

„Képi alapú többszintű környezetelemzés”

c. MTA Doktori Értekezéséről

Az értekezés egy kifejezetten aktuális témakörben íródott. A modern térérzékelő eszközök segítségével egyre több mérési lehetőség nyílik meg, és ezekből sokrétű és lényegesebb nagyobb tömegű - a korábbi érzékelő architektúrák minőségi és mennyiségi képességén lényegesen túlmutató- adattípusok, illetve nagy adathalmazok jelennek meg. Ezen szenzorok megjelenésével új mérési és környezet megismerési/feltérképezési valamint változáskövetési, igények is megjelennek. A dolgozatban adott megoldások részben a térképészet, a városrendezés, részben a közlekedésirányítás és a közlekedésbiztonság, részben az okos város, részben pedig az ipari kihívásaira adott tudományos válaszokként születtek.

Felépítés

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.

Az értekezés fő része – az eredményeket bemutató 3-as, 4-es, 5-ös, és 6-os fejezetek – két fő csoportra oszthatók. Az első csoport három fejezetében a Jelölt egy nagy ívben bemutat három módszer családot, amelyet távérzékelési - elsősorban légifelvételek, illetve légifelvétel párok - kiértékelésére fejlesztett ki. Az első egy pixel alapú szegmentálás, a második már modell alapú detekció, míg a harmadikban a detektált alakzatok egymáshoz képesti hierarchiájának vizsgálatára is sor kerül. Ugyanakkor mindegyik módszernek megmutatja az általánosítási lehetőségeit is más modalitású adatokkal.

A második csoportba a 6. fejezet tartozik, amely csokorra gyűjt három - különböző szenzorokkal elért - egymástól nagyjából független új távérzékelési eredményt. Itt kerül bemutatásra egy több kamera által megfigyelt térrészben az emberek követésére, illetve magasságuk meghatározására szolgáló algoritmus, valamint egy álló platformos forgó LIDAR alapú térfigyelő algoritmus, továbbá egy mozgó platformos forgó LIDAR alapú városi dinamikus látvány analízis. Míg az első csoportban az Jelölt Markov Random Field alapú módszereket hagyományos matematikai módszerekkel egészített ki, addig a másodikban már konvolúciós neurális hálózatokat is alkalmazott.

Formai szempontok:

A Jelölt a doktori értekezést angolul írta, mégpedig igen jó angolsággal, és jó stílusban. Az ábrák informatívak, jól szerkesztettek. Bár az értekezés - mint a műszaki szövegek általában - tele van rövidítéssel, az értelmezhetőséget, illetve követhetőséget egy rövidítés, illetve jelölés gyűjtemény segíti.

Formai szempontból az egyetlen aggályos pont az anyag mérete. A doktori szabályzat 100 oldalban maximálja az értekezés tartalmi részét, ami ebben az anyagban 140 oldal. Az egész dolgozat közel 200

oldal, a függelékekkel, illetve a tartalomjegyzékkel együtt. Ez ugyanakkor nem egy túlságosan bőbeszédűre eresztett stílus eredménye, hanem annak, hogy a Jelöltnek nagyon sok eredménye született az elmúlt időszakban, és ezeket foglalta össze az anyagba.

Tartalmi szempontok és tézisek:

Az első téziscsoportban a Jelölt bemutat két algoritmust, amelyek többretegű címkefűzión alapulnak. Mindkettő eljárás olyan képpárokon keresi a releváns különbséget, amelyeket különböző időpontokban készítettek, nagyjából - de nem pontosan - ugyanabból a kamera pozícióból. Mindkét eljárás speciálisan az alkalmazásra tervezett tulajdonság kinyeréssel indul, amelyet MRF alapú többretegű szegmentáció követ. Az első algoritmus elmozduló autókat talál meg egy mozgó kamera egymás utáni képkockáin. Itt két tulajdonságvektort alkalmaz a Jelölt, az egyik differencia alapú, a másik korreláció alapú. 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 (α_s) kivágott képszegmens korrelációjának magas korrelációs csúcsot kellene adni. Ugyanakkor a képen csak a mozgó objektumok, illetve a nagy homogén területek emelkednek ki korrelációs csúcsként.

Számomra a második algoritmus eredménye volt igazán izgalmas. Itt kertvárosi településekről több éves időkülönbséggel készített légifelvételeket hasonlított össze a Jelölt, és kereste meg rajta az újonnan beépített területeket. A képek természetesen nem ugyanabból a pontból készültek, nem feltétlenül ugyanabban a napszakban és nem is ugyanabban az évszakban, azaz a megvilágítási, illetve a fák lombosodási viszonyai is különbözhetnek. Itt a három kiemelt tulajdonság kombinációjára használta a Jelölt az imént bemutatott többretegű szegmentációs technológiát. A kérdésem ennél a példánál az, hogy hogyan értelmezhető-e az algoritmus egy téli és egy nyári felvételen páron, ahol a vegetáció drasztikusan különbözik?

A második téziscsoportban megjelenő tér-időbeli jelölt pontfolyam modellek már nem pixel szintű, hanem objektum szintű szegmentálást eredményeznek. Az algoritmus itt már lényegesen komplexebb, mivel különböző geometriai modell alapú leírókat használ, illetve alkalmazza a Jelölt által továbbfejlesztett többszörös születési és halál optimalizációs eljárást. Itt is két gyakorlati példán keresztül ismerjük meg az algoritmust. Az első ismét a többéves különbséggel felvett légifelvételek kiértékelése. Itt a leírók között megjelenik a ház alaprajzának geometriai modellje, a tető színe és alakja, és a ház mellett megjelenő árnyék. Az eredmény nagyon jó minőségű, objektum szintű változás térkép. Itt az lenne a kérdésem, hogy mennyire limitálja az algoritmus alkalmazását, ha a képet felhős időszakban készítették, azaz az árnyékok nem látszanak?

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.

A harmadik téziscsoport kiterjeszti a második téziscsoportban ismertetett *jelölt pontfolyam modellt* azáltal, hogy több rétegben beágyazza azokat, így azok le tudnak írni már populációkat, illetve a populációkon belüli hierarchiákat - csoportokat, illetve azon belül is szülő-gyerek relációkat.

Az így kapott *beágyazott jelölt pontfolyam modell* három példán keresztül kerül bemutatásra. Az első szintén a kertvárosi légifelvételen az épületek megtalálása, illetve alakjuk, állásuk és tető típusuk szerinti osztályozása, valamint a kémények kiemelése. A Jelölt itt is alaposan kihasználja pl. a tető geometriát, illetve a rajta levő kéménynek még az árnyékát is. Ezzel az algoritmus nagyon szelektívvé válik, és kifejezetten jól felismeri a mintaképeken az épületeket. 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 téziscsoport keretében bemutatott másik két példa a légi LIDAR felvételeken a gépjárművek felismerését és hierarchikus csoportosítását, illetve nyomtatott áramkörök analízisét célozzák meg.

Az első három téziscsoportban bemutatott megoldások nagyrészt emberi nyelven is megfogalmazható egyéni vagy csoportos többszintű tulajdonság leírók matematika nyelvére történő virtuóz lefordításával alakultak ki. Ezek ennek megfelelően precízen a problémára (és talán a képállományra is) optimalizált tulajdonságok összegyűjtése (1-es téziscsoport), az ember által megfogalmazható geometriai leírók alkalmazása (2-es téziscsoport), illetve az ember által szintén verbálisan leírható képi hierarchiák megfogalmazása (3. téziscsoport). Ezek egyrészt abból a szempontból üdvözlendők, hogy nem bekötött szemmel megyünk neki egy komplex probléma megoldásának, ahogy azt a neurális hálókat alkalmazó megoldások teszik, hanem valamilyen szinten logikailag megfogalmazható és ember számára érthető egy-egy képi döntés. Ugyanakkor, ezeknek az ember által is megfogalmazható leíróknak a matematikai megfogalmazása és behangolása valószínűleg egy időigényes művelet, ami ráadásul nem garantált, hogy egy más paraméterekkel rendelkező képseregen hangolás nélkül működik. Egy filozófiai kérdés: hogyan lehetne automatizálni, vagy félautomatikusá tenni egy tulajdonságokon, geometriákon, hierarchiákon alapuló gondolkodást úgy, hogy az emberi leírási szintet még megtartsa, ugyanakkor az implementációja fájdalommentesebbé váljon? A másik kérdés, hogy ezek a módszerek minőségben jobbak-e a megfogalmazott tulajdonságokat nem tartalmazó, emberi nyelven nem körülírható neurális hálózatos megoldásoknál? 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.

A negyedik téziscsoport már nem tartozik az előző három ívébe, sőt az itt bemutatott három eljárás is csak annyiban kapcsolódik össze, hogy mindegyik 4D-s adatsorok értelmezésével, illetve kiértékelésével foglalkozik. Az első egy emberi alakokat kezelő jelölt pontfolyam modell, amely többkamerás perspektívából készített képeken henger modelleket alkalmazva alkalmassá válik emberek 3D lokalizációjára, illetve magasságuk becslésére.

A második két algoritmus LIDAR képek értékelésével foglalkozik. Az első különlegessége, hogy a forgó LIDAR viszonylag ritka pontfelhőjéből emberek biometrikus jellemzőinek kinyerését, és ez alapján való követését teszi lehetővé, és még szokásos emberi cselekedeteket is felismer.

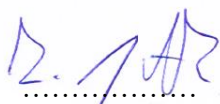
A harmadik algoritmus városi környezetben autóról készített ritka pontfelhőt regisztrál mobil lézerszkenneléssel készített sűrű pontfelhőhöz.

A negyedik téziscsoport algoritmusai már nemcsak az eddig alkalmazott Markov mezőkön, illetve más, klasszikusnak számító algoritmusokon alapulnak, hanem azokat konvolúciós hálózatokkal kombinálják.

Összefoglalás

Összefoglalásként elmondható, hogy az értekezés egy nagyon fontos és korszerű területen mutat be jelentős tudományos hozzájárulásokat. Ezek egyrészt az anyagban bemutatásra kerülnek, másrészt az eredményeket nagyszámú, kiugróan magas impakt faktorú publikáció is alátámasztja. Ennek megfelelően javaslom az értekezés nyilvános vitára bocsátását, és az összes bemutatott tudományos tézis elfogadását.

Budapest, 2020. március 31.


.....
Zarándy Ákos
MTA Doktora