a multiframe Marked Point Process model," IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 4, pp. 2234–2246, 2014

[149] R. White and M. Williams, "Processing ISAR and spotlight SAR data to very high resolution," in Proc. IEEE Int. Geosci. Remote Sens. Symp., Jul. 1999, pp. 32–34.

[Giusti et. al., 2011] E. Giusti, M. Martorella, and A. Capria, "Polarimetrically-persistent scatterer-based automatic target recognition," IEEE Trans. Geosci. Remote Sens., vol. 49, no. 11, pp. 4588–4599, Nov. 2011

[Baghaie, 2016] Ahmadreza Baghaie Markov Random Field Model-Based Salt and Pepper Noise Removal arXiv 1609.06341 2016

[Kumar&Herbert, 2006] Kumar, S., Hebert, M. Discriminative Random Fields. Int J Comput Vision 68, 179–201 (2006).

## *21. Népszerű összehasonlítási platformként felmerül, hogy a bemutatott – és alapos – összehasonlításon túl sikerült-e a módszereket verseny (pl. Kaggle challenge) jellegű fórumokon tesztelni?*

A 4.1 altézisben bemutatott többkamerás emberdetekciós modellt egy ismert benchmark adatbázison (PETS2009) definiált lokalizációs problémára fejlesztettük ki, ahol a verseny fő szempontja a 3D detekció pontossága volt. Bár nem közvetlenül a PETS versenyen indultunk el, 2011-ben az aktuális *state-of-the-art*-nál jobb számszerű eredményeket elérve bejutottunk a CVPR (fő)konferenciára.

A Lidar adatok feldolgozásával kapcsolatos kutatásainkban az eddigiekben elsősorban az volt a célunk, hogy újszerű feladatokat oldjunk meg (például Lidar alapú járásfelismerés), kevésbé elterjedt adattípusokat dolgozzunk fel (Budapest Közút mobil lézerszkenneléssel nyert MLS pontfelhői), vagy kombináljunk (valósídejű forgó Lidar és MLS adatok), így ezekhez a feladatokhoz nem léteztek még versenyek. A 6.4.3. fejezetben bemutatott objektumosztályozásnál is fontos szempont volt, hogy egy teljes észlelési folyamatot átfogó (*end-to-end*), automatikus detekciót és osztályozást együttesen megvalósító pontfelhőfeldolgozó megoldást hozzunk létre, míg az akkoriban meghirdetett nyílt versenyek többnyire kézzel szegmentált objektumokat tartalmazó adatbázisokon írtak elő osztályozási feladatokat.

## *22. Metodológiaiag kiterjeszthetőek-e a módszerek időben alakjukat változtató objektumokra (állatok, szervi működés, stb.), továbbá egymást takaró objektumokra (pl. repülő, drón, robotkar átmenetileg takarja az alatta lévő objektumot)?*

A jelölt pontfolyamatmodellek (JPM) erős geometriai modellezőképességét az inverz leírás adja, ahol a megoldáshoz vezető út közvetlen megadása helyett a különböző javasolt megoldások kiértékelésre írunk fel fitness- vagy energiafüggvényt, a legjobb megoldás megtalálását pedig a populációterben végzett keresési műveletre bízunk. Az előbbieket miatt fontos, hogy a geometriai alakzat kevés paraméterrel jól leírható legyen, hiszen a paraméterek számának a növelése exponenciálisan növeli a paraméterter méretét, ahol a keresést végezzük. Bár ez a feltétel jelentős megkötést ad például az aktív kontúr modellekkel (*active contour model*), vagy az aktív alakmodellekkel (*active shape models*) szemben, az utóbb említett két eljárás során viszont globális optimalizáció helyett valamilyen gradiens alapú szuboptimális keresést végzünk, ahol kritikus a jó kezdeti állapot meghatározása, ami szintén nem könnyű minden esetben.

A szakirodalomban találunk néhány példát JPM és aktív kontúr modellek ötvözésére. [40] konferenciáinkban például egyszerűen egymás után tettük a két modellt: JPM segítségével detektáltuk a házak befoglaló téglalapjait, majd innen indított aktív kontúrokkal pontosítottuk a tetők alakját. Valamivel összetettebb megoldás szerepel [Kulikova 2011]-ben, ahol a szerzők az energiafüggvénybe beépítették az aktív kontúrt definiáló adattagokat, azonban a mintavételezés véletlenszerűen elhelyezett körökből kiinduló determinisztikus (tehát csak lokálisan optimális) aktív kontúrszámításon alapul. Releváns megközelítést javasolt [Ghanta et.al. 2018], ami a prior alakmodelleket előzetes rögzítés helyett tanulás útján határozza meg referenciaadatok segítségével.

A drónos-robotkaros alkalmazásokban többnyire valós idejű megoldásra van szükség, ezért a JPM alapú módszerek csak korlátozottan használhatók. Általában a takarási jelenségek modellezők lehetnek JPM segítségével megfelelő adattagok meghatározásával, például úgy, hogy az elvárt geometriai objektumhoz akkor is alacsony energiaértéket rendelünk, ha csak egy részében, de ott nagyon pontosan illeszkedik az észlelt képi struktúrára, vagy az objektummodellt részekből építjük fel, és az egyes részeket külön-külön próbáljuk a képre illeszteni [Gamal 2011].

## Referenciák

[40] A. Kovács, C. Benedek, and T. Szirányi, "A joint approach of building localization and outline extraction," in IASTED International Conference on Signal Processing, Pattern Recognition and Applications (SPPRA), (Innsbruck, Austria), February 2011

[Kulikova 2011] M. S. Kulikova, I.H. Jermyn, X. Descombes, E. Zhizhina and J. Zerubia A Marked Point Process Model Including Strong Prior Shape Information Applied to Multiple Object Extraction From Images. International Journal of Computer Vision and Image Processing, 1(2): pages 1-12, 2011.

[Ghanta et.al. 2018] Ghanta, Dy, Niu and M. Jordan. Latent Marked Poisson Process with Applications to Object Segmentation. Bayesian Analysis. 13. 10.1214/16-BA1035. 2018.

[Gamal 2011] Ahmed Gamal Eldin Point process and graph cut applied to 2D and 3D object extraction PhD thesis, INRIA Sophia Antipolis, 2011

*23. Minimális számú gépelési hibát találtam, azt azonban jeleznem kell, hogy a magyar nyelvű tézisfüzetben a 3.1. altézis leírása duplán szerepel, minimális módosítás mellett; a hiba az altézis értelmezését nem zavarja.*

Sajnos ez a szerkesztési hiba nem tűnt fel a többszörös ellenőrzések során (véltetően egy a bekezdésduplikálást előidéző billentyűzetkombináció véletlenül előhívása okozta a szövegszerkesztőben). Érdemi különbség szerencsére nincs a két változat között, stilisztikai szempontokat figyelembe véve inkább a második állítás tekinthető véglegesnek.

*24. A 3. téziscsoportnál a két altézis összevonás megfontolható lehetett volna, viszont véleményem szerint a nyomtatott áramkörökkel kapcsolatos járulékos munka és eredmények miatt a 3.1 altézis is megáll önálló eredményként*

A téziscsoportban bemutatott eredmények hátterét adó kutatómunka három különböző fázisban folyt:

1. Áramköri elemek vizsgálata szülő-gyermek kapcsolatban álló objektumokat tartalmazó Jelölt pontfolyamat modellekkel. Kapcsolódó folyóiratpublikációk: egyszerűs Pattern Recognition Letters cikk [11] és elektronikai technológiai szakemberekkel közösen publikált IEEE Trans. Industrial Electronics cikk [8].
2. Légi Lidar mérésekről járművek és járműcsoportok együttes kinyerésére alkalmas kétrétegű jelölt pontfolyamatmodell kidolgozása. Kapcsolódó folyóiratpublikáció: Börös Attila doktoranduszommal közösen írt IEEE Trans. Geoscience and Remote Sensing cikk [5].
3. A két előző modell összekapcsolása, és az egyesített modell kiterjesztése három különböző alkalmazásra. Kapcsolódó folyóiratpublikáció: egyszerűs IEEE Trans. Image Processing cikk [2].

Az első két fázis párhuzamosan folyt, a fent is említett különböző társzerzőkkel. Mivel az áramköri elemekről szóló modellváltozat publikációi olyan általam kidolgozott lépéseket és modellrészletet is tartalmaztak, amelyek az általános [2] cikkben nem szerepeltek, külön tézispontot hoztam létre a megoldásnak. A légi Lidar-os modell alkalmazás-specifikus elemeinek a kidolgozása Börcs Attila PhD hallgatómmal közösen történt [5], az általam hozzátett modellrészletek jelentős része pedig az általános modell prezentációja során is megjelent [2], ezért a második fázishoz nem tartottam szükségesnek különálló saját tézispont létrehozását. Végül a harmadik fázisban végzett modellkiterjesztés, és részletes kísérleti kiértékelés újabb jelentős, tudományos eszközökre épülő munkát igényelt.

Budapest, 2020. június 9-én

Benedek Csaba