PÁZMÁNY PÉTER KATOLIKUS EGYETEM
ROSKA TAMÁS MŰSZAKI ÉS TERMÉSZETTUDOMÁNYI
DOKTORI ISKOLA

Kovács Lóránt

3D-s változásdetekció és emberpozíció-becslés Lidar érzékelésben

PhD Disszertáció tézisei

Témavezető:
Prof. Dr. Benedek Csaba DSc

Budapest, 2024

1. Bevezetés

A 3D érzékelési technológia jelentős fejlődésen ment keresztül az elmúlt évtizedben, ami jelentősen javította az összetett környezetek automatizált elemzésének és megértésének lehetőségeit. A disszertációban a 3D gépi érzékelés két kutatási területén mutatok be új eredményeket elsősorban földi mobil lézerszkennerekkel rögzített Lidar pontfelhők feldolgozására fókuszálva.

Az első kutatási téma különböző időpontokban rögzített Lidar pontfelhők összehasonlításával történő 3D változásérzékelés, amely kulcsfontosságú lépés különböző alkalmazások esetén, többek között a városervezésben, a környezetfigyelésben és a városi infrastruktúra karbantartása során.

A második kutatási téma az emberi póz becslése, amely a különböző testrészek helyzetének felismerését és becslését jelenti. Míg a pózbecslés a szakirodalomban általában optikai kameraképek felhasználásával történik, kutatásomban kizárólag a Lidar-adatok e célra történő felhasználásának megvalósíthatóságát és előnyeit vizsgálom.

A disszertációban bemutatott eredmények rávilágítanak a Lidar technológia alkalmazásában rejlő lehetőségekre az automatizált rendszerek 3D érzékelési képességeinek fejlesztése során, megnyitva az utat új innovatív alkalmazások és továbbfejlesztett módszerek elterjedése előtt a különböző területeken.

A két témát az 1.1. és 1.2. fejezetben mutatom be, majd a kutatásomhoz használt két különböző típusú Lidar-szenzor leírása következik a 2.1. és 2.2. fejezetben.

1.1. Változásfelismerés

A növekvő városi népsűrűség, valamint az intelligens városi alkalmazások és az autonóm járműtechnológiák gyors fejlődése miatt egyre nagyobb igény mutatkozik a városi infrastruktúrák automatikus monitorozása és a térfigyelő alkalmazások széleskörű intelligens funkciókkal történő ellátása iránt. Alapvető fontosságú környezetünkben a potenciálisan veszélyes helyzetek felismerése, amelyeket például a hiányzó közlekedési táblák, a megfakult felfestések és a sérült utcabútorok okoznak. A városüzemeltetési szolgáltatóknak forrás- és időigényes feladat nagy kiterjedésű városi területeket lefedő, rendszeresen készített és frissített képi és 3D téradatok formájában rendelkezésre álló mérések folyamatos elemzése és összehasonlítása, amely elengedhetetlen lépés a releváns

környezeti változások felfedezéséhez.

A gépi érzékelés tudományterülete szempontjából a fenti feladat változásfelismerési problémaként fogalmazható meg. A videó alapú térfigyelő alkalmazásokban [16,17] a környezetelemzés egy elterjedt megközelítése a változásérzékelésen, azaz a helyszín háttér modelljének becslésén és az új mérések ezen háttérrel való összehasonlításán alapul. A változásdetekció számos távérzékelési alkalmazásban is gyakori feladat, beleértve a légi képek, pontfelhők vagy más mérési módozatok közötti különbségek kinyerését [18, 19]. A rendelkezésre álló megközelítések túlnyomó többsége azonban feltételezi, hogy az összehasonlított pontfelhők vagy 3D pontfelhő mérések regisztráltak, azaz térben pontosan illeszkednek egymáshoz, mivel a méréssorozat során az érzékelők vagy mozdulatlanok, vagy a pontos pozíció- és orientációs paraméterei ismertek az egyes mérések időpontjában. Kutatásom során olyan helyzetekkel foglalkoztam, amikor ezek a feltevések nem teljesülnek, tehát a különböző időpontban készített téradatok közötti eltéréseket a pontos regisztráció feltételezése nélkül kell kinyerni.

1.2. Emberi testhelyzet becslése

A pózbecslés fő feladata az emberi test anatómiai kulcspontjainak lokalizálása. Az emberi testhelyzet becslése a gépi érzékelés egyik alapvető feladata, és számos valós alkalmazással rendelkezik többek között a robotikában [20], a biztonságtechnikában, térfigyelésben [21, 22] és az autonóm vezetés [23] területén.

Az emberi póz becslését általában kamera alapú módszerekkel oldják meg [24–26] a kamerák 2D képterében. Az ilyen megoldásokat azonban eleve korlátozza az, hogy a kamera nem alkalmas a valós távolság közvetlen mérésére, a rögzített képek érzékenyek a különböző fény- és időjárási viszonyokra, valamint a környezetben előforduló 3D tárgyak és alakzatok 2D-s megjelenése csak korlátozott, kameranézet-függő információt tartalmaz.

A mélységi információ figyelembevétele növelheti a pózbecslés robusztusságát, ahogyan azt a [20] tanulmány is mutatja, amely RGBD-kamerát használ az emberi póz 3D-s becslésére, és melynek eredménye felülmúlja a kamera-alapú 3D-s pózbecslőket és a kizárólag mélység információt használó módszereket. Az ilyen elven működő eszközök kültéri használata azonban erősen korlátozott az infravörös fény alapú mérési technológia korlátai miatt, így a Lidar alkalmazása természetes igényként merülhet fel a területen.

A Lidar alapú megközelítések fontos szerephez juthatnak olyan alkalmazásokban is, ahol a magánélet védelme kiemelt jelentőségű, mivel a megfigyelt embereket a kezelő személyzet nem tudja azonosítani a Lidar pontfelhő ritka jellege miatt.

2. Lidar szenzor

A Lidar egy aktív érzékelő, amely lézersugarak kibocsátásával világítja meg a környezetet. A távolságok pontos mérése a felületekről érkező lézerreflexiók feldolgozásával történik.

Általánosságban a Lidar úgy működik, hogy a látómezőjét egy vagy több közeli infravörös (NIR) lézersugárral pásztázza.

A lézersugár a környezetből visszaverődik az érzékelőre, a visszavert jelet egy fotodetektor fogadja. Egy gyors elektronika megszűri a visszaverődött jelet, és megméri a kibocsájtott és a visszavert jelek közötti különbséget, amely arányos a megfigyelt tárgy távolságával. A távolságot az érzékelő modellje e számított időkülönbség alapján becsüli meg. A Lidar 3D pontfelhőket rögzít, amelyek megfelelnek a letapogatott környezetnek, valamint rögzíti a visszavert lézer intenzitását, ami a visszavert lézernyalábok energiájával feleltethető meg [27]. A Lidar maximális érzékelési távolságát a lézer teljesítményére vonatkozó, szem védelme miatt alkalmazott határértékek korlátozzák.

A Lidar-érzékelő pásztázó rendszere a megfigyelt tér gyors feltárásáért felelős. A különböző Lidar-típusoknál többféle pásztázási megközelítés létezik, ezek közül néhányat az alábbiakban bemutatok.

A *mechanikus forgó típusú* érzékelőkben (rotating multi-beam (RMB) Lidar) a lézersugarakat egy forgó érzékelőfejen keresztül irányítják, amelynek belsejében egy mozgó tükör és optika található. Az általam használt Lidar, amelyet a változásérzékelési kutatásaimhoz használtam (1. Tézis), ezen elv szerint működik, részletesen a 2.1. fejezetben mutatom be.

Egy másik mechanikus megközelítés *prizmák forgatását* használja a lézersugarak irányítására. Az általam a csak Lidarral végzett, emberi pózbecsléssel kapcsolatos kutatásomhoz használt Lidar (2. Tézis) ezt a pásztázási módszert követve működik, amelyet részletesen a 2.2. fejezetben írok le.

A pásztázás egy chipben lévő „tükör” rugó- és elektromágneses erőkkel történő mozgatásával is megvalósítható egy *Mikroeletromechanikus rendszerben* (MEMS) [28].

(a) **(b)**

1. ábra. Velodyne HDL-64 forgó többsugaras Lidar érzékelő és az általa rögzített pontfelhő városi környezetben

A *Flash Lidar* nem rendelkezik mozgó komponensekkel [29]. Egyetlen kibocsátott lézersugarat egy optikai diffúzor szór szét, hogy megvilágítsa az egész megfigyelt környezetet, és a visszaverődéseket egy fotodióda mátrix érzékeli.

2.1. Velodyne HDL-64 forgó többsugaras Lidar

A Velodyne HDL-64 érzékelő (látható az 1a ábrán) egy nagy felbontású és nagy teljesítményű, forgó többsugaras Lidar-érzékelő, amelyet arra terveztek, hogy segítse az autonóm eszközök valós idejű környezetérzékelését. Nagy felbontású és valós idejű 3D méréseket rögzít a környezetéből. Az érzékelő 64 lézersugárral rendelkezik, amely $26,9°$ függőleges látómezőt (FoV) határoz meg. Az érzékelő forgó fejének köszönhetően a vízszintes látómezője 360°. A mért adatok térbeli pontossága 1-2 cm. Az érzékelő karakterisztikája miatt a pontsűrűség az érzékelőtől való távolsággal gyorsan csökken.

A Velodyne HDL-64 az RMB Lidarok úttörője. Újabb generációs RMB Lidar szenzorok elérhetőek már (pl. az Ouster gyártótól), amelyek hasonló műszaki jellemzőkkel rendelkeznek, azonban méretük és fogyasztásuk jelentősen csökkent. Így a Velodyne Lidarral végzett mérések és a jelen kutatásban végzett kutatások még mindig aktuálisak [30].

A rögzített pontfelhőkben körkörös mintázatok figyelhetőek meg,

amint az 1b ábrán is látható, ahogy a lézersugarak az érzékelő függőleges tengelye mentén forognak. Az érzékelő folyamatosan továbbítja a 3D méréseket, amelyeket különálló pontfelhőkbe gyűjt az érzékelőfej minden egyes vízszintes körbefordulását követően.

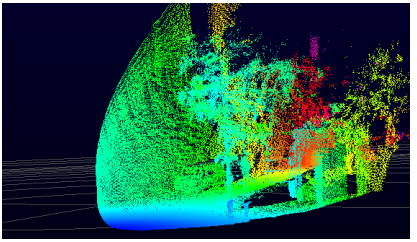
2.2. Livox Avia Lidar érzékelő nem ismétlődő körkörös pásztázással

A 2. ábrán látható Livox Avia érzékelő [31] egy kis- és könnyű Lidar-érzékelő, amelyet egyedülálló, nem ismétlődő körkörös szkennelés (NRCS) jellemez. Az érzékelő hat Lidar-sugárral rendelkezik, amelyek egy vonalban vannak elhelyezve, és az érzékelő belsejében egy prizmát mozgatva és forgatva pásztázza a látómezőjét (vízszintesen 70°, függőlegesen 77°).

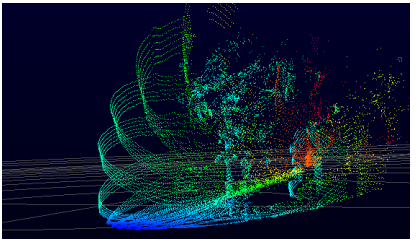
2. ábra. Livox Avia Lidar érzékelő

A legtöbb RMB Lidarral ellentétben, amelyek ismétlődő pásztázási mintát használnak, az Avia nem ismétli meg a pontos pásztázási pályákat minden pontfelhőn, hanem a lézerek a látómező új részeit fedik le. Ez a kulcsfontosságú különbség egyszerre előnyös és hátrányokkal is jár. Az NRCS Lidarok az idő múlásával a teljes látómezőt lefedik, részletes térbeli információt szolgáltatva, különösen egyhelyben álló szenzor esetén. Másrészt, mivel ugyanazt a területet ritkábban pásztázzák, mint a „hagyományos" RMB Lidarok, a dinamikus objektumok, például az emberek, kihívást jelenthetnek, mivel a mozgásuk miatt jelentős elmosódással készül róluk a pontfelhő. Az érzékelő folyamatosan rögzíti a távolságméréseket a megfelelő időbélyegekkel együtt, a látómezőjében a nem ismétlődő körkörös mintázatot követve. Egy rögzített integrálási idő beállításával az egymás után gyűjtött pontok külön Lidar-pontfelhőkbe csoportosíthatók. A fő kihívás a rögzített távolsági adatok térbeli és időbeli felbontása közötti hatékony egyensúly megteremtése.

Miközben nagyobb integrálási idő használata esetén a lézersugarak a látómező nagyobb részét fedik le, ami a pontfelhő nagy térbeli felbontását eredményezi, a megfigyelt területen lévő dinamikus objektumok mozgása különböző nem várt helyzeteket (pl. elmosódott gyalogos sziluetteket) idéz elő, amelyek nem teszik lehetővé a dinamikus események hatékony elemzését. A Livox Avia érzékelő 240000 pontot gyűjt $1s$ időablakon



(a) 1s integrálási idővel rögzített pontfelhő



(b) 100 ms integrálási idővel rögzített pontfelhő

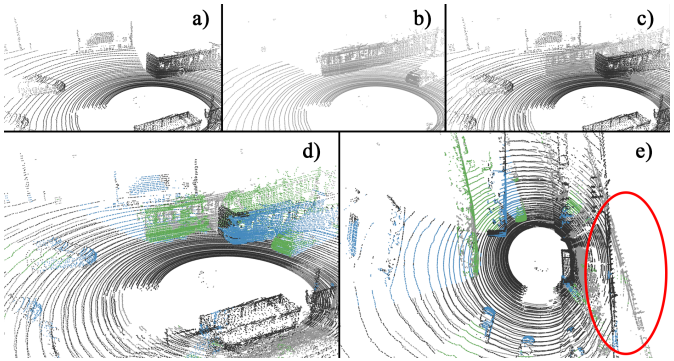
3. ábra. Livox Avia pontfelhők, a pontok távolsága szerint színezve

belül, amint az a 3a ábrán látható. Másrészt, ha a mérések csak egy szűk időablakban kerülnek gyűjtésre (pl. 100 ms alatt), az így kapott pontfelhők nagyon ritkák és kevésbé részletgazdagok. Egy 24000 pontból álló minta pontfelhő látható a 3b ábrán.

3. Új tudományos eredmények

1. Tézis: Új változásdetekciós módszert javasoltam komplex városi környezetben rögzített, egymáshoz pontatlanul regisztrált RMB Lidar pontfelhők alapján. A pontfelhő méréseket mélységképekként tárolom, és a módszer kimenete egy olyan bináris maszkpár, amely mindkét bemeneti mélységképen megmutatja a változások régióit. Az így származtatott változásmaszk információvesztés nélkül visszavetíthető a bemeneti pontfelhőkre. A javasolt eljárást különböző gyakorlati példákon értékeltem ki, és megmutattam, hogy a módszer jobb eredményre képes, mint a meglevő legkorszerűbb változásfelismerő eljárások.

A módszert egy folyóiratban [1] és egy benyújtott szabadalmi bejelentésben [3] publikáltam. A kutatás kezdeti fázisaként az [5] konferenciapublikációban egy módszert írtam le több objektum felismerésére városi környezetben 3D háttértérképek és objektum-követés felhasználásával. A módszer egy sűrű 3D várostérképet használ a Lidar-érzékelőből szárma-

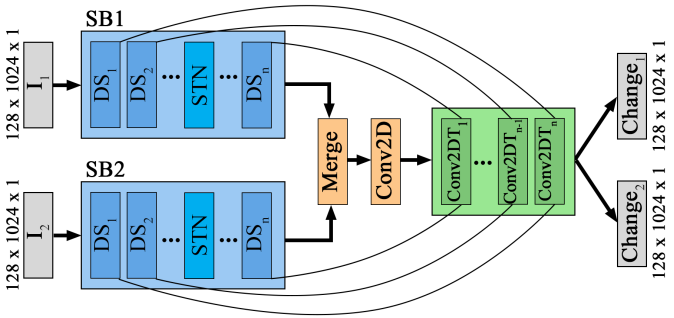


4. ábra. A javasolt módszer (*ChangeGAN*) által észlelt változások egy pontatlanul regisztrált pontfelhőpáron. (a) és (b) a két bemeneti pontfelhőt mutatja, (c) a pontatlanul regisztrált bemeneti pontfelhőket mutatja közös koordinátarendszerben. (d),(e) a változásdetekció eredményeit mutatja be: a kék és zöld színű pontok az első és a második pontfelhőben változásként megjelölt pontokat jelölik. A piros ellipszis a két pontatlanul regisztrált pontfelhő közötti pozíció- és irány-különbségre hívja fel a figyelmet.

zó ritka pontfelhőn történő objektum detekció pontosságának növelésére. Ez a módszer kibővítheti a [6]-ban leírt kamera alapú járműérzékelést. Az objektumok pályájának kiértékeléséhez egy pályát-pályához rendelő kiértékelési módszer [7] használható.

A pontatlanul regisztrált pontfelhőpárok között egyes alakzatok elmozdulásával keletkező különbségek pontonkénti detekciójának szükségességét olyan gyakorlati példákkal lehet hangsúlyozni, ahol a jelenleg rendelkezésre álló módszerekkel nem lehet megbízható regisztrációt és így változásdetekciót elérni. A változásdetekciós problémafelvetéshez egy új megközelítési módot fogalmaztam meg: egy pontatlanul regisztrált pontfelhőpár közötti különbség anélkül felismerhető, hogy a bemeneti pontfelhők mérési körülményei (pozíció, irány) pontosan egyeznének.

A javasolt módszer (ChangeGAN) kulcsfontosságú jellemzője, hogy nem szükséges hozzá a pontfelhőpárok pontos regisztrációja. Kísérleteim alapján a javasolt módszer hatékonyabb, mint a létező megoldások, és hatékonyan képes kezelni a megfelelő 3D pontfelhők közötti akár 1 m-es eltolást és az akár 10°-os forgatási eltéréseket. A 4. ábrán láthatóak a Velodyne HDL-64 forgó többsugaras Lidarral rögzített bemeneti pontfelhők és a javasolt módszer eredményei.



5. ábra. A javasolt *ChangeGAN* architektúra. A komponensek jelölései: SB1, SB2: Siamese ágak, DS: leskálázás, STN: térbeli transzformátor-hálózat, Conv2DT: transzponált 2D konvolúció

1.1. Altézis: Létrehoztam egy mély neurális hálózati struktúrát, amely képes megtanulni és robusztusan felismerni a változásokat pontatlanul regisztrált ritka 3D pontfelhők között, amelyek egy komplex városi környezetből származnak. A módszer tanításához részben automatikus módszert javasoltam pontatlanul regisztrált pontfelhőpárokból álló adatbázis létrehozására, szimulált regisztrációs hibák felhasználásával.

A javasolt mély tanulási módszer bemenete két pontatlanul regisztrált 3D pontfelhő, amelyeket egy RMB Lidar szenzorral rögzítettek ($\mathcal{P}_1$ és $\mathcal{P}_2$), ezeket I_1 és I_2 mélységképként reprezentálom (Ez látható az 6a és 6b ábrán). A javasolt architektúra feltételezi, hogy az I_1 és I_2 mélységképek ugyanakkora pixelrácson helyezkednek el.

A fenti célra egy új, generatív, versengő neurális hálózat (GAN) architektúrát javasoltam, pontosabban egy diszkriminatív módszert, amely egy további versengő-megkülönböztető komponenssel rendelkezik. A módszer elnevezése *ChangeGAN*, amelynek struktúrája a 5. ábrán látható.

Mivel a fő cél az I_1 és I_2 bemeneti mélységképek közötti összefüggések megtalálása, a bemeneti mélységképpárokból a releváns jellemzők kinyerésére Siamese-jellegű [32] architektúrát alkalmaztam. A Siamese architektúrát úgy tervezték, hogy a háló súlyparamétereit a párhuzamos ágak között megosztja, ami lehetővé teszi, hogy a bemeneti adatokból hasonló jellemzőket nyerjen ki, valamint egyúttal csökkenti a memóriahasználatot és a tanítási időt. A Siamese hálózat mindkét ága leskálázó

konvolúciós (downsampling) blokkokból áll.

A javasolt modell második része transzponált konvolúciós rétegek sorozatát tartalmazza, hogy a kinyert leírókat az alacsony dimenziós leírótérből a 2D-s bemeneti mélységképek eredeti méretére felskálázzuk. Végül egy 1×1 konvolúciós réteg, amelynek szigmoid az aktiválási függvénye, létrehozza a két bináris Λ_1 és Λ_2 változás-térképet.

A hálózat általánosítási képességének növelése- és a túltanulás megakadályozása érdekében az első két transzponált konvolúciós réteg után egy dropout réteg következik. A változásfelismerés eredményének javítása érdekében az U-háló [33] egy ötletét adaptáltam, amikor a DS blokkokból származó nagyobb felbontású jellemzőket is adunk a megfelelő transzponált konvolúciós réteg bemenetére.

Ebben az esetben, mivel a pontfelhők pontatlanul vannak regisztrálva, a bemeneti mélységképek azonos régiói nem feltétlenül korrelálnak egymással. A pontosabb jellemző-illesztés elérése érdekében mindkét Siamese ághoz térbeli transzformációs hálózati (STN) blokkok [34] kerültek hozzáadásra. Az STN képes megtanulni egy optimális affin transzformációt a jellemzők között, hogy csökkentse a bemeneti mélységképek közötti térbeli regisztrációs hibát. Továbbá az STN dinamikusan transzformálja a bemeneti mélységképeket, ami szintén előnyös augmentációt eredményez.

Létrehoztam egy új, Lidar-alapú városi adathalmazt *Change3D* néven. A mérések Budapest belvárosában, két különböző napon kerültek rögzítésre egy autó tetejére szerelt Velodyne HDL-64 forgó, többsugaras Lidarral.

Mivel a javasolt *ChangeGAN* módszer fő célja a változások kinyerése a pontatlanul regisztrált pontfelhőkből, a modell betanításához és kiértékeléséhez nagy számú és pontosan felcímkézett, ugyanazon a területen gyűjtött, különböző térbeli eltolási és forgatási különbségekkel rendelkező pontfelhőpárok adatbázisára van szükség. Az annotáció pontosan megjelöli az olyan objektumokat vagy pontfelhő részleteket, amelyek vagy csak az első pontfelhőn, vagy csak a második pontfelhőn vannak jelen, valamint azokat is, amelyek változatlanok, így mindkét pontfelhőn megfigyelhetőek.

A pontfelhő-különbségek kézi bejelölése nagy kihívást jelent még akkor is, ha a pontfelhők azonos koordináta-rendszerből származnak. A tanító adatbázis pontosságának biztosítása érdekében a változások bejelölését az azonos szenzorpozícióból és orientációból rögzített regisztrált pontfelhőpárokon kezdem, majd véletlenszerűen módosítom a második pontfelhő pozícióját és irányát, így nagyszámú, pontosan címkézett, pon-

(a) I_1: a $\mathcal{P}_1$ mélységképe

(b) I_2: a $\mathcal{P}_2$ mélységképe

(c) Λ_1: az I_1 mélységképen megfigyelhető változások maszkja

(d) Λ_2: az I_2 mélységképen megfigyelhető változások maszkja

6. ábra. ChangeGAN bemeneti adatok. (a), (b): I_1, I_2 mélységképek egy pontatlanul regisztrált pontfelhőpárról ($\mathcal{P}_1$, $\mathcal{P}_2$). (c), (d): Λ_1, Λ_2 bináris változási maszkok a I_1 és I_2 mélységképekhez.

tatlanul regisztrált pontfelhőpárt kapunk.

A pontatlanul regisztrált pontfelhőpárok szimulálására véletlenszerűen alkalmaztam egy legfeljebb ± 1 m eltolást és egy legfeljebb $\pm 10°$-os forgatást a z-tengely körül minden egyes pontfelhőpár második pontfelhőjére ($\mathcal{P}_2$). A módszer teljesítményét kiértékeltem fixen beállított forgatási és eltolási paraméterekkel létrehozott adatbázisokon. A GT-címkék a $p \in \mathcal{P}_2$ pontokhoz csatolva maradtak, és velük együtt kerültek transzformálásra, amint az az 6c és 6d ábrán látható.

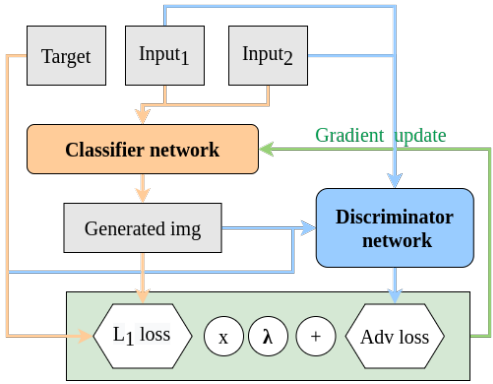
A következő lépésben térbeli szűrést alkalmaztam, csak az 5 m-nél alacsonyabban levő és a 40 m-nél közelebbi pontokat tartottam meg. A fennmaradó pontfelhőben a távolságokat a $[0-1]$ tartományra normalizáltam.

A transzformált 3D pontfelhőket az I_1 és I_2 2D-s mélységképekre vetítettem, amint az az 6a és 6b ábrán látható. A Lidar vízszintes 360° látómezőjét 1024 pixelre, a levágott pontfelhő 5 m függőleges magasságát pedig 128 pixelre képeztem le, így az előállított mélységkép mérete 1024×128.

A tanító adatbázis 20000 pontfelhőpárt tartalmaz 50 helyszínről, míg a teszt adatbázis 2000 pontfelhőpárból áll, amelyek teljesen más mérési helyszínekről származnak.

Összefoglalva, létrehoztam egy új adathalmazt, amely alkalmas új változásfelismerő módszerek tanítására és kiértékelésére, olyan alkalmazások számára, ahol nem elvárható az összehasonlított pontfelhők pontos regisztrálása.

1.2. Altézis: Új, versengő osztályozó - diszkriminátor alapú tanítási módszert javasoltam a változásdetekciós feladatra pontatlanul regisztrált 3D pontfelhőpárokon.



7. ábra. Az *ChangeGAN* architektúra javasolt versengő tanító stratégiája.

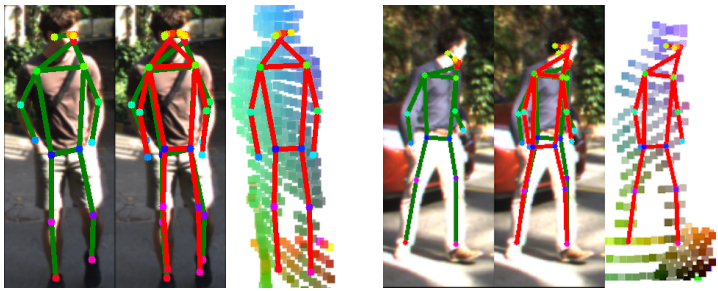
Az *osztályozó* hálózat felelős a mélységkép-párok közötti változások megtanulásáért és a köztük levő különbség felismeréséért. Az osztályozó háló aktuális állapota állítja elő a validációs adatokat, amelyeket a diszkriminátor modell dolgoz fel.

A *diszkriminátor* hálózat egy konvolúciós hálózat, amely az osztályozó hálózat kimenetét osztályozza. A diszkriminátor modell a bemeneti képet foltokra osztja, és minden egyes folt esetében eldönti, hogy az azon jelzett különbség valódi vagy hamis. A tanítás során a diszkriminátor háló arra kényszeríti az osztályozó modellt, hogy egyre jobb és jobb változás-becsléseket hozzon létre.

A 7. ábrán látható a javasolt versengő tanító stratégia. Az L1 Loss (L_{L1}) értékét a generált kép és a tanító kép közötti átlagos abszolút hibaként számolom ki, és definiálom a Adversarial költségfüggvényt (L_{Adv}), amely a diszkriminátor által generált leírók és egy egyesekből álló tömb szigmoid keresztentrópiája. A módszer végső költségfüggvénye (L) az Adversarial Loss és az L1 Loss súlyozott kombinációja: $L = L_{\text{Adv}} + \lambda * L_{L1}$.

2. Tézis: Új módszert javasoltam valós idejű előtér-háttér szegmentálásra és emberi póz becslésére, amely kizárólag egy nem ismétlődő körkörösen pásztázó Lidar szenzor pontfelhőit használja fel a becsléshez.

A javasolt módszer a ViTPose architektúrán alapul [35], amely egy transzformátor alapú [36, 37], optikai kameraképeket felhasználó emberi pózbecslési módszer.



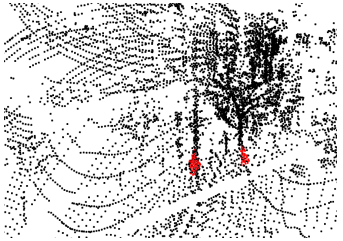
(a) **(b)**

8. ábra. *LidPose* pirossal jelölve a becslések, a bemeneti Lidar-pontfelhőn megjelenítve (*jobb oldalt*). A kapcsolódó kamerakép, és zölddel a tanító adatok (*bal oldalt*). Az becslés és a tanító adat együtt megjelenítve a Lidar pontfelhőn (*középen*).

A javasolt módszer a ViTPose továbbfejlesztése, amely 3D pontfelhőket használ, és hatékonyan képes kezelni a pontfelhők ritka jellegét és az NRCS Lidarok szokatlan, rozetta-szerű pásztázási mintázatát. A javasolt módszer [2] első lépése előtér-háttér szegmentáció [8] az NRCS Lidar pontfelhőjén, az előtérpontok kiválasztására. A következő lépésben a *LidPose* pózbecslő hálózat megbecsüli az emberi pózt az NRCS Lidar pontfelhő előtér részein. A kifejlesztett eljárás egy teljes megoldást ad az emberi póz becslésére statikus szenzor által rögzített NRCS Lidar mérési szekvenciákból. A módszer értékeléséhez létrehoztam egy új, valós, multimodális adathalmazt, amely egy Livox Avia Lidar pontfelhőit, egy kamera képeit és az azokhoz tartozó 2D és 3D emberi pózok annotált tanító adatait tartalmazza.

A módszer egy folyóiratban [2] és egy konferencia kiadványban [8] került publikálásra. A 8. ábrán példák láthatóak a javasolt *LidPose* módszer becsléseire.

2.1. Altézis: Javasoltam egy pontszintű előtér-háttér szegmentálási módszert nem ismétlődő körkörös szkennelésű Lidarral, statikus szenzorkonfigurációban rögzített pontfelhőkre. Bebizonyítottam, hogy a javasolt módszer képes kezelni az NRCS Lidar mérések inhomogenitását és ritkaságát. A javasolt megközelítés teszteléséhez és kiértékeléséhez létrehoztam egy adatbázist, és bizonyítottam annak hatékonyságát [8].



(a) Előtér pontok pirossal megjelítve egy NRCS Lidar pontfelhőn

(b) Előtér pontok pirossal megjelítve a nagyfelbontású háttérmodell pontfelhőjével együtt

9. ábra. Detektált előtér pontok a *Városi* adatbázisból származó NRCS Lidar adatokon.

A pontonkénti előtér-háttér szegmentálási feladat megoldásához egyensúlyt kell teremteni az NRCS Lidar adatok térbeli és időbeli felbontása között, melyekre példa a 3. ábrán látható. A térbeli felbontás növelése érdekében a szenzor látómezőjének nagyfelbontású háttérmodelljét készítem el, és frissítem naprakészen egy Gauss eloszlások keverékén alapuló valószínűségi módszer [8] segítségével. Az így szintetizált háttérmodellre egy példa a 9b ábrán látható. Másrészt, a dinamikus objektumok valós idejű elemzésének lehetővé tétele érdekében alacsony integrálási időt használok az egymást követő Lidar-pontfelhők létrehozásához. Ennek eredményeként az előtérben lévő objektumokról ritka, de geometriailag pontos pontfelhők készülnek, melyre példa a 9a ábrán látható. Az így nyert pontszintű szegmentációs eredmények bemenetként szolgálhatnak további, magasabb szintű felismerési lépéseknek, mint az alakfelismerés, objektumdetektció és póz becslés, ami a 2.3. Altézisben bemutatásra kerül. Az új megközelítés hatékonyságát különböző valós NRCS Lidar mérési szekvenciákon mutattam be.

2.2. Altézis: Félautomata módszert javasoltam emberi pózok adatbázisának létrehozására kameraképek és NRCS Lidar mérések felhasználásával.

A Lidar-pontfelhők annotálása kihívást jelentő feladat, mivel a ritka 3D pontfelhők vizuális értelmezése nem magától értetődő. Az általunk alkalmazott NRCS szenzor inhomogén mintázata miatt különösen nehezen vizsgálhatók a Lidar pontfelhők, ami tovább nehezíti az annotációs feladatot. Ezért a tanító adatok rögzítése során az NRCS Lidar-érzékelő platformjára egy kamera is felszerelésre került. Az együtt rögzített Lidar pontfelhő és kameraképet a 10. ábrán látható. A kameraképeket csak az emberi póz becsléséhez szükséges referencia adatok létrehozására, valamint a *LidPose* becsléseinek vizuális kiértékeléséhez használtam. Az annotálás során a kameraképek a csontváz csompópontok pozícióinak jelölésére és ellenőrzésére voltak használva.

10. ábra. NRCS Lidar pontfelhő 100 ms integrálási idővel, 2D mélységképként ábrázolva a hozzá tartozó kameraképen megjelenítve.

Mivel a kísérleti konfiguráció mind a kamera-, mind a Lidar-adatokat felhasználja a tanító adatok létrehozásához és az eredmények validálásához, a két szenzor koordináta-rendszere közötti térbeli transzformációs paramétereket kalibrációs eljárással határoztam meg. A kamera külső- és belső paramétereit az OpenCV könyvtárak és egy Livox-specifikus, kalibrációs tábla nélküli kalibrációs módszer [38] segítségével határoztam meg. Ezt követően a kameraképeket és a Lidar pontfelhőkből készített mélységképeket közös koordináta-rendszerbe transzformáltam. A kamera- és a Lidar-adatokat időbélyeggel láttam el a *Precision Time Protocol daemon (PTPd)* [39] használatával.

A kamera és a Lidar adatait különböző, szenzorspecifikus adatgyűjtési sebességgel rögzítettem, a kamera esetében 30 Hz-es, a Lidar esetében 10 Hz-es frekvenciával. Az adatgyűjtés sebességét a Lidar lassabb mintavételi sebességéhez igazítottam.

A referencia adatok generálását félautomata módon valósítottam meg, kihasználva a bevált kamera alapú emberdetekciós- és pózillesztési technikákat.

1. Minden egyes adatcsomag kameraképén lefuttattam a YOLOv8 [40] hálót a képen szereplő személyek detektálására.
2. A kezdeti pózbecslést a kivágott kameraképeken a legkorszerűbb 2D emberi pózbecslővel, a ViTPose [35] háló detekciójával hoztam létre.
3. A kameraképek segítségével kézzel ellenőriztem, validáltam, szűrtem és finomhangoltam az egyes becsült 2D emberi pózokat, így kaptam meg az emberi pózok 2D-s referencia adatait.
4. A szűrt kamera alapú 2D-s emberi pózok szolgálnak referenciaként a regisztrált Lidar mélységképén.
5. A 3D-s emberi pózok tanító adatait a 2D csontvázak kiterjesztésével hoztam létre, így minden egyes ízülethez mélységértéket próbáltam rendelni a Lidar érzékelőnek az ízület 2D pozíciója körüli mélységmérései alapján.
6. Tér- és idő interpolációt alkalmaztam a közvetlen távolságmérés nélküli ízületekre más közeli ízületek és közeli pontok mélységértékeiből.

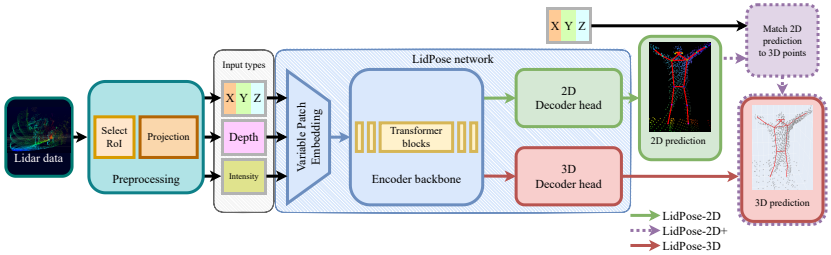
A létrehozott új adathalmaz összesen 9500 csontvázat és 161000 ízületet tartalmaz. Az adatbázist független tanító, validáló és teszt halmazra osztottam, amelyekben 5500, 490 és 3400 csontváz volt.

Összefoglalva, létrehoztam egy új adatbázist, amely alkalmas egy új emberi pózbecslési módszer létrehozására és kiértékelésére, amely csak NRCS Lidar pontfelhőt használ bemenetként. Az adatbázis használhatóságának bizonyítására egy vizuális transzformátor-alapú neurális hálózatot javasoltam az emberi póz becslésére, amelynek részleteit 2.3. Altézisben ismertetem.

2.3. Altézis: Új, vizuális transzformátor-alapú mély tanuló módszert javasoltam valós idejű emberi pózbecslésre inhomogén és ritka Lidar pontfelhőkből, amelyeket egy NRCS Lidar szenzor rögzít.

A publikált *LidPose* módszer [2] kizárólag NRCS Lidar mérések felhasználásával oldja meg az emberi póz becslési feladatot olyan környezetben, ahol a szenzor fixen rögzített pozícióba van felszerelve. A *LidPose* módszer komponensei a 11. ábrán láthatóak.

A javasolt módszer a legkorszerűbb kamera alapú emberi pózbecslési módszeren alapul (ViTPose [35]) és a Vision Transformer (ViT) architektúrából [37] indul ki, amelyet a COCO adathalmazon [41] tanítottak és teszteltek.



11. ábra. *LidPose* emberi pózbecslő ritka és inhomogén Lidar pontfelhőkre: Lidar-data: teljes Lidar-pontfelhő. Select ROI: kiválasztja az ember környezetében levő 3D pontokat. A 3D pontfelhő vetítéssel egy 2D-s tömbben. Bemeneti adatok: 3D XYZ koordináták (`XYZ`), mélység (`D`) és Lidar intenzitás (`I`). *LidPose* hálózat: Mind a *LidPose–2D*, mind a *LidPose–3D* a képszegmens feldolgozó modulokat és a kékkel jelölt kódoló vázat használja. A *LidPose–2D* és a *LidPose–3D* a megfelelő dekódoló modult használja, a *LidPose–2D+* pedig a 2D becslésből és a bemeneti pontfelhőből kerül kiszámításra.

Először a mozgó objektumok és előtér pontfelhő részek kerülnek elkülönítésre az NRCS Lidar pontfelhőn a statikus régióktól, amint azt a 2.1. Altézisben bemutattam.

A második lépésben az előtérpontok régióit szegmentálom az egyes mozgó objektumok elkülönítése érdekében, és megbecsülöm az észlelt emberek talppontját. Ennek a lépésnek az eredménye a személyek befoglaló doboza, amely mind a 2D, mind a 3D térben kiszámításra kerül.

A következő lépésben az NRCS Lidar pontfelhő és a mélységképe a fenti dobozon belül kivágásra kerülnek.

A különböző rendelkezésre álló mérési módozatok együttes reprezentálására egy új 2D adatstruktúrát javasoltam, amely a nyers Lidar-mérésekből egyszerűen összeállítható, és hatékonyan használható a javasolt *LidPose* modell tanítására és tesztelésére. A bemeneti pontfelhőből egy ötcsatornás kép készül a Lidar-érzékelő 2D-s mélységkép pixelrácsán, ahol két csatorna közvetlenül a Lidar-mérések mélység- és intenzitásértékeit, míg a fennmaradó három csatorna a Lidar-pontok `X,Y,Z` 3D koordinátáit tartalmazza.

A ViTPose [35] hálózatból kiindulva javasoltam a *LidPose* módszert a 3D-s pózbecslési feladat megoldásának. A javasolt *LidPose* módszer főbb ismérvei:

- A hálózat tulajdonság-kinyerő részében egy, a különböző bemeneti csatornaszámokhoz könnyen és dinamikusan adaptálódó modult

vezettem be.

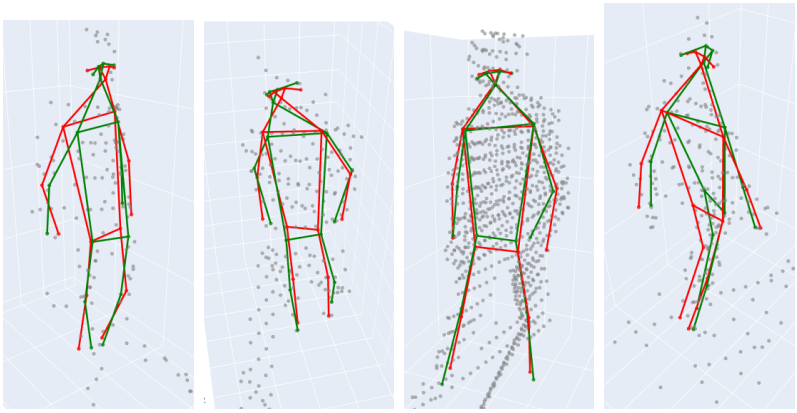
- A *LidPose* hálózat első részében használt transzformátorblokkok számát megnöveltem, hogy a hálózat általánosítási képességeit a több paraméter bevezetésével megnöveljem.
- A LidPose–3D konfiguráció kimeneti paramétereit is kibővítettem, hogy a 2D becslések mellett a csuklópontok mélységeit is képes legyen megbecsülni.

Amint a 11. ábrán mutatja, a *LidPose* hálózati struktúra különböző bemeneti és kimeneti konfigurációkat tud kezelni. Az optimális csatornakonfiguráció a módszer hiperparamétere, amely a [2]-ban leírt kísérleti kiértékelés alapján került kiválasztásra.

A *LidPose–3D* hálózat esetében a tanítás költségfüggvénye két komponensből áll: A csuklópontok 2D becslési pontosságáért az átlagos négyzetes hiba ($L_{\text{csukló2D}}$) felel, míg a másik komponens a mélységbecslés pontosságát számszerűsíti ($L_{\text{mélység}}$). A teljes költségfüggvény a pozíció- és mélység hibák súlyozott összege:

$$L_{\text{LidPose–3D}} = W_{\text{csukló2D}} \cdot L_{\text{csukló2D}} + W_{\text{mélység}} \cdot L_{\text{mélység}} \qquad (1)$$

A mélység hiba ($L_{\text{mélység}}$) három különböző módon került kiszámításra: *L1 loss*, *L2 loss* és *Structural Similarity Index Measure (SSIM)* [42]. A kiértékelések alapján és a számítási időt figyelembe véve a *SSIM*-t választottam a javasolt *LidPose–3D* hálózat mélység-hibájának mérésére.



(a) (b) (c) (d)

12. ábra. *LidPose3D* becsült csontvázak. Piros csontváz: 3D becslés. Zöld csontváz: tanító adat. Szürke pontok: NRCS Lidar pontok.

A módszer számszerű kiértékeléséhez több metrika került kiszámításra. A legjobb létrehozott modell a *helyesen becsült kulcspontok százalékos aránya görbe alatti terület* (Area Under the Percentage of Correct Keypoints Curve) metrikára $0,694$ értéket ért el. Az eljárás által becsült csontvázak átlagos távolsági hibáját is kiszámítottam, ahol a legjobb modell $0,158$ m értéket ért el. A javasolt módszerrel becsült 3D emberi pózokat néhány releváns mintaképen a 12. ábrán mutatja, lehetőséget nyújtva az eredmények vizuális kiértékelésére. A *LidPose* publikációmban [2] bemutatott eredmények megerősítik, hogy a javasolt módszer képes emberi csontvázak helyes detekciójára ritka és inhomogén NRCS Lidar pontfelhőkben.

A javasolt módszer pontosan és hatékonyan becsül 3D-s emberi pózt valós időben az NRCS Lidar pontfelhőkön. A becsült pózok megfigyelőrendszerek magasabb szintű feldolgozási lépéseiben (például cselekvésfelismerés, biometria) felhasználhatóak.

4. Az eredmények alkalmazása

4.1. ChangeGAN

A javasolt *ChangeGAN* [1], [3] képes robusztusan észlelni a változásokat a komplex utcai környezetben kapott ritka pontfelhők között. A javasolt módszer kulcsfontosságú jellemzője, hogy nem szükséges a hatékony működéséhez a pontfelhő párok pontos regisztrációja. Kísérleteim alapján a módszer hatékonyan képes kezelni a 3D pontfelhők közötti akár 1 m transzlációs és 10° forgatási eltéréseket. Ez teszi a javasolt módszert alkalmassá olyan valós alkalmazásokhoz, ahol a pontfelhők pontos regisztrálása a környezet összetettsége vagy a szenzorok korlátai miatt nem kivitelezhető. A módszer alkalmazható automatikus infrastruktúrafelügyeletben, ahol az esetlegesen veszélyes helyzetek, például a hiányzó közlekedési táblák és a sérült utcai tárgyak felismerése kulcsfontosságú. Használható a módszer nagyfelbontású 3D térképek hatékony frissítéséhez az autonóm járműveken. A városüzemeltető szolgáltatóknál költséges és időigényes erőfeszítések csökkenthetők a módszer alkalmazásával, amely nagy területekről rendszeresen érkező felvételek automatikus és folyamatos elemzésére és összehasonlítására szolgál a releváns környezeti változások megtalálása érdekében.

4.2. LidPose

A *LidPose* módszert leíró publikációban [2] bizonyítottam, hogy a Livox Avia [31] NRCS Lidar szenzor az alacsony ára miatt széles körben alkalmazható valós mérési helyzetekben, használható komplex emberi pózbecslési feladatok megoldására. A javasolt módszer előnye, hogy nagymértékben tiszteletben tartja a megfigyelt személyek magánéletét, mivel a rögzített ritka pontfelhőkből nem lehet azonosítani az embereket.

A változásérzékelés pontossága növelhető egy új mélységképkiegészítési technika alkalmazásával, amely kiküszöböli az NRCS Lidaradatok egyenetlen ritkaságát, amint azt az egyik általunk benyújtott szabadalmi bejelentés [4] leírja.

4.3. Publikációk és előadások

A kutatási eredményeim elsősorban rangos folyóiratokban és konferenciákon jelentek meg, a disszertációmban és a fentiekben hivatkozottak szerint. [1, 2] [3, 4] [5–8]

Ezeken felül a kutatási eredményeimet bemutattam a *Képfeldolgozók és Alakfelismerők Társasága (KÉPAF)* kétévente megrendezésre kerülő konferenciáin [9–11] és a *Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Karának PhD proceedings* évenkénti kiadványában [12–15].

Az eredményeimet bemutattam többek között a Kutatók éjszakáján[1], valamint a *Mesterséges Intelligencia Nemzeti Laboratórium (MILAB)* és *Autonóm Rendszerek Nemzeti Laboratórium (ARNL)* által szervezett rendezvényeken, többek között az *AI & Aut Expo*-n[2].

5. Köszönetnyilvánítás

Ez a disszertáció nem készülhetett volna el a körülöttem lévők támogatása nélkül. Szeretném kifejezni legmélyebb hálámat és elismerésemet PhD témavezetőmnek, **Benedek Csabának**. Éleslátó útmutatásai és visszajelzései arra ösztönöztek, hogy kritikusan gondolkodjak, feszegessem a tudásom határait, és munkámban kiválóságra törekedjek.
Őszinte köszönetemet fejezem ki a *Roska Tamás Műszaki és Természettudományi Doktori Iskola* jelenlegi és korábbi vezetőinek, **Szederkényi**

[1] https://sztaki.hun-ren.hu/kutatok-ejszakaja-2022#xr

[2] https://www.facebook.com/photo/?fbid=877979741876158 2&set=pcb.8780186178722706

Gábornak és **Szolgay Péternek**. Valamint a *Pázmány Péter Katolikus Egyetem (PPKE) Információs Technológiai és Bionikai Kar* dékánjainak, **Cserey Györgynek** és **Iván Kristófnak**. És nem szabad megfeledkeznem **Vida Tivadarné Katinka néniről**, aki minden adminisztratív kérdésemben és kihívásomban segítőkész és türelmes volt.
Továbbá mély elismerésemet fejezem ki **Szirányi Tamásnak** és a *Gépi Érzékelés Kutatlaboratóriumnak (MPLAB)* a *Számítástechnikai és Automatizálási Kutatóintézetben (HUN-REN SZTAKI)*, amiért minden kutatási eszközt, szenzort, számítógépet, szervert és az ösztönző kutatási környezetet biztosították számomra.
Szívből köszönöm **Nagy Balázs, Bódis Balázs, Kégl Marcell, H. Zováthi Örkény** társszerzőimnek a publikációimhoz való értékes hozzájárulásukat. Hálás vagyok kollégáimnak és barátaimnak a rendületlen támogatásukért.
Külön köszönet azoknak, akiket talán nem említettem itt név szerint, de akik közvetlenül vagy közvetve támogattak kutatásom megvalósításában.
Az anyagi támogatásért köszönet az Európai Uniónak melyet az Autonóm Rendszerek Nemzeti Laboratórium (RRF-2.3.1-21-2022-00002) és a Mesterséges Intelligencia Nemzeti Laboratórium (RRF-2.3.1-21-2022-00004) programok keretében nyújtott. További támogatást nyújtottak a TKP2021-NVA-27 és TKP2021-NVA-01 pályázatok, valamint a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal (NKFI Hivatal) OTKA #143274 projektje.
Szeretnék köszönetet mondani szüleimnek, nagypapámnak és testvéreimnek a feltétlen támogatásukért, és mindazért, amit értem tettek nem csak a PhD-tanulmányaim alatt, hanem azt megelőzően is.
Végül, de nem utolsósorban szeretnék köszönetet mondani a családomnak, és legfőképpen szeretett *Klári*mnak, aki biztos hátteret biztosított számomra, és egy olyan otthont, ahol mindig felfrissülhettem és kipihenhettem magam. Mindig támogattatok, függetlenül az akadályoktól, és megértettétek, hogy valóban el akarom érni ezt a célt. Végül hálás vagyok lányomnak *Bíbor*nak, és fiaimnak *Özsé*bnek és *Donát*nak, akik elfogadták, hogy a nehéz időszakokban többet dolgoztam, mint ők szerették volna. A szeretetük és a megértésük mindig ott volt, és ők voltak azok, akik a legsötétebb pillanatokban is mosolyra fakasztottak és megnevettettek.

6. Irodalomjegyzék

Folyóirat publikációk

[1] B. Nagy, **L. Kovács**, and C. Benedek, „ChangeGAN: A deep network for change detection in coarsely registered point clouds," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8277–8284, 2021, **IF: 4.3, Scimago Q1/D1**. (Hivatkozva: 7, 19, and 20. oldal)

[2] **L. Kovács**, B. M. Bódis, and C. Benedek, „LidPose: Real-time 3d human pose estimation in sparse lidar point clouds with non-repetitive circular scanning pattern," *Sensors*, vol. 24, no. 11, 2024, **IF: 3.9, Scimago Q1**. (Hivatkozva: 13, 16, 18, 19, and 20. oldal)

Szabadalmak

[3] B. Nagy, **L. Kovács**, C. Benedek, T. Szirányi, Ö. Zováthi, and L. Tizedes, „Training method for training a change detection system, training set generating method therefor, and change detection system," WO Patent application, WO/2023/007198, International Filing Date: 08.07.2022, Priority data: P2100280, 27.07.2021, HU. (Hivatkozva: 7, 19, and 20. oldal)

[4] Ö. Zováthi, Z. Rózsa, B. Pálffy, Z. Jankó, C. Benedek, T. Szirányi, **L. Kovács**, and M. Kégl, „Methods for spatial and temporal densification of Lidar measurements," Patent application, Priority data: P2300075, 01.03.2023, HU. (Hivatkozva: 20. oldal)

Konferencia publikációk

[5] Ö. Zováthi, **L. Kovács**, B. Nagy, and C. Benedek, „Multi-object detection in urban scenes utilizing 3d background maps and tracking," in *2019 International Conference on Control, Artificial Intelligence, Robotics and Optimization (ICCAIRO)*, 2019, pp. 231–236. (Hivatkozva: 7 and 20. oldal)

[6] A. Horvath, I. Horvath, A. Kiss, D. Huszar, A. Palffy, **L. Kovács**, D. Babicz, B. Farkas, G. Majoros, and C. Rekeczky, „Cellular vision based adas applications," in *CNNA 2016; 15th International Workshop on Cellular Nanoscale Networks and their Applications*, 2016. (Hivatkozva: 8 and 20. oldal)

[7] **L. Kovács**, L. Lindenmaier, H. Nemeth, V. Tihanyi, and A. Zarandy, „Performance evaluation of a track to track sensor fusion algorithm,"

in *CNNA 2018; The 16th International Workshop on Cellular Nanoscale Networks and their Applications*, 2018. (Hivatkozva: 8 and 20. oldal)

[8] **L. Kovács**, M. Kégl, and C. Benedek, „Real-time foreground segmentation for surveillance applications in nrcs lidar sequences," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B1-2022, pp. 45–51, 2022. (Hivatkozva: 13, 14, and 20. oldal)

További publikációk

[9] **L. Kovács**, B. M. Bódis, and C. Benedek, „LidPose: Real-time 3d human pose estimation in sparse lidar point clouds with non-repetitive circular scanning pattern," in *15th Conference of the Hungarian Association for Image Analysis and Pattern Recognition*, Hévíz, 2025. (Hivatkozva: 20. oldal)

[10] **L. Kovács**, M. Kégl, and C. Benedek, „Real-time foreground segmentation for surveillance applications in sequences from a non-repetitive circular scanning lidar," in *14th Conference of the Hungarian Association for Image Analysis and Pattern Recognition*, Gyula, 2023. (Hivatkozva: 20. oldal)

[11] **L. Kovács**, B. Nagy, and C. Benedek, „Demonstration of changegan: change detection in unregistered point clouds using neural networks," in *13th Conference of the Hungarian Association for Image Analysis and Pattern Recognition*, Budapest, 2021. (Hivatkozva: 20. oldal)

[12] **L. Kovács**, „Change detection in unregistered 3d point clouds," in *PhD proceedings, annual issues of the Doctoral School, Faculty of Information Technology and Bionics*, vol. 16, 2021, p. 67. (Hivatkozva: 20. oldal)

[13] **L. Kovács**, „Change detection in lidar point clouds," in *PhD proceedings, annual issues of the Doctoral School, Faculty of Information Technology and Bionics*, vol. 15, 2020, p. 68. (Hivatkozva: 20. oldal)

[14] **L. Kovács**, „Challenges in track to track sensor fusion using neural networks," in *PhD proceedings, annual issues of the Doctoral School, Faculty of Information Technology and Bionics*, vol. 14, 2019, p. 60. (Hivatkozva: 20. oldal)

[15] **L. Kovács**, „Challenges in sensor fusion," in *PhD proceedings, annual issues of the Doctoral School, Faculty of Information Technology and Bionics*, vol. 13, 2018, p. 55. (Hivatkozva: 20. oldal)

Hivatkozások

[16] C. Benedek, B. Gálai, B. Nagy, and Z. Jankó, „Lidar-based gait analysis and activity recognition in a 4d surveillance system," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 28, no. 1, pp. 101–113, 2018. (Hivatkozva: 3. oldal)

[17] F. Oberti, L. Marcenaro, and C. S. Regazzoni, „Real-time change detection methods for video-surveillance systems with mobile camera," in *European Signal Processing Conference*, 2002, pp. 1–4. (Hivatkozva: 3. oldal)

[18] C. Benedek, X. Descombes, and J. Zerubia, „Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 33–50, 2012. (Hivatkozva: 3. oldal)

[19] S. Ji, Y. Shen, M. Lu, and Y. Zhang, „Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples," *Remote Sensing*, vol. 11, no. 11, 2019. (Hivatkozva: 3. oldal)

[20] C. Zimmermann, T. Welschehold, C. Dornhege, W. Burgard, and T. Brox, „3d human pose estimation in rgbd images for robotic task learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 1986–1992. (Hivatkozva: 3. oldal)

[21] M. Cormier, A. Clepe, A. Specker, and J. Beyerer, „Where are we with human pose estimation in real-world surveillance?" in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 2022, pp. 591–601. (Hivatkozva: 3. oldal)

[22] W. Hu, T. Tan, L. Wang, and S. Maybank, „A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 3, pp. 334–352, 2004. (Hivatkozva: 3. oldal)

[23] A. Zanfir, M. Zanfir, A. Gorban, J. Ji, Y. Zhou, D. Anguelov, and C. Sminchisescu, „Hum3dil: Semi-supervised multi-modal 3d human

pose estimation for autonomous driving," in *Proceedings of The 6th Conference on Robot Learning*, vol. 205, 2022, pp. 1114–1124. (Hivatkozva: 3. oldal)

[24] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, „Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 01, pp. 172–186, 2021. (Hivatkozva: 3. oldal)

[25] H.-S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y.-L. Li, and C. Lu, „Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 6, pp. 7157–7173, 2023. (Hivatkozva: 3. oldal)

[26] P. Lu, T. Jiang, Y. Li, X. Li, K. Chen, and W. Yang, „RTMO: Towards high-performance one-stage real-time multi-person pose estimation," 2023. (Hivatkozva: 3. oldal)

[27] Y. Li and J. Ibanez-Guzman, „Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 50–61, 2020. (Hivatkozva: 4. oldal)

[28] H. W. Yoo, N. Druml, D. Brunner, C. Schwarzl, T. Thurner, M. Hennecke, and G. Schitter, „Mems-based lidar for autonomous driving," *e & i Elektrotechnik und Informationstechnik*, vol. 135, no. 6, pp. 408–415, Oct 2018. (Hivatkozva: 4. oldal)

[29] F. Amzajerdian, V. E. Roback, A. Bulyshev, P. F. Brewster, and G. D. Hines, „Imaging flash lidar for autonomous safe landing and spacecraft proximity operation," 2016. (Hivatkozva: 5. oldal)

[30] A. Palffy, E. Pool, S. Baratam, J. Kooij, and D. Gavrila, „Multi-class road user detection with 3+1d radar in the view-of-delft dataset," *IEEE Robotics and Automation Letters*, pp. 1–1, 2022. (Hivatkozva: 5. oldal)

[31] „Livox avia specifications," https://www.livoxtech.com/avia/specs, accessed: 2024-03-11. (Hivatkozva: 6 and 20. oldal)

[32] J. Bromley, J. Bentz, L. Bottou, I. Guyon, Y. Lecun, C. Moore, E. Sackinger, and R. Shah, „Signature verification using a "siamese" time delay neural network," *International Journal of Pattern*

Recognition and Artificial Intelligence, vol. 7, p. 25, 1993. (Hivatkozva: 9. oldal)

[33] O. Ronneberger, P. Fischer, and T. Brox, „U-net: Convolutional networks for biomedical image segmentation," in *Int. Conf. Medical Image Computing and Comp.-Ass. Intervention*, 2015, pp. 234–241. (Hivatkozva: 10. oldal)

[34] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, „Spatial transformer networks," *Advances in Neural Information Processing Systems (NIPS)*, 2015. (Hivatkozva: 10. oldal)

[35] Y. Xu, J. Zhang, Q. Zhang, and D. Tao, „Vitpose: Simple vision transformer baselines for human pose estimation," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 38 571–38 584. (Hivatkozva: 13, 16, and 17. oldal)

[36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, „Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17, 2017, p. 6000–6010. (Hivatkozva: 13. oldal)

[37] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, „An image is worth 16x16 words: Transformers for image recognition at scale," 2021. (Hivatkozva: 13 and 16. oldal)

[38] C. Yuan, X. Liu, X. Hong, and F. Zhang, „Pixel-level extrinsic self calibration of high resolution lidar and camera in targetless environments," *CoRR*, 2021. (Hivatkozva: 15. oldal)

[39] K. Lao and G. Yan, „Implementation and analysis of ieee 1588 ptp daemon based on embedded system," in *2020 39th Chinese Control Conference (CCC)*, 2020, pp. 4377–4382. (Hivatkozva: 15. oldal)

[40] G. Jocher, A. Chaurasia, and J. Qiu, „Ultralytics YOLOv8," https://github.com/ultralytics/ultralytics, 2023. (Hivatkozva: 16. oldal)

[41] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, „Microsoft coco: Common objects in context," in *Computer Vision – ECCV 2014*, 2014, pp. 740–755. (Hivatkozva: 16. oldal)

[42] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, „Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. (Hivatkozva: 18. oldal)