

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

Válasz Prof. Zarándy Ákosnak

az MTA doktori tézis opponensének

Benedek Csaba

2020. június 9.

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

- 1. Az értekezés a bevezetést követő fejezetében a Jelölt dióhéjban bemutatja az alkalmazott módszerek matematikai alapjait. Itt egy rövid leírása jelenik meg a klasszikus Markov modelleknek, a Markov alapú szegmentációnak, illetve a Jelölt pontfolyamat modelleknek. Sajnos ez a fejezet nem elég részletes ahhoz, hogy a témát mélyen nem ismerő olvasó ebből megérthesse az alkalmazott matematikai módszereket. Például a Jelölt többször említi, hogy az adathalmazának egy részét tanításra használja, ugyanakkor ennek mikéntje nem derül ki érdemben sem itt sem később a doktori műből.*

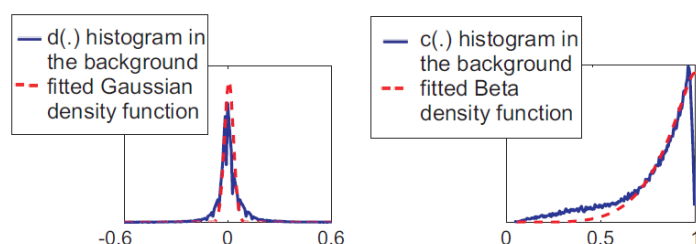
A bevezető két fejezetben valóban csak röviden esik szó Markovi modellek paraméterbeállításáról, néhány releváns szakmai referencia megemlítésével [90], [91]. Ugyanakkor a dolgozat további fejezeteiben a konkrét modellek ismertetése után visszatérek a paraméterbecslés kérdéseire: mind a 3. fejezet (többrétegű Markov mezők) mind a 4. fejezet (tér-időbeli jelölt pontfolyamat modellek) végén egy-egy alfejezetben (3.4, illetve 4.4) összefoglalom az aktuális modellek paraméterbeállításának a folyamatát:

3.4 fejezet. A többrétegű Markov mezős modellek adattagjai parametrikus sűrűségfüggvényekből állnak össze. Például az L^3MRF modellben az $f_d(s)$ jellemző a háttér régiókban Gaussi, míg $f_c(s)$ Beta eloszlást követ (Figure 3.2 (c) és (e)). Ezeknek a sűrűségfüggvényeknek a paraméterbecslése standard (például Matlab-ban is implementált) maximum likelihood becslő eljárásokkal elvégezhető néhány kézzel felcímkézett tanítóképről összegyűjtött mintahalmazon (1 ábra). A prior simasági tényező beállítására a szakirodalomból ismert stratégiákat használtam [73], mivel ezek a hiperparaméterek általában széles tartományban, a konkrét adatoktól és jellemzőválasztástól függetlenül megbízható működést biztosítanak (bár megemlítettem, hogy a szakirodalomban erre is léteznek automatikus becslőeljárások [136]). Az L^3MRF modell esetén a további hiperparaméterek (a korrelációs ablak v és a mozgó ablak l oldalmérete) beállítása kísérleti úton történt. Erre a disszertációban csak rövid utalás található, viszont a kapcsolódó publikáció [12] részletezi a folyamatot: különböző paraméterbeállításokkal kiszámoltuk és ábráztuk a kapott F-score értékeket (2. ábra, balra, forrás: [12]), majd az optimálisnak bizonyult $v = 7$, $l = 7$ konfigurációt választottuk. A CXM modellnél az eljárás intenzitásmodelljének hibáját rajzoltuk ki a Gaussi keverékeloszlás különböző komponensszámai függvényében (2. ábra, jobbra, forrás: [13]).

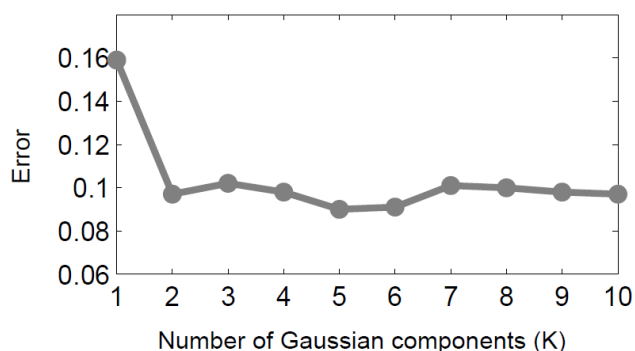
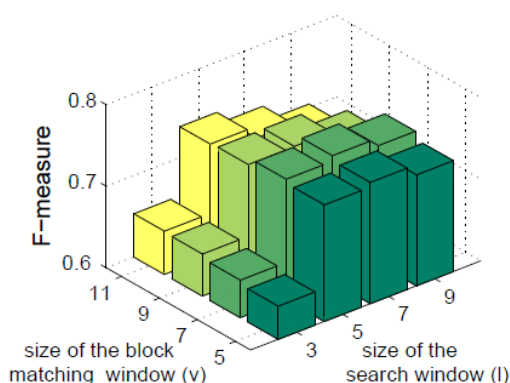
4.4 fejezet: A jelölt pontfolyamat modellek (JPM) esetén szintén meg kell határoznunk az objektumok adatfüggő és prior energiatagjainak a szabad paramétereit, valamint a többszörös születés-halálozás optimalizációs eljárás paramétereit. A prior és az optimalizációs paraméterek meghatározása itt is

szakirodalmi bevett gyakorlatok (*best practice*) alkalmazásával történt főként [87] alapján, míg az adattagok meghatározásának elveit 4.4 fejezet utolsó bekezdése tömören leírja.

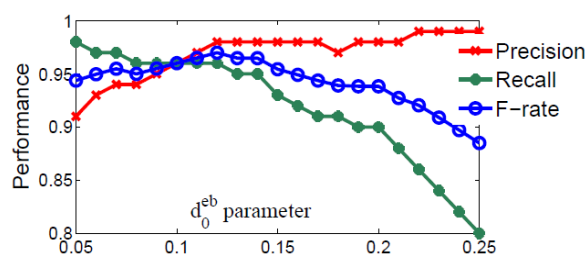
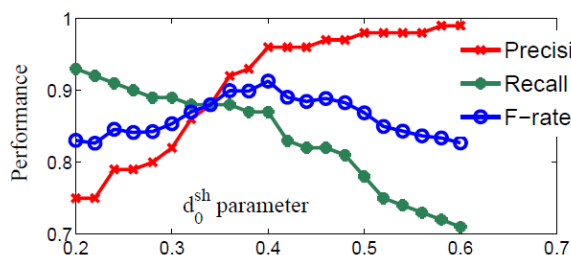
Az 5. fejezet szintén jelölt pontfolyamat modellekkel foglalkozik, itt a paraméterbeállítás módszerei megegyeznek a 4. fejezetben leírtakkal, ezért nem részleteztem külön a disszertációban. Ugyanakkor a beágyazott JMPP modell tesztelése során a [2] publikációban több alkalmazás több különböző paraméteréhez is ábrázoltam a precizitás (precision, Pr), fedés (Recall, Rc) és F-score görbéket a releváns dinamikartományyaikon, amit a 3. ábra az épületdetekciós és járműdetekciós példák egy-egy paraméterére mutat be.



1. ábra Az L^3 MRF model adatfüggő energiatajgaiban használt eloszlások meghatározása a tanító mintákból származtatott hisztogramok alapján Expectation Maximization eljárásokat alkalmazva



2. ábra: Balra: az L^3 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, árnyékletírójának paramétere, jobbra: Lidar alapú járműdetekció külső evidencia jellemző paramétere

Paraméterbeállítással kapcsolatos referenciák a dolgozatban:

[90] S. Ben Hadj, F. Chatelain, X. Descombes, and J. Zerubia, "Parameter estimation for a marked point process within a framework of multidimensional shape extraction from remote sensing images," in ISPRS Technical Commission III Symposium on Photogrammetry Computer Vision and Image Analysis (PCV), (Paris, France), 2010

[91] F. Chatelain, X. Descombes, and J. Zerubia, "Parameter estimation for marked point processes. application to object extraction from remote sensing images," in Energy Minimization Methods in Comp. Vision and Pattern Recogn., vol. 5681 of Lecture Notes in Computer Science, pp. 221–234, Bonn, Germany: Springer, 2009.

[73] Z. Kato and J. Zerubia, Markov random fields in image segmentation. Foundations and Trends in Signal Processing, Now Publishers, 2012.

[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

[136] V. Hodge and J. Austin, "A survey of outlier detection methodologies," Artif. Intell. Rev., vol. 22, no. 2, pp. 85–126, 2004.

2. *Formai szempontból az egyetlen aggályos pont az anyag mérete. A doktori szabályzat 100 oldalon maximálja az értekezés tartalmi részét, ami ebben az anyagban 140 oldal. Az egész dolgozat közel 200 oldalas, a függelékekkel, illetve a tartalomjegyzékkel együtt. Ez ugyanakkor nem egy túlságosan bőbeszédűre ereszített stílus eredménye, hanem annak, hogy a Jelöltnek nagyon sok eredménye született az elmúlt időszakban, és ezeket foglalta össze az anyagba.*

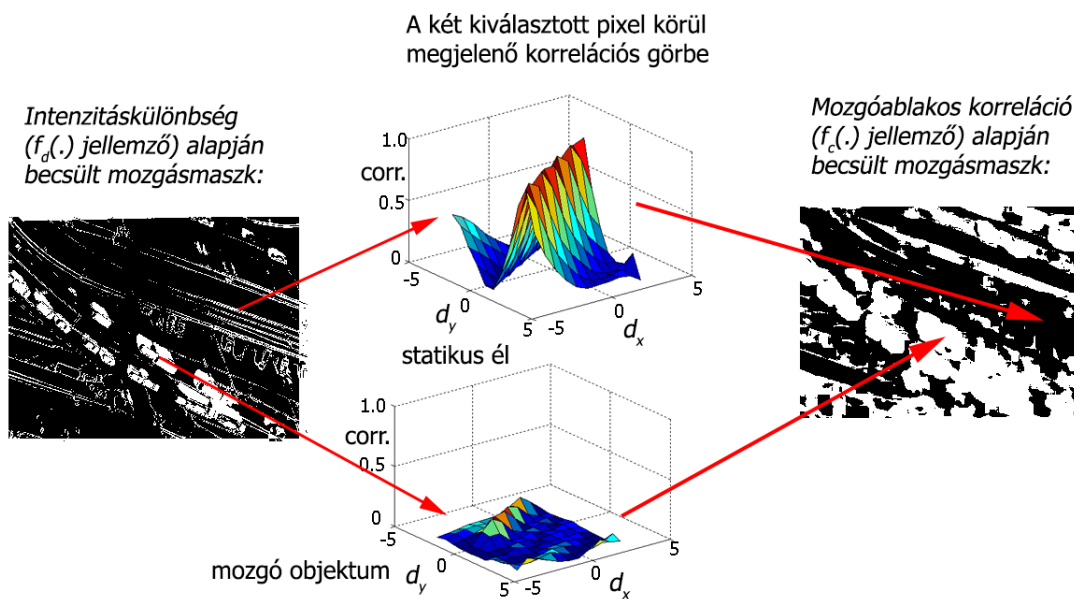
A dolgozat készítése során komoly mérlegelés után alakítottam ki a végleges terjedelmet. Egyrészt nem szerettem volna egyik általam fontosnak tartott eredményt sem teljesen elhagyni, ugyanakkor túlzottan össze sem tudtam tömöríteni a tárgyalásukat, mivel az MTA MTO doktori követelményrendszerének 1. pontja kimondja, hogy „Az értekezés [a korábban publikált] eredményeket egységes jelölésrendszerben és szemléletben foglalja össze, beleértve az eredmények előzményeit pontos irodalmi hivatkozásokkal, az eredmények helyét a szakterületen, *valamint áttekinthető bizonyításukat, igazolásukat részletes levezetésekkel*”. A fentieken túl a szabályzat valóban előírja, hogy „Az értekezés *érdemi fejezeteinek terjedelme ne haladja meg a 100 oldalt*”. Esetemben a bizonyítások kísérleti úton történtek, az ezeket bizonyító összehasonlító táblázatok, szöveges elemzések és demonstrációk ábrák jelentősen növelték a terjedelmet. Ugyanakkor arra ügyeltem, hogy amennyiben az *érdemi fejezetekből* kihagyjuk a bevezetést, az ismert matematikai háttér bemutatását és a konklúziót, és csupán a négy téziscsoport háttérét leíró 3-4-5-6 fejezeteket számítjuk bele, ezek együttes hossza *nem haladja meg a 100 oldalt* (21-120 sorszámú oldalak a disszertációban).

3. *[3. fejezet, L3MRF] A korreláció alapúval kapcsolatban az a kérdésem, hogy itt az éleket miért nem találjuk meg? Ha ugyanis két egymástól egyenes éllel elválasztott homogén terület van a képen, ott az él irányába mutató offset eltolással (os) kivágott képszegmens korrelációjának magas korrelációs csúcsot kellene adni. Ugyanakkor a képen csak a mozgó objektumok, illetve a nagy homogén területek emelkednek ki korrelációs csúcsként.*

A dolgozatbeli 3.2(f) ábra (itt az 4. ábra jobb oldali képe) egy korreláció alapú „mozgásmaszkot” mutat, ahol fehérrel az alacsony korrelációjú (mozgó objektum), feketével a magas korrelációjú (háttérhez

tartozó) lokális képrészleteket jelöltem. (Lehet, hogy ez a jelölés félreérthető volt a disszertációban). Amint azt a 4. ábra is mutatja, képeken a parallaxis miatt „elcsúszott” párhuzamos élek esetén létezik olyan offset érték, ahol magas az összehasonlított régiók korrelációja, ezért ezek nem kerülnek rá a mozgásmaszkra, hanem statikus régiókként kerülnek megjelölésre. A modell felállításkor tett kezdeti feltevések szerint az objektumok elmozdulása nagyobb, mint az offset keresési ablakának a sugara, így a mozgó objektumok régiói helyesen alacsony korrelációval fognak rendelkezni. A jellemző hibája, hogy a homogén statikus régióknál a normalizált keresztkorreláció érzékenyvé válik a zajra, így ezek hamis pozitív találatokat adnak a jellemző által szolgáltatott mozgásmaszkhoz.

A 4. ábra bemutatja a parallaxis miatt elcsúszott statikus él (fenn) és a mozgó alakzat (lenn) okozta intenzitáskülönbségek korreláció alapú megkülönböztetésének a folyamatát.



4. ábra Bal oldalon: a két regisztrált kép különbségképe ($f_d(s)$ jellemző). Középen: a mozgó korrelációs ablak különböző offset értékeihez tartozó korrelációs értékek görbéje, két olyan pixelhez, amit a bal oldalon látható intenzitás-különbségkép változásnak jelöl. A felső pixel a parallaxis miatt elcsúszott statikus él, míg az alsó egy mozgó alakzat része. Jobbra: a mozgó korrelációs jellemző térképe (a dolgozatbeli 3.2(f) ábra), a nyíl vége mutatja az előbb két pixelt, melyek közül a statikus él háttérpont, míg a mozgó alakzat előtérpont lesz.

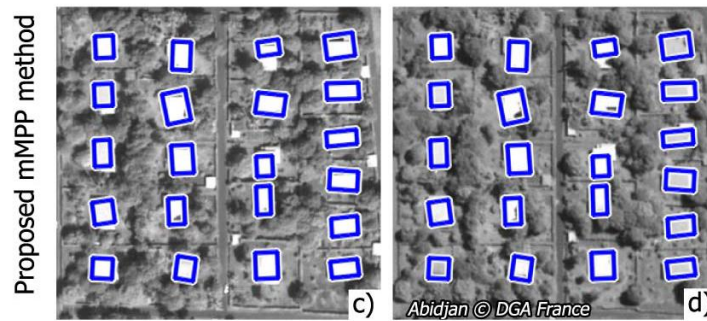
[3. fejezet, CXM] A kérdésem ennél a példánál az, hogy hogyan értelmezhető-e az algoritmus egy téli és egy nyári felvételen páron, ahol a vegetáció drasztikusan különbözik?

Bár téli-nyári képpárjaink nem voltak a kutatás során, elvi akadályát nem látom az ilyen helyzetek kezelésének, hiszen az $f_g(s)$ leíróból származtatott $P(f_g(s)|bg)$ Gaussi keverékeloszlással jellemzett energiatényező az előforduló intenzitáspárok együttes eloszlását méri a két képen, tehát akár nagyon különböző értékpárokhoz is tartozhat magas gyakoriság, így magas fitnessz érték. Persze amennyiben a téli képen az egész táj fehérnek látszik a hó miatt, a leíró alacsony információtartalommal fog rendelkezni.

A kérdés további részletei a kutatócsoportunk számára is érdekesek, ezért a közelmúltban benyújtottunk a Lechner Tudásközpont Űrtávérzékelési Osztályával közösen egy kapcsolódó pályázatot, ami ha támogatást nyer, a CXM módszerünk széleskörű tesztelését is el fogjuk végezni különféle évszakokban készített űrfelvételeket összevetve.

4. [4. fejezet, tJPM] Itt az lenne a kérdésem, hogy mennyire limitálja az algoritmus alkalmazását, ha a képet felhős időszakban készítették, azaz az árnyékok nem látszanak?

Az épületek jelenlétét több különböző jellemző felhasználásával is igazolhatjuk, az árnyék ezek közül csupán egy, ami bizonyos esetekben sokat segíthet, máskor akár el is hagyható. A [10] publikációban mutattunk árnyékmentes képeket is, melyen az algoritmusunk jól működött, ilyen például az Abidjan-i képpár (5. ábra)



5. ábra Épületdetekció az Abidjan képpáron

Egy adott képhalmazhoz a 47. oldalon ismertetett rugalmasan konfigurálható jellemzőfúziós algoritmussal lehet különböző épületprototípusokat definiálni, itt a nem releváns jellemzők kihagyhatók. Például a BUDAPEST képpáron az egyik prototípus kizárólag a piros tetőszín alapján detektálja az épületeket, míg a másik az élkép és az árnyékok együttes jelenlétét ellenőrzi, Abidjan esetén pedig a gradiens alapú önmagában alkalmas a tetők elkülönítésére. A fentiekén túl [10]-ben használtunk egy negyedik, úgynevezett tetőhomogenitás jellemzőt is.

5. [4. fejezet] A második példa Inverz Szintetikus Apertura Radar (ISAR) képek kiértékelése. Ez a példa számomra nehezebben volt áttekinthető, mert nagyon hiányzott egy rövid fejezet az ISAR képek háttéréről és tulajdonságairól, a feladat nehézségeiről.

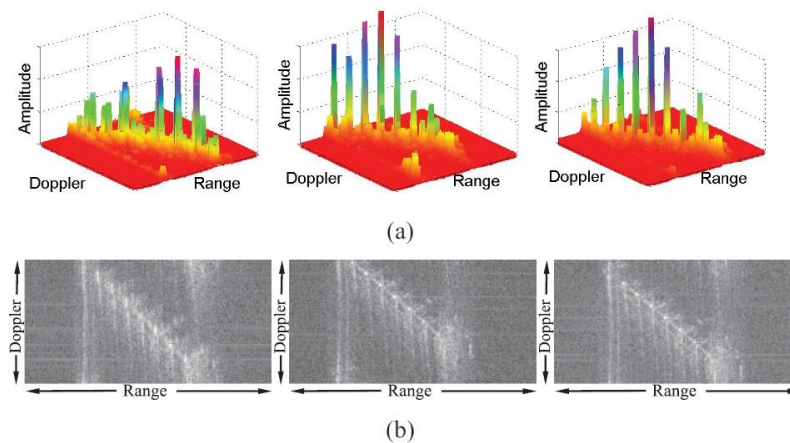
Az olvasónak valóban lehet itt hiányérzete, azonban az ISAR képképzés meglehetősen bonyolult folyamat (és kívül esik a disszertáció tárgyán), amivel nem szerettem volna tovább növelni a fejezet terjedelmét. A közölt eredmény egy 2010 és 2012 között futó Európai Védelmi Ügynökség által finanszírozott projekthez kapcsolódik (APIS), amiben több neves ISAR képképzéssel foglalkozó kutatócsoport is részt vett, többek között a Pisai és a spanyol Alcalá-i Egyetemekről, és a feladat relevanciáját ők validálták.

A [7] folyóiratpublikációban egy valamivel részletesebb bevezető elemzést közlünk az ISAR képképzés alapjairól és kihívásairól, amely társszerzőm, a Pisai Egyetem Radar laboratóriumának kutatója, Marco Martorella segítségével készült. A radarképek analízise során a szegmentáláshoz felhasználtuk az ISAR képfeldolgozás szakirodalomban elterjedt log-normal megközelítését [White & Williams, 1999]. A mért amplitúdóértékeket logaritmikus skálán vizsgáltuk (6. ábra), és az egyes osztályok modelljeit ebben a logaritmikus amplitúdóterben Gaussi eloszlással írtuk le, aminek érvényességét a 7. ábrán bemutatott saját kísérleteim is igazolták. A 6. ábra szintén megfigyelhető, hogy a karakterisztikus jellemzőpontokat okozó úgynevezett domináns visszaverő felületek („dominant scatterers”) a logaritmikus tartományban is jól kivehetők (Giusti et. al., 2011).

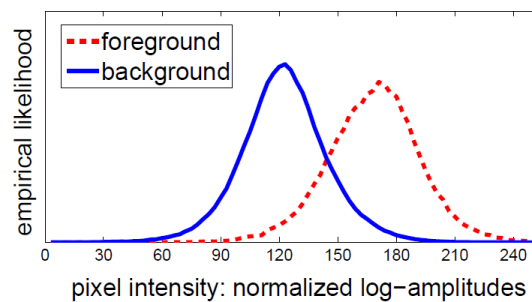
A képfelismerési feladat bonyolultságát tapasztalatom szerint reálisan demonstrálják az itt bemutatott 6. ábra képei mellett a disszertáció 4.3 fejezetében és C.2 függelékében bemutatott ábrák és az összehasonlító eredmények.

Referencia:

(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



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



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

6. [5. fejezet, EMPP, épületdetekciós példa] A Jelölt itt is alaposan kihasználja pl. a tető geometriát, illetve a rajta levő kéménynek még az árnyékát is. Ugyanakkor egy ennyire célirányosan behangolt algoritmusnál felmerül a kérdés, hogy a generalizációs képessége mennyire marad meg, mennyire alkalmazható valamilyen paraméterében más képi inputokon azonos feladatra

A jelölt pontfolyamat modell (JPM) fontos tulajdonsága az energiafüggvény moduláris felépítése, ami a hatékony optimalizációs módszerek alkalmazhatósága mellett azért is hasznos, mert a tervezés során jól elkülöníthetők a különböző modellezésbe bevonni kívánt tényezők: például az adatfüggő objektummegjelenés és a főként geometriai tudásunkból levezethető prior interakciós kényszerek. Az

adattag ($A(u)$ a (2.21) képletben) többnyire alkalmazás-specifikus, tehát más modell szükséges épületeknél, áramköri elemeknél, vagy flamingóknál, és egy alkalmazási területen belül is lecserélhetjük vagy periodikusan váltogathatjuk az adattagokat a megváltozott körülmények függvényében, ahogy azt a 4. kérdésnél korábban részleteztem. Ugyanakkor a prior energiatag ($I(u, v)$ a (2.21) képletben) akár valamennyi előbb említett alkalmazásnál ugyanaz (vagy részben ugyanaz) lehet megegyező paraméterekkel. Elterjedt például az objektumok átlapolódását büntető tag (2.22) képlettel definiált változata, amely a dolgozatban bemutatott szinte valamennyi JPM modellben szerepelt a prior energia egyik részttagjaként.

A dolgozatban bemutatott példákban az automatikus modellválasztást, és megváltozott jellegű bemeneti képek esetén a modellek automatikus adaptációját valóban megnehezíti, hogy az adattag-prototípusok elsősorban kézimunkával szerkesztett („handcrafted”) jellemzőkön alapultak. Itt fontos mindenekelőtt megemlíteni, hogy a dolgozat 3-5 fejezeteiben bemutatott munka nagyobb, alapozó része még a „deep learning” éra előtt, illetve annak kezdetén folyt, amikor természetes választásnak számított ez a megközelítés. Ugyanakkor napjainkban a mélytanuló hálózatok elterjedése kézenfekvő lehetőségeket ad az adattagok általánosabb helyzetekben történő hatékony felírására is. Például, az egyszerűség kedvéért, tekintsük az épületdetekcióhoz használt *tetőszín* jellemzőt. Itt a dolgozatban azzal a feltevéssel éltem, hogy a piros tetők régiói egy egyszerű színszűrővel is kinyerhetők: előbb binarizáltam a képet, majd az illesztett téglalap adattagját a 8. ábrán látható módon a belső R_u rész fehér és a T_u rész fekete pixelinek arányával definiáltam (a dolgozat 46. oldalán ezek a f_{in}^{co} és f_{ex}^{co} jellemzőtagok.)



8. ábra egyszerű színszűrővel működő tetőszegmentáció, majd a szegmentált kép alapján felírható objektum-energiát meghatározó R_u és T_u képrégiók

Általánosabb esetben azonban a 8 ábrán látottakhoz hasonló bináris objektum/háttér maszkok különböző, szemantikus szegmentációt végző mélytanuló hálózatokkal (például U-Net-tel) is generálhatók. Megfelelő mennyiségű és minőségű tanítóadat rendelkezésre állása esetén a szín- vagy árnyékszűrőn alapuló ad-hoc jellemződefiníciós rész akár ki is küszöbölhető az adattag felírása során, a bemutatott jelölt pontfolyamat modell (JPM) többi elemének változatlanul hagyása mellett. Hasonló módon taníthatók a kémények, és egyéb objektumok vagy részobjektumok is, ekkor a modell általánosításának a minősége a neurális hálózat általánosító képességétől függ. (Li et.al 2019) néhány hónapja bemutatott egy megoldást úthálózat-kinyerési feladatra, amely a fenti elveken nyugszik. U-Net-tel detektált egy zajos útmaszkot, amit a JPM modellel adattagjába épített be.

A *Bírálati* kérdés által érintett általános „beágyazott jelölt pontfolyamat modell” (BJPM, vagy angolul EMPP) bevezetésének az volt a fő motivációja a részemről, hogy egy bonyolult hierarchikus szemléletben is képesek legyünk konkrét alkalmazások széles körét egységesen kezelni. A modell strukturális elemeit és az energioptimalizációs algoritmust ezért absztrakt szinten definiáltam és implementáltam, míg a különböző alkalmazások felé egyszerű interfészeket biztosítottam, lehetővé téve a megváltozott körülményekhez, vagy felhasználási területhez igazított rugalmas adaptációt. A BJPM modell adott alkalmazáshoz való illesztéséhez három modellelemet kell csak külön-külön

megadni, minden más egységes: 1) Mindegyik alkalmazás esetén definiálnunk kell az adatfüggő jellemzőket a szülő és a gyermek szintű adattagok kiszámításához – akár a dolgozatban bemutatott *handcrafted* megközelítéssel, akár *mélytanuló* eljárások felhasználásával. Ezen felül meg kell határoznunk a 2) gyermek-szülő interakciókat és 3) az objektumcsoportokba tartozás szabályait az objektum-konfiguráció-szegmens távolság definiálásán keresztül, ezek a dolgozatban szintén eseti módon történtek, de itt is lehet tanuló eljárásokat bevezetni, amire a következő kérdésnél még részletesebben kitérek.

Referencia

(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

7. Egy filozófiai kérdés: hogyan lehetne automatizálni, vagy félautomatikussá tenni egy tulajdonságokon, geometriákon, hierarchiakon alapuló gondolkodást úgy, hogy az emberi leírási szintet még megtartsa, ugyanakkor az implementációja fájdalommentesebbé váljon?

A kérdés bár általánosan feltéve tényleg filozófiai, egyes tartalmi részei jól konkretizálhatók.

Kezdeném az BJPM előző pontban is felsorolt alapvető alkalmazás-specifikus feladataival. Az adattagok kézi jellemzőválasztásának kiküszöbölésére az előző válaszomban javasoltam egy lehetséges megoldást mélytanulás alapú szemantikus szegmentációk felhasználásával. A szülő-gyermek kapcsolati energiatagokat, valamint az objektumokcsoportokat kialakító energiatagokat meghatározó kényszerfeltételek szintén kialakíthatók tanulás útján, az egy szülőhöz tartozó gyermekek vagy az adott csoporthoz tartozó objektumok geometriai/szín/textúra paraméterei között végzett korrelációs vagy hisztogram alapú vizsgálattal, kézzel ellenőrzött referenciaadatokból kiindulva. Például autós forgalmi helyzeteket vizsgálva a tanítómintákból is kinyerhető, hogy az objektumok azonos orientációja jó csoportosító tényező, de például az objektumok színe és mérete nem releváns. Itt a konkrét alkalmazásokat szem előtt tartva érdemes azt is átgondolni, hogy egy adott esetben egy bonyolult hierarchikus modellhez megfelelően címkézett tanítóhalmazt tényleg fájdalommentesebb-e generálni, mint néhány jellemzőt kikísérletezni és ezekre néhány logikus szabályt felírni. Természetesen a mérlegelésnél az is kérdés, hogy a módszert tapasztalt szakemberek paraméterezhetik, vagy célunk, hogy bármilyen felhasználó azonnal tudja őket használni egyedi helyzetekben is.

A kérdés bonyolultabbá válik, ha az általam is vázolt beágyazott pontfolyamat modell (BJPM) funkcionalitását ki akarjuk bővíteni, például a hierarchiaszintek növelésével. Ennek az automatizálása már nem csak számítógépes látással kapcsolatos vizsgálatokat igényel, hanem ontológia modellezést, azaz egy adott feladatra vonatkozó fogalomalkotást, fogalomfeltérképezést is: meg kell határozni például, hogy hány hierarchikus szinten értelmezhetők a vele kapcsolatos részfeladatok egyáltalán. Itt ismét felmerül, hogy a bonyolultabb hierarchiához a referenciaadat (*ground truth*) generálása is egyre bonyolultabb és drágább/hosszadalmasabb feladat is lehet.

8. A másik kérdés, hogy ezek a módszerek minőségben jobbak-e a megfogalmazott tulajdonságokat nem tartalmazó, emberi nyelven nem körülíráható neurális hálózatos megoldásoknál?

A dolgozatban erről a kérdésről valóban kevés szó esik, mivel a tézisek alapját képező tudományos publikációk közlése során mindig a cikkbeküldés idején legjobbnak tartott aktuális szakirodalmi

eljárásokkal szemben igyekeztem igazolni az új eljárásaink előnyeit. A Markov mezős és a jelölt pontfolyamat modell (JPM) alapú megoldások (például az épületdetekciós módszerek) kifejlesztése még a mélytanuló megoldások berobbanása előtt történt, az akkor ismert rivális state-of-the-art módszerek sokkal kevésbé épültek gépi tanulásra. Kivételt képez a CXM hosszútávú változásdetekciós modell (1.2 tézis), amelynek kiértékelésekor három neurális hálózat alapú megoldással is történt összevetés: Hopfield [131], Parzen [132] and MLP [133, 134], itt az általunk javasolt nem-neurális megoldás jobb eredményeket adott ugyanazon az adathalmazon tanítva és tesztelve a módszereket. A 6. fejezetbeli legújabb, Lidaros eredmények szakirodalmi környezete szintén tartalmaz neurális hálós módszereket, melyek elemzése és részben az új modellekkel szembeni összehasonlításuk is megtörtént.

Természetesen az érintett tématerületek mai átfogó elemzése már szinte sehol sem nem kerülheti meg a *deep learning* módszereket, különösen a távérzékelte adatok feldolgozása esetén, ahol nagy mennyiségű adathalmaz és referenciaadat egyaránt rendelkezésre áll az ipari szereplőknél. A 2.3 fejezetben szerepel is egy rövid összevetés a Bayesi modellek és mélytanulási eljárások használhatóságáról a megcélzott területeken. Általánosságban elmondható, hogy a JPM-en alapuló modellezési folyamatban a neurális hálós megoldásoknál sokkal inkább használható a geometriai tudás, top-down modellezési képesség. Mivel nem feketedobozos technikákat kapunk, azok limitációi is jobban körülírhatók, így a hibák bekövetkezésére jobb becslés kaphatunk, illetve célzottan dolgozhatunk a modellek fejlesztésén a szükséges irányban. A JPM módszerek bár feladatspecifikusak, azonban a megszerzett modellezési tudás gyakorlat hatékonyan átvihető egyik feladatról a másikra.

A *Bírálati* kérdésre még néhány releváns új mélytanulós módszert megemlítenék a távérzékelési területről: (Liu *et al* 2018) hierarchikus CNN modellt javasolt épülettetők megtalálására és befoglaló téglalapokkal történő közelítésére, különböző felbontású képeket és különböző méretű épületeket is feltételezve. A cikkben szereplő eredmények ígéretesek, de sok hamis negatív találatot tartalmaznak. Ilyen hibákkal én is találkoztam a módszereim fejlesztése közben, de a hamis negatív találatok oka sok esetben könnyen megtalálható volt analitikus eszközökkel, és új ad-hoc jellemzőkön alapuló újabb prototípus-modellek kifejlesztésével jelentős részben orvosolni tudtuk a hibákat. Ez a konstruktív megoldás a feketedobozos neurális hálózatos modellnél kevésbé tűnik járható útnak, tipikusan új tanulást, gyakran újabb adatok bevonását igényli, és a siker így sem garantált. (Vakalopoulou *et al* 2017) a SegNet mélytanuló hálózat eredményeit kombinálja hagyományos él-jellemzőkkel egy feltételes Markov mezős (Conditional Random Field, CRF) modellben, és egy korlátozott méretű adathalmazon megmutatja, hogy geometriai (szoft-)kényszerek alapján szabályok figyelembe vétele valóban javítani tud a neurális hálós eredményeken. Két friss rövid konferenciatick bemutatja, hogy optikai képek elemzésén túl a mélytanulás alkalmas lehet 3D épületmodellek rekonstrukciójára (Teo 2019), és változásdetekciós feladatok (Yuan *et al* 2019) megoldására távérzékelte képeken – ezeknek a módszereknek a részletes analízise, használhatóságának felmérése és továbbfejlesztése azonban a területen dolgozó kutatóközösség (közel)jövőbeli feladatai közé tartozik.

A Markovi és a mélytanulós módszerek ötvözése távérzékelte adatelemzéshez saját és a kutatólaboratóriumunk jövőbeli tervei között is szerepel, többek között központi témája a korábban már említett, Európai Űrügynökséghez benyújtott pályázatunknak is.

A dolgozatban a CXM modellel összehasonlított neurális hálózatos és nemparametrikus sűrűség-modellező megoldások

[131] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, and A. Ghosh, "A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, pp. 778–789, March 2007. 3.3.3

- [132] L. Bruzzone and D. Fernández-Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images," *IEEE Trans. on Image Processing*, vol. 11, no. 4, pp. 452–466, 2002. 3.3.3 111
- [133] L. Bruzzone, D. Fernandez Prieto, and S. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, pp. 1350–1359, May 1999. 3.3.3
- [134] L. Castellana, A. D'Addabbo, and G. Pasquariello, "A composed supervised/unsupervised approach to improve change detection from remote sensing," *Pattern Recogn. Lett.*, vol. 28, no. 4, pp. 405–413, 2007. 3.3.3

Új mélytanulmányos eredmények épületdetekciós témában

(Liu *et al* 2018) Y. Liu *et al.*, "Multilevel Building Detection Framework in Remote Sensing Images Based on Convolutional Neural Networks," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 10, pp. 3688–3700, Oct. 2018, doi: 10.1109/JSTARS.2018.2866284.

(Vakalopoulou *et al* 2017) M. Vakalopoulou, N. Bus, K. Karantzas and N. Paragios, "Integrating edge/boundary priors with classification scores for building detection in very high resolution data," *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, TX, 2017, pp. 3309–3312, doi: 10.1109/IGARSS.2017.8127705.

(Teo, 2019) T. Teo, "Deep-Learning for Lod1 Building Reconstruction from Airborne Lidar Data," *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, 2019, pp. 86–89, doi: 10.1109/IGARSS.2019.8897810.

(Yuan *et. al* 2019) X. Yuan, J. Tian and P. Reinartz, "Building Change Detection Based on Deep Learning and Belief Function," *2019 Joint Urban Remote Sensing Event (JURSE)*, Vannes, France, 2019, pp. 1–4, doi: 10.1109/JURSE.2019.8808968.

9. Az egyik előnye az itt bemutatott módszereknek bizonyára az, hogy itt 50-80 annotált képpel, és algoritmus fejlesztéssel eljutunk egy olyan szintre, ahova a neurális hálózatok talán eljutnának 5-10 ezer annotált képpel.

Ezzel a megállapítással egyetértek. A valódi termelési folyamatban (például a teljes országról negyedévenként készített műholdképes felmérés osztályozása során a Lechner Tudásközpont Úrtávérzékelési osztályán) is fontos feladat lehet a modellek gyors konfigurálhatósága, márpedig több ezer új tanítóminta generálása rendkívül idő- és erőforrásigényes feladat, míg néhány szabály felírása, és jellemzők néhány annotált kép alapján történő félautomatikus beállítása gyakorlattal rendelkező operátoroknak sokkal gyorsabb lehet.

Budapest, 2020. június 9-én

Benedek Csaba