Benedek Csaba
Image based multi-level environment analysis
(Képi alapú többszintű környezetelemzés)
című MTA doktori értekezésének bírálata

1. Tartalmi bemutatás, kérdések

Az értekezés a képfeldolgozás szakterületén gépi látással kapcsolatos problémákra fókuszál. Alapvetően objektumdetektálási feladatokra építve magasabb szintű környezeti elemzést tűz ki célul, ami az egyszerűbb objektumleírókon túl figyelembe veszi az objektumok egymáshoz való viszonyát, valamint a helyszín időbeli változását is. Az összetett metodológiai megközelítésen túl az értekezésben komoly hangsúlyt kap, hogy a korszerű szenzoradatokra hagyatkozva az adott alkalmazáshoz optimális hardver/szoftverkörnyezetből származó adatok elemzése történhessen. Ez különösen azért nagy kihívás, mert sokszor zajjal terhelt képekkel kell dolgozni, illetve szükségessé válik a különböző képalkotókkal – például többkamerás rendszerek, multispektrális, mélység- és hőkamerák, radarok, lézerszkennerek – méréseinek fuzionálása is egy közös keretrendszerben. Értekezésében a Szerző korábbi eredményeire is támaszkodva alapvetően a képi osztályozásra, objektumdetektálásra, szegmentálásra és bővíthetőségre alkalmas Markov véletlen mezőkön (MRF) alapuló keretrendszert használja. Ezt a rendszert bővíti az adott, valós világból származó feladatok minél pontosabb megoldására. Ennek megfelelően teszi a keretrendszert többrétegűvé, illetve a korábbi következtetési modellek fejlesztéséhez dinamikus markovi gráfokat vezet be a képszegmentációs módszertanba. A MVM megközelítést kiterjesztve a jelölt pontfolyamat modellek (MPP) képpontok helyett geometriai objektumokat tekintenek, mely módszertan lehetővé teszi az objektumszintű viszonyokat is a keretrendszerbe ágyazni. Az MPP alapmodellt a szerző az objektumok időbeli változását, elmozdulását, illetve egymásba ágyazott (hierarchikus) voltának leírását támogató modellé alakítja. Az MRF alapú módszertanoknál nagyon hangsúlyos szerepet kap a modellek hatékony optimalizálása, amely feladatról a Szerző egyik kiterjesztésnél sem feledkezett meg.

A dolgozat egy bevezető, egy metodológiai hátteret áttekintő, az új eredményeket taglaló négy és egy összegző fejezetből áll. A bevezető fejezet célja, hogy a gyakorlati alkalmazások bemutatásán át ismertesse a fő célkitűzéseket és az ezek eléréséhez tekintett módszertani megoldásokat a korábbi eredmények továbbfejlesztésével. A 2. fejezet egy nagyon hasznos összegzést ad a Markov véletlen mezők elméletéről és azok alapszintű alkalmazhatósági módjáról a célfeladatokra, beleértve az optimalizációs kérdéseket is. Példákkal illusztrálva definiálásra kerül a jelölésrendszer is, továbbá maga a fejezet a szakterületen kevésbé jártasok számára is biztosít lényegében egy bevezető kurzust.

A 3. fejezet többrétegű, címkefúzión alapuló Bayesi eljárásokat vezet be alapvetően két távérzékelési (légifotók) feladatra koncentrálva. Az első feladat pár másodperces eltéréssel rögzített légifotókon való objektum-elmozdulás automatikus érzékelésére vonatkozik, ahol értelemszerűen a kameramozgás is nehézséget jelent. A Szerző a 3.2. szakaszban adott megoldásában az alap MRF keretrendszert háromrétegű Markov véletlen mező alapú modellé bővítette (L3MRF). Az eljárás lényege, hogy a két szélső réteg szegmentálása a különböző jellemzőkön alapul, míg a középső réteg a végső változást jeleníti meg. A simaságot a rétegen belüli kapcsolatok biztosítják, míg a szemantikai helyes címkézést a rétegek közötti kapcsolatok adják. A rétegek előállításához a bemeneti képpárt regisztrálni kell, amit a Szerző egy, a Fourier eltolási tételen alapuló eljárással végez. A bevezetett módszer validálása három különböző adathalmazon történt, öt más megközelítéssel összehasonlítva. A 3. fejezet második feladatát (3.3. szakasz) a jelentős időkülönbséggel készül légi fényképek automatikus összevetése adja. Itt a környezeti (pl. szezonális) változások mellett a beépítettség változása jelenti a fő kihívást. A

különböző képi (textúra stb.) jellemzők alapján szegmentált rétegek címke szintű fuzionálása itt is kézenfekvő, viszont indokolt kontextusfüggő következtetési lépések felhasználása. Ennek biztosításához a Szerző a kevert Markov modelleket bővítve feltételes kevert Markov modellt (CXM) vezet be létrehozva adatfüggő dinamikus kapcsolódási lehetőségeket a csomópontok között. A kiértékelés légi- és műholdképek három különböző teszthalmazán történt, négy, a szakterület vezető folyóirataiban közölt korábbi szakirodalmi módszerrel összevetve.

Kérdéseim, megjegyzéseim a 3. fejezethez (1.1. altézis és 1.2. altézis) kötődően:

- 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?
- 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?
- 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 optimalizáción azzal, hogy azon csomópontok címkéi, amelyek az előző valahány iterációban nem változtak rögzítésre kerülnek? (Esetleg egy nagyobb szomszédság figyelembe vételével.)
- Felmerült a 3.3. szakaszban (Task 2) bemutatott bővített modell első feladatra (Task 1) alkalmazhatósága? Összehasonlítható a két modell egy feladaton?
- 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?
- Nagyon pozitív, hogy a szerző a CXM modell esetében összehasonlításokat végzett azóta megjelent eljárásokkal. Történt ugyanilyen összehasonlítás az L3MRF modell esetében, illetve ez a megoldás vajon hogyan viszonyul az azóta megjelent módszerek pontosságához?
- A 3. fejezetben tárgyalt mindkét módszert illetően felmerül, hogy a futási időkkel kapcsolatban mi mondható el róluk (összehasonlítva például a többi tárgyalt külső módszerrel)?

Az értekezés 4. fejezetében a Szerző már olyan modelleket vizsgál, amik támogatják geometriai feltevések integrálását, valamint a szomszédos objektumok kölcsönhatásainak figyelembe vételét a környezetben. A kivitelezés jelölt pontfolyamat modellek (MPP) kiterjesztésével történik, ami a 4.2. szakaszban tárgyalt első célfeladat esetében egy terület házakkal való beépítettsége változásának vizsgálata. Ebben a feladatban a házak téglalapokkal modellezhetőek és figyelembe vehető a kölcsönös viszonyuk. Az újonnan épült házak felismerése egy többidejű jelölt pontfolyamat (mMPP) modell segítségével történik, amely párhuzamosan használ fel az időrétegek között kinyerhető alacsonyszintű változásjellemzőket. Az mMPP modell nyolc, jelentősen különböző légi- és műholdképeket tartalmazó adathalmazon került összehasonlításra a létező szakirodalmi eljárásokkal mind képpont szintű, mind objektum szintű detektálási pontosságra nézve. A 4. fejezet 4.3. szakaszban tárgyalt második feladata ISAR (radar) képsorozaton alapuló objektumdetektálást céloz. A feladatban különösen nagy kihívást jelent a felvételek meglehetősen alacsony szintű minősége, így a leírók esetében alapvetően a fő alakzati tengelyekre, illetve ú.n. karakterisztikus pontokra lehet hagyatkozni, amelyek megbízhatósága azonban képkockánkét változhat. A nehézségek kiküszöböléséhez a szerző egy vonalszakaszokat és

ponthalmazokokat együttesen kezelő többkeretes pontfolyamat modell sémát (FmMPP) vezetett be hajók és repülők struktúráinak automatikus észlelésére és követésére. A módszer kvantitatív kiértékelése nyolc különböző valódi ISAR képszekvencián történt.

Kérdéseim, megjegyzéseim a 4. fejezethez (2.1. altézis és 2.2. altézis) kötődően:

- 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, másrész az utakhoz való geometriai viszonya a házaknak (élek merőlegessége/párhuzamosság) javíthatja a becslést.
- 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?

Az 5. fejezetben szintén egy objektumok viszonyának vizsgálatát követelő kérdéskörrel találkozunk a beágyazott (hierarchikusan felépülő) objektumok detektálásával kapcsolatban. A Szerző egy általános keretrendszert dolgoz ki, ami lehetővé teszi objektumcsoportok és szülő-gyermek kapcsolatban álló objektumok együttes vizsgálatát. A megvalósítás egy általános háromrétegű beágyazott jelölt pontfolyamat (EMPP) modell struktúra kialakításával történik, amely konkrét alkalmazások széles köréhez illeszthető, ahogyan azt a Szerző három gyakorlati területen is illusztrálja: az 5.1. szakaszban bemutatott gyakorlati probléma áramköri elemek felismerésére vonatkozik, melyek belsejében további hibás terület jelenhet meg (ellipszis alakú régiók); az 5.2. szakasz a modell háztető + kémény formájában beágyazott téglalap-alakú objektumok detektálásával foglalkozik; az 5.3. szakasz LIDAR mélységi képeken járművek detektálására fókuszál (téglalap alakú jármű + szélvédő alaki feltételekkel). Mind a három alkalmazási területen részletes kvantitatív kiértékelés történt a hagyományosnak tekinthető szekvenciális megközelítéssel összehasonlítva.

Kérdéseim, megjegyzéseim az 5. fejezethez (3.1. altézis és 3.2. altézis) kötődően:

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

A dolgozat 6. fejezete 4D környezetanalízissel kapcsolatos problémákra koncentrál többkamerás- és LIDAR-ral készült felvételek elemzésével. A 6.2. szakasz egy többkamerás rendszert tekint, ahol a feladat emberek detektálás és mozgásának követése. Módszertani szempontból az előző eredményekhez kapcsolódva – számos előfeldolgozó lépést követően – a megoldást egy henger objektumokat kezelő jelölt pontfolyamat (3DMPP) modell szolgáltatja, részben egymást takaró és egymással érintkező személyek csoportjainak jellemzésére, ami a személyek 3D lokalizációja mellett magasságuk becslésére is alkalmas. A javasolt megoldás két publikusan elérhető adatbázison került kiértékelésre egy aktuális szakirodalmi referenciamódszerrel összehasonlítva. A 6.3. szakasz már forgó többsugaras LIDAR szenzorral készített felvételekkel foglalkozik, ahol az eszköz statikus (földön álló). A feladat itt is emberek detektálása, az adott megoldás pedig alkalmas mozgó személyek észlelésére, követésére, sőt, bizonyos szintű biometriai azonosítására is. A megoldás fontos részét képezi a kritikus térbeli szűrés 3D pontfelhőből származtatott 2D mélységképen való végzése, a pontrácsra történő projekció kvantálási hibáinak a valódi 3D pontpozíciók és a 2D címkék visszavetítésének együttes figyelembevételével való kezelése, valamint a személyek azonosításának rövid- és hosszú távú összerendelése. Bár a szakirodalom a LIDAR alapú emberi viselkedés detektálására nézve szegényes

(például saját tesztadatbázis összeállítására volt szükség), a kvantitatív összehasonlítást sikerült megoldani hasonló eszközöket (pl. Kinect) használó eljárások bevonásával. A 6.4. szakasz a 6.3. szakaszhoz hasonlóan LIDAR adatokkal dolgozik, ahol azonban a szenzor mozgó (autóra szerelt), így számos további nehézség is felmerül. A célul kitűzött dinamikus városi környezet analízisére a Szerző egy új eljárássorozatot dolgozott ki, melyhez referenciaként mobil lézerszkenneléssel (MLS) kapott felvételeket használt sűrű pontfelhőként. Az eljárás több részfeladat megoldására bontható beleértve az objektumdetektálást és -osztályozást, a háttér szegmentálását, a multimodális pontfelhőregisztrációt és a változások detektálását. A korábban más ismertetett MRF alapú módszerek mellett (változás detektálása) 3D szemantikus szegmentációs célra a 6.3. szakaszhoz hasonlóan itt is megjelennek a mély konvolúciós háló (CNN) alapú megközelítések.

Kérdéseim, megjegyzéseim a 6. fejezethez (4.1. altézis, 4.2. altézis és 4.3. altézis) kötődően:

- 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 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?
    - a 6.17. ábra táblázatának Task oszlopában szereplő rövidítéseket nehéz értelmezni.
- A 6.3.6. szakaszban a k=100 és l=60 paraméterbeállítás mi alapján történt?
- A 6.3.4. szakaszban említett CNN esetében történ hiperparaméter-optimalizáció (pl. a struktúrára vagy a szabad paraméterekre nézve)?
- 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 tevékenységcsoportok esetleg valamilyen klaszterezéssel is kialakíthatóak lehetnek, történt ilyen jellegű vizsgálat? 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 6.4. szakaszban a KxKxK-s voxel környezethez a K=23 beállítás hogyan került meghatározásra?
- A 6.4.3. szakaszban bemutatott CNN esetében említésre kerül, hogy különféle architektúrák kipróbálása után véglegesedett a háló szerkezet. Nem lett volna egyszerűbb itt is hiperparaméter-optimalizációval meghatározni az architektúrát?

A dolgozat 7. fejezete az új tudományos eredményeket összegzi három téziscsoport formájában altézisekre bontva. Az ezt követő függelék a felhasznált rövidítések és jelölések listáját, valamint a korábbi fejezetekhez tartozó további részleteket tartalmazza, beleértve az algoritmusok bemutatását is. Az értekezés irodalomjegyzékkel zárul. Az angol nyelvű értekezés mellé magyar nyelvű tézisfüzet is készült.

Az egyes tézispontokhoz kötődően túl, a dolgozat egészét érintően a következő kérdések fogalmazódtak még meg bennem:

- 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?

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

2. *Értékelés*

A nyelvezete szakmailag precíz, de kellőképpen olvasmányos. A munka nagyon jól szerkesztett, kiemelhető a 2. fejezet nagyon hasznos szakterületi bemutatása, a fejezetek végén az eljárások szabad paramétereiről folytatott diszkusszió, valamint az olvasást megtörő részek Függelékbe való átemelése. 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.

A megfogalmazott téziscsoportok altéziseit (1.1, 1.2, 2.1, 2.2, 3.1, 3.2, 4.1, 4.2, 4.3) mindenhol elfogadom, azokat a Szerző tudományos eredményeinek tartom. 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.

Minden egyes téziscsoport/altézis esetében kijelenthető, hogy a Szerző eredményeit nagyon nívós szakmai fórumokon közölte (vezető folyóiratok és konferenciák); ezért is nem tértem ki erre a tartalmi összegzőmben. Ugyanígy elmondható, hogy a szerző mindenhol törekedett a precíz elméleti alapozásra, valamint hogy az eljárások hatékonyságának bizonyítása minden esetben meggyőző kvantitatív összehasonlításokkal, a szakterületi elvárásoknak megfelelően történt. Kifejezetten pozitív továbbá, hogy minden altézis konkrét gyakorlati feladathoz kapcsolódik, a szükséges modellek kialakításának szükségességét valós problémák indukálták.

A dolgozat meggyőzően demonstrálja, hogy a szerző kiváló ismerője a területnek és tudományos munkája mellett az eredményekhez kapcsolódó tudományszervezői/oktatási tevékenysége is kiemelkedő. Szakmai elismertségét az értekezésben szereplő eredmények jelentős hivatkozottsága is mutatja.

A dolgozat eredményeit elegendőnek tartom az MTA doktori cím megszerzéséhez, a nyilvános védés kitűzését és a fokozat odaítélését támogatom.

Debrecen, 2020. április 24.

Prof. Dr. Hajdu András
tszv. egyetemi tanár