

Óbudai Egyetem

Doktori (PhD) értekezés
tézisfüzete



**Többirányú, rögzített felbontású képi vetületek
objektumok felismerésére és párosítására**

Kertész Gábor

Témavezetők:

Dr. Vámosy Zoltán

Dr. habil. Szénási Sándor

**Alkalmazott Informatikai és Alkalmazott
Matematikai Doktori Iskola**

Budapest, 2019.

1 A kutatás előzményei

A gépi látás és képfeldolgozás használata a forgalomelemzés és -irányítás területén az elmúlt években jelentős növekedésnek indult [Rinner és Wolf, 2008]. A hagyományos megoldások (például forgalomszámlálás, balesetészlelés) mellett a járművek azonosításán és követésén alapuló több-kamerás alkalmazások is megjelentek [Sanchez és tsai., 2011].

Egy gépjármű azonosítása nem minden esetben lehetséges annak rendszáma alapján. Időjárási tényezők mellett a természetes megvilágítás hiánya is okozhat rossz látási viszonyokat. Kihívást jelenthet például a távoli azonosító leolvasása mesterseges megvilágítással rendelkező alagutakban elhelyezett kamerák gyenge minőségű képén, amely kevés színi információ mellett alacsony felbontású és zajos lehet.

A feladat a gépi látás területén belül az objektum-felismerés problémakörébe tartozik. Objektum detektálás a feladat, amennyiben az ismert megjelenésű objektum képi azonosítására van szükség. Objektum klasszifikációként definiálják a szegmált ismeretlen objektum típusának meghatározását.

Amennyiben a feladat egy korábban látott objektum ismételt felismerése, újraazonosítása, akkor a feladatra példány felismerésként, vagy objektum párosításként hivatkoznak [Szeliski, 2010].

Az elmúlt évtizedek során többféle irányból is megközelítették az objektumok párosításának problémáját. Az irodalomban rengeteg példát találunk kulcspontokon alapuló jellemző leírók használatával történő megoldásokra [Shi és Tomasi, 1993]; [Lindeberg, 1998]; [Schmid, Mohr és Bauckhage, 2000]; [Lowe és tsai., 1999]; [Bay és tsai., 2008], kifejezetten gépjárművek felismerése esetében is [Guo, Hsu és tsai., 2007]; [Guo, Rao és tsai., 2008]; [Hou, Wang és Qin, 2009]; [Choe, M. W. Lee és Haering, 2010].

Amennyiben a kép gyenge minőségű, és alacsony felbontású, nem célszerű magas szintű, kulcspontokon alapuló megoldásokkal próbálkozni. Ebben az esetben pixel-intenzitásokon alapuló megoldások jöhetnek szóba, például a mintaillesztés [Brunelli, 2009], a Haar-szerű jellemzők [Oren és tsai., 1997]; [Papageorgiou, Oren és Poggio, 1998]; [Viola és Jones, 2001a]; [Viola és Jones, 2001b]; [Yuan és Sclaroff, 2009]; [Rios-Cabrera, Tuytelaars és Van Gool, 2012], vagy a vetületi jellemzők használata [Betke, Haritaoglu és Davis, 2000]; [Liu, Collins és Tsin, 2002]; [S. Lee, Liu és Collins, 2007]; [Shan, Sawhney és Kumar, 2005].

[Jelača és tsai., 2013] publikációjukban egy összetett modellt mutattak be, amely segítségével alacsony hibarárával voltak képesek gépjárművek képi azonosítására egymástól távol elhelyezett, nem-átfedő nézetű kamerák képe alapján. A módszer alapja egy négy-dimenziós vetületi szignatúra, amely vízszintes, függőleges, átlós és mellék-átlós vetületekből áll. A vetületi szignatúrák összehasonlíthatóak, értékeik alapján kiszámítható az objektumok hasonlósági mértéke.

A vetületi szignatúra dimenziószáma növelhető több vetületi irány bevezetésével. Több-irányú képi vetületekre való leképezésre alkalmazható a Radon-transzformáció [Radon, 1917], amely elsősorban inverz formulája miatt vált ismertté [Deans, 1983]. Kapcsolódó leképezés a Trace transzformáció [Kadyrov és Petrou, 2001], amely a Radon egy általánosított formáját adja. A Hough transzformáció [Hough, 1962] egy

vonalak és alakzatok detektálására szolgáló módszer, amely a Radonhoz hasonlóan forgatáson alapuló leképezést alkalmaz [Deans, 1983]; [Ginkel, Hendriks és Vliet, 2004].

Az elmúlt években robbanásszerű mértékben növekedett a képi klasszifikáció hatékonysága [Krizhevsky, Sutskever és G. E. Hinton, 2012], amely a gépi tanulás, azon belül is a mélytanulás [Bengio, 2012a]; [Y. LeCun, Bengio és G. Hinton, 2015]; [Schmidhuber, 2015]; [Goodfellow és tsai., 2016] megjelenésének eredménye.

A modern képi objektumfelismerés konvolúciós neurális hálózatokon¹ alapul [Y. LeCun, Boser és tsai., 1989]; [Y. LeCun, Bengio és tsai., 1995], amely nagy neuron és rétegszám esetén kiemelkedő hatékonyságra képes [Y. A. LeCun és tsai., 2012]; [Bengio, 2012b]; [Szegedy és tsai., 2015]; [Karpathy, Li és Johnson, 2017]. A nagy elemszám és az ezzel járó nagyszámú tanítóminta feldolgozása komoly számítási kapacitást igényel. A *deep learning* módszerek a párhuzamosíthatóságuknak köszönhetően váltak elterjedtté, grafikus processzorok segítségével a modellek betanításához szükséges idő a töredékére csökkenthető [Raina, Madhavan és Ng, 2009].

Képi objektumok párosítására konvolúciós neurális hálózatokon alapuló architektúrát alkalmazott [Bromley és tsai., 1994], amelyet Sziámi neurális hálózatnak² neveztek el. A módszert sikeresen alkalmazták kézi aláírások hitelességének vizsgálatára.

Hatékonyan alkalmaznak sziámi architektúrájú megoldásokat arcfelismerésen alapuló azonosításra [Chopra, Hadsell és Y. LeCun, 2005]; [Taigman, Yaniv and Yang, Ming and Ranzato, Marc'Aurelio and Wolf, Lior, 2014]; [Schroff, Kalenichenko és Philbin, 2015], de további *one-shot learning*, azaz kevés tanítómintán alapuló problémák esetén is alkalmazhatóak [Oriol Vinyals and Charles Blundell and Timothy P. Lillicrap and Koray Kavukcuoglu and Daan Wierstra, 2016]; [Koch, Zemel és Salakhutdinov, 2015].

2 Célkitűzések

A kutatásom elsődleges célja a többirányú vetületi leírók alkalmazhatóságának vizsgálata objektumok párosítására. A négy-dimenziós vetületi szignatúrák jól alkalmazhatóak objektum párosításra, feltételezhető, hogy több vetületi irány bevezetésével a módszer pontossága javítható.

Radon transzformáció esetén az egyes vetületek hossza függ a bemeneti kép méretétől és a leképezés irányától. Amennyiben a hasonlósági mutató a különböző hosszúságú vetületek legjobb illeszkedése alapján kerül meghatározásra, akkor különböző megfigyelések esetén is tévesen magas hasonlóságot jelezhet.

Céлом egy olyan többirányú képi vetületi módszer kidolgozása, amely vetületi szeletek hossza fix méretű, előre meghatározható. Ennek eredményeképp a forgatási szögtől és egyben a kép méretétől független leképezés jöjjön létre.

A módszer párhuzamosítási lehetőségeinek vizsgálata is fontos feladat, hiszen a modern számítógép architektúrák nagy mértékű támogatást nyújtanak adatpárhuzamos végrehajtásra, feladattól függően a valós idejű feldolgozásra. Jellemzően többdimenziós tömbök, mátrixok feldolgozása esetén alkalmazhatóak a grafikus kártyák multiprocesszorai is a párhuzamos végrehajtásra.

¹Convolutional Neural Network, CNN

²Siamese Neural Network, SNN

További célom a kidolgozott vetületi módszert képi egyezésvizsgálatra alkalmazva teljesítmény, számítási komplexitás és memóriaköltség szempontok szerint összehasonlítani a további vetületi módszerekkel.

A kutatás folytatásaként elemzem a gépi tanulás alapú módszereket objektumok párosítására, és megvizsgálom a többirányú vetületi leírókon alapuló szíami modellek alkalmazhatóságát. A különböző vetületi irányok jelentősége objektum párosítás esetén probléma típusától és kamerabeállításaitól függően más és más lehet. Feltételezhető továbbá, hogy a paramétertér különböző jellemzői alapján is elvégezhető a hasonlóságvizsgálat.

A vetületi jellemzőkön alapuló neurális hálózatok hatékonyságának megállapítására egy átfogó kísérletet kell elvégezni. Ehhez szükséges a különböző típusú és méretű bemenetekre neurális architektúrákat generáló módszert kidolgozni.

Ezt követően a nagyszámú neurális modell tanítása és kiértékelése egy nagy számításigényű probléma, amely megoldása célszerű lehet elosztott környezetben. Jelen probléma több, egymástól független modell tanításának olyan megtervezése, amely felépítés és ütemezés kialakításával a párhuzamos hatékonyságot maximalizálja.

Célom tehát egy összetett kísérlet megtervezése és kivitelezése többirányú képi vetületek alapján történő objektumpárosításra neurális hálózatok alkalmazásával. Ennek részeként megoldást keresek neurális hálózati architektúrák generálására, a modellek elosztott tanítására, valamint kiértékelésére.

3 Vizsgálati módszerek

3.1. Több-irányú képi vetületek

A vetületi szakasz hosszát a bemeneti kép mérete és a vetületi szög határozza meg. Trigonometriai függvények alapján adott α szög szerinti vetület hossza $\cos(\alpha)N + \sin(\alpha)N$, ahol N jelöli a négyzetes kép oldalhosszát. Ezen szakaszt S darab egyenlő hosszúságú rekeszre bontva α szögtől és N hosszától független felbontás adható.

A rekeszek meghatározását követően az egyes vizsgált pixelek értéke a fedett rekeszek értékéhez adódik, a fedettség arányában. Külön esetként célszerű kezelni amennyiben az érintett rekeszek száma 1, 2, vagy több. Pontosan egy érintett rekesz esetén az intenzitásérték teljes egészében a rekeszhez adódik. Pontosan két rekesz esetén a fedettség arányában történik a rekeszek értékének növelése. Több érintett rekesz esetén a teljesen lefedettek a teljes értékkel, a részben fedettek a vetület arányában kerülnek növelésre. Az érintett elemszámokkal történő normalizáció követően a skála egységessé válik.

Adatpárhuzamos implementáció esetén szükséges figyelembe venni a multiprocesszoros környezet architektúrais sajátosságait. GPGPU³ programozás esetén fontos szempont a gazdagép és a grafikus gyorsító memóriája közötti átvitelek minimalizálása, az átviteli idő *overhead*jének csökkentése miatt.

Hasonlóképp fontos a GPU⁴ eszköz memória architektúrájának széleskörű kihasználása, lévén a processzorok hatékonyabban érik el a helyi *on-chip* memóriát mint a nagyobb kapacitású globális memóriát.

³General-purpose computing on graphics processing units

⁴Graphics processing units

A multiprocesszorok által futtatott különböző szálak közötti közös változónak felszámolása fontos tervezési lépés: közös változók használata a versenyhelyzet miatt számítási hibát okozhat. A versenyhelyzet klasszikusan szinkronizációval történő kezelése a hatékonyság csökkenésével jár, így az algoritmus átalakítása jobb tervezési irány. Az eredmények felhalmozásakor redukció, vagy atomi műveletek is alkalmazhatóak.

Vetületi függvények hasonlósági tényezőjének megállapítására a Pearson korrelációs koefficiens kerül kiszámításra. A kapott korrelációs értékek alapján a vetületi szignatúra dimenziószámával normalizált euklideszi norma adja a hasonlósági mutatót.

A klasszifikáció hatékonyságának meghatározására alkalmazható olyan küszöbölési technika, ahol az adathalmazban szereplő valós párokra számított hasonlósági értékek pontosan 50 vagy 80 százaléka túlmutat. Ilyen határértékek számítása után vizsgálandó, hogy hamis párok milyen arányban találhatók meg a küszöbérték felett, tehát fals pozitívként.

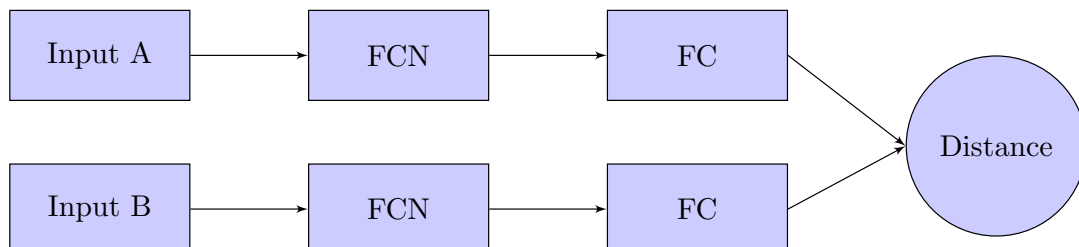
3.2. Sziámi Konvolúciós Neurális Hálózatok

A Sziámi architektúrájú konvolúciós neurális hálózatok két azonos konvolúciós fejből állnak, amelyekben az architektúra mellett a paraméterek súly és eltolás értékei is megegyeznek (Ábra 3.1).

Konvolúciós neurális hálózatok jól ismert rétegzett architektúrájuk, konvolúciós rétegek használatával végzik a jellemzők felismerését, majd ezt követően esetenként pooling rétegek csökkentik a reprezentáció méretét. A kimeneti réteg előtt gyakran alkalmaznak hagyományos, teljesen összekötött rétegeket is. A hálózatok ilyen felépítésében többféle tervezési sajátosság is megfigyelhető a kernel- vagy ablakmérettől a szűrők számán át a különféle hiperparaméterekig.

A konvolúciós hálózati architektúrák generálására kidolgozott módszer három lépésen alapul:

1. a maximális kernel- és ablakméretek meghatározása a bemeneti mátrix mérete és a konvolúciós rétegek száma alapján;
2. a konvolúciós és pooling rétegpárok generálása ügyelve a valid reprezentációra, a maximális memóriaköltség betartása mellett;



3.1. ábra. A "két-fejű" Sziámi Neurális Hálózat architektúrális felépítése. A teljesen konvolúciós (fully-convolutional, FCN) rétegeket teljesen összekötött (fully connected, FC) rétege követik. A fejekben található súlyparaméterek megegyeznek, kimeneteik több-dimenziójú vektorok. A kimeneti vektorok között számolt távolság adja a bemenetek hasonlóságát.

3. adott számú architektúra gyűjtése és a tanítóminták kötegméretének optimalizálása a maximális memóriakihasználtság érdekében.

A generált modellek elosztott környezetben történő tanítására alkalmas felépítés a *Master/Worker* tervezési minta, hiszen az egyes feldolgozások függetlenek egymástól, csak a bemeneti tanítóminták közösek.

A minta jellemzője, hogy a terhelés elosztás automatikus, a *Worker* egységek kérésre feladatot kapnak a várósból. A párhuzamos hatékonyság növelhető ütemezéssel, például a feladatok végrehajtási idejeinek csökkenő sorrendben történő kiosztásán alapuló heurisztikával. A végrehajtási idők előre nem ismertek, azonban az architektúrák paraméterei alapján becsülhetőek.

A modellek párosításának teljesítménye mérhető a *one-shot* klasszifikáció használatával, amely a mért szemantikus távolság értékek vizsgálata helyett a párosítási precizitást veszi figyelembe. Előbbi esetben a Sziámi hálózat kimeneteként megjelenő jellemzővektorok közötti távolság küszöbölésével történik az azonosság prediktálása. Ezzel szemben *one-shot* klasszifikáció esetén a referenciaobjektum több, különböző példányokhoz tartozó objektumokkal kerül összehasonlításra, ahogy valós alkalmazás esetén is történne. A klasszifikáció helyes, amennyiben a referencia és a valós párja közötti távolság minimális a további objektumokhoz képest.

A modellek összehasonlíthatóak továbbá memóriaköltség alapján is, amely jól közelíthető a modellben lévő súly és eltolás paraméterek számával. Többcélú optimalizációval megadhatóak a klasszifikáció pontossága és modellek memóriaköltsége alapján optimális modellek.

Amennyiben egy m_1 modell mért predikciós teljesítménye és becsült memória használata szempontjából is jobb m_2 modellnél, akkor m_1 Pareto dominálja m_2 modellt. A Pareto-front elemei azok a modellek, amelyeket nem dominál másik modell, azaz Pareto-optimálisak a precizitás és a memóriaköltség szempontjából.

4 Új tudományos eredmények

I. téziscsoport: Többirányú képi vetületek terén elért eredményeim

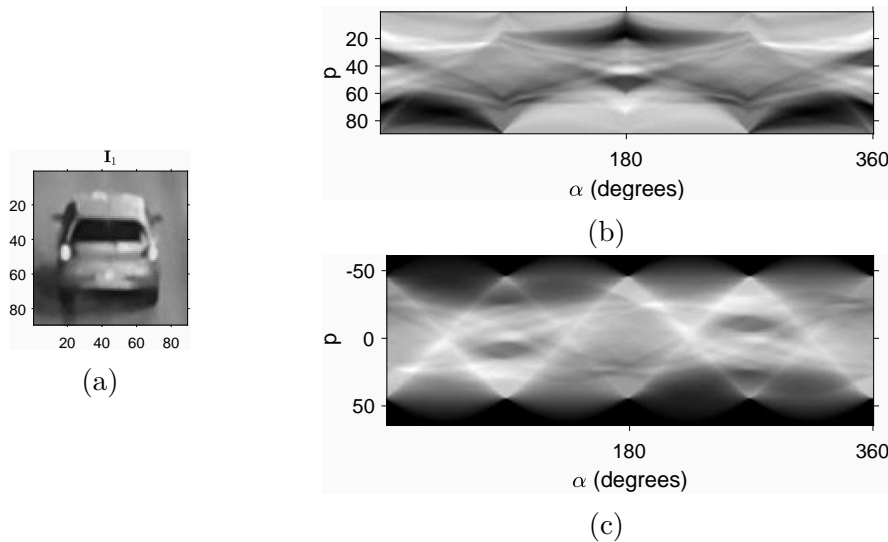
1.1 tézis

Megterveztem és implementáltam egy többirányú, fix rekeszszámot használó, forgatási szögtől függetlenül azonos hosszúságú vetületi vektorokra leképező módszert. Az eredmény memóriagigénye a kép méretétől független, csak a forgatás lépésszáma és a rekeszszám befolyásolja.

A trigonometriai függvényeken alapuló módszer számítási komplexitása szekvenciális feldolgozás esetén megközelítőleg azonos a korábbi módszerekével, a vetületi irányok számának függvényében meghatározható.

A vetületi szeletek hossza különböző leképezési irányok esetén is fix, ennél fogva a memóriaköltség előre ismert. Alacsony rekeszszám esetén a módszer tömörítésként is értelmezhető, a Radon transzformációnál kisebb memóriaköltségű leképezést ad. A módszer által előállított egy kimenet a 4.1 ábrán látható.

Tézishez kapcsolódó saját publikációk: [1], [2], [3].



4.1. ábra. A fix rekeszszámot használó vetületi leképező módszer egy kimenete. (a) ábrán látható az eredeti kép, (b) ábrán pedig a vetületi térkép. Összehasonlításképp a (c) ábrán a Radon transzformáció szinogramja látható.

1.2 tézis

Megterveztem és grafikus processzorokra implementáltam a fix rekeszszámú többirányú képi vetületi algoritmus adatpárhuzamos változatát, amellyel a végrehajtó egységek számával arányos gyorsítás érhető el.

A megoldás tervezésekor figyelembe vettem a GPU eszközök sajátos memória architektúráját, és törekedtem a gazdagép és az eszköz közötti memória átvitelek minimalizálására. A közös változók közötti versenyhelyzet elkerülésére az algoritmus áttervezése mellett döntöttem, így a helyes működést nem szükséges szinkronizációval biztosítani. Az eredmények összehordásakor atomi műveletek biztosítják a helyességet.

Az algoritmus alapját a bemeneti mátrix logikai blokkokra való osztása képezi, amely blokkok a GPU multiprocesszorainak *on-chip* memóriájába kerülnek. A blokkok feldolgozása szimultán történik, az előállított eredmények pedig a blokkok eredeti pozíciójából számolt eltolással kerülnek összehordásra.

A párhuzamos implementáció futásidőjét összehasonlítottam a szekvenciális változat, valamint a Radon transzformáció *Matlab* környezetben implementált GPU-gyorsított változatával is, amely utóbbi esetről is rövidebb idő alatt végezhető el a vetületekre való leképezés, amennyiben a rekeszszám a kép méreténél alacsonyabb.

Tézishez kapcsolódó saját publikációk: [2], [3].

1.3 tézis

Kiértékeltem a fix vektorhosszúságú többirányú képi vetületi módszer hatékonyságát objektumok párosítására, összevetve az eredményeket a hasonló vetületi alapú, alacsonyabb dimenziójú képi szignatúrákkal, és megállapítottam, hogy a párosítás pontossága szignifikánsan megnő.

A módszert objektumpárosításra alkalmazva összehasonlítottam a kettő- és négydimenziójú vetületi szignatúrákon alapuló módszerrel, valamint a Radon-transzfor-

mációval is. Az eredmények szerint a fix rekeszszámot alkalmazó módszerek esetén a valós és fals párokra számított hasonlósági mértékek távolsága nagyobb mint a 2D vagy 4D vetületi szignatúrákon alapuló módszerek esetén. A módszer összehasonlításra került a magasabb dimenziójú vetületi szignatúrát adó Radon transzformációval is, ugyanezzel az eredménnyel.

A fix felbontás eredményeként a memóriaköltség független a kép méretétől, ennélfogva a nem szükséges csúszóablakot alkalmazni a legjobb illeszkedés megtalálására, így is csökkentve a potenciális hibalehetőségeket.

Különböző felbontású vetületekkel elvégzett kísérletek alapján az is megfigyelhető, hogy a klasszifikáció teljesítménye nem romlik alacsony rekeszszámok alkalmazása esetén sem, memóriaköltség tekintetében is hatékony megoldást eredményezve.

Tézishez kapcsolódó saját publikációk: [3].

II. téziscsoport: Képi vetületek alkalmazása előfeldolgozásként Sziámi architektúrájú konvolúciós neurális hálózatokban

2.1 tézis

Kidolgoztam egy visszalépéses keresésen alapuló módszert, amely megadott méretű bemenet, rétegszám és memória költség mellett megadja az összes ezeknek megfelelő konvolúciós neurális hálózati architektúrát.

A korszerű konvolúciós tervezési minták vizsgálatát követően meghatároztam azokat a rétegekre vonatkozó követelményeket, amelyek alapján az optimalizáció elvégezhető. A megoldásban rétegenként sorban történik az aktivációs térképek méretének meghatározása, és a következő rétegek keresése. A visszalépéses keresésnek köszönhetően az összes lehetőség áttekintésére nincs szükség, azon réteg-kombinációk amelyek nem vezetnek érvényes megoldáshoz kihagyásra kerülnek.

Az algoritmus előbb megállapítja a konvolúciós és pooling rétegpárok kernel- és ablakméreteit, majd a paraméterek száma alapján a modell várható memóriaköltségét is megbecsli. Memória költség tekintetében felső korlát meghatározásával célhardverre optimalizált architektúrák generálhatóak.

A modell memóriafoglalásának becslését követően a kötegelt tanítás során alkalmazott kötegméret is optimalizálható. Nagyobb kötegméret alkalmazásával a tanítás futásideje rövidebb.

Tézishez kapcsolódó saját publikációk: [4].

2.2 tézis

Megterveztem és implementáltam egy termelő-fogyasztó elvű modellt Sziámi konvolúciós neurális hálózati architektúrák vizsgálatára elosztott környezetben, leghosszabb műveleti időközön alapuló ütemezéssel. A gyakorlati vizsgálatok során a generált neurális hálózati architektúrák feldolgozásának párhuzamosítása 99,87%-os hatékonyságú volt.

Az elosztott tanítás grafikus gyorsítókkal felszerelt számítógépes klaszterben történt. A megoldást a Master/Worker tervezési minta alapján alakítottam ki.

A terheléelosztás optimalizálására komplexitás becslésen alapuló LPT⁵ heurisztika szerinti ütemezést dolgoztam ki. A modellek tanítási komplexitását a tanítható paraméterek száma, valamint a kötegméret alapján határoztam meg.

A mérések igazolták, hogy a komplexitás-becslésen alapuló ütemezés hatásos, a teljes folyamatra vetített gyorsítása közel megegyezik a feldolgozásban résztvevő munkaállomások számával. A számított komplexitások és a mért feldolgozási idők közötti kapcsolat erős, a korrelációs koefficiens értéke 0.749.

A mért feldolgozási idők alapján az ütemezés hatékonyságának vizsgálatára szimulációkat végeztem, és megállapítottam, hogy a párhuzamos módszer automatikus terheléelosztása 90% feletti hatékonyságot eredményez véletlenszerű feladatkiosztás esetén is, a bemutatott ütemezés figyelembevételével azonban ez a mutató 99,87%.

Tézishez kapcsolódó saját publikációk: [5].

2.3 tézis

Megvizsgáltam az egyes többirányú képi vetületi módszereket, hogy azokat a bemenő adatok előfeldolgozásaként használva, milyen hatással vannak a Sziámi konvolúciós hálózatok hatékonyságára. Az eredmények alapján megállapítottam, hogy a fix rekeszszámú módszer hatékonyság és memóriaigény szempontjából Pareto optimális a referenciaként vizsgált nyers képi módszerekkel szemben.

A megvalósított kísérlet során a Radon, a Trace valamint az általam definiált transzformációk különböző paraméterezésekkel kerültek összehasonlításra. A definiált leképezések méretei alapján Sziámi konvolúciós architektúrák lettek generálva, amelyek elosztott környezetben kerültek feldolgozásra.

A betanított modellek objektum párosításának pontosságát *one-shot* klasszifikációs vizsgálatokkal határoztam meg, ahol a referenciával összehasonlított objektumok száma $2 \leq N \leq 10$.

Az eredmények pontosság és memóriafoglalás szerinti kiértékelése után megállapítottam, hogy a fix rekeszszámokon alapuló módszer Pareto optimális választás a képi bemeneteken alapuló módszerekkel szemben. A két szempont szerinti összehasonlítás eredményei a 4.2 ábrán láthatóak.

Tézishez kapcsolódó saját publikációk: [6].

5 Az eredmények hasznosítási lehetősége

Az objektumazonosítás vagy újraazonosítás⁶ egy fontos terület, nem csak a disszertációban tárgyalt gépjárművek követése és párosítása esetén. Elterjedt megoldások léteznek beléptetőrendszerek és kültéri kamerarendszerek esetén személyek azonosítására is.

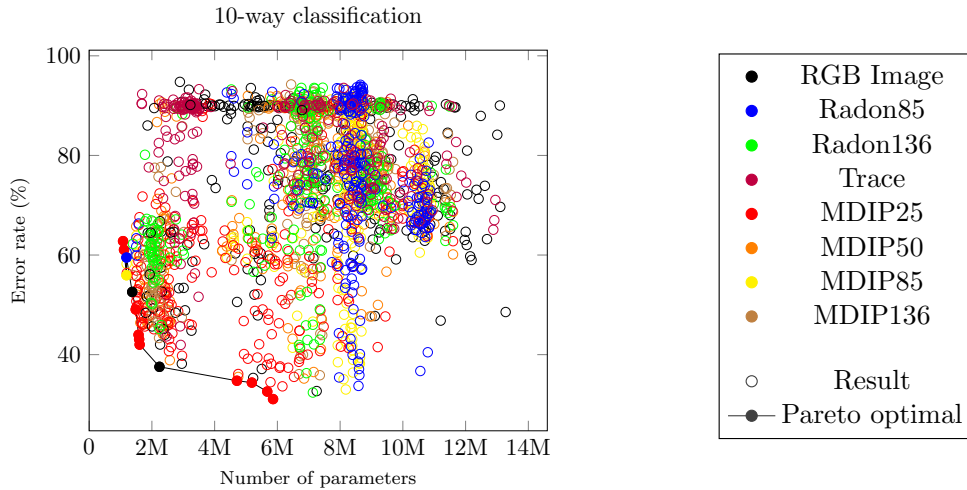
A kidolgozott módszer legnagyobb előnye, hogy nagyszerűen illeszkedik a mai IoT⁷ okoskamera⁸ rendszerekhez. A kamera a kép rögzítése mellett a feldolgozásában is részt vesz: detektálja a mozgó objektumokat, szegmentál, és klasszifikál. Ez

⁵Longest Processing Times

⁶object reidentification

⁷Internet of Things

⁸smart-camera



4.2. ábra. A modellek paraméterszáma és a validációs adatokon mért 10-osztályú one-shot klasszifikáció hibája szerinti eredmények. Az egyes modellek 10000 validációs vizsgálattal lettek tesztelve. A Pareto optimális modellek – azok, amelyeket nem dominál más elem – a bal alsó sarokban jelennek meg, mint Pareto-front.

utóbbi lépés helyett az objektum vetületi szignatúrája is kiszámítható és továbbítható hálózati kapcsolaton keresztül egy másik egységnek.

A fix rekeszszám előnye, hogy a memóriaköltség a bemeneti képi méretétől független, így a kommunikációs költségek jól meghatározhatóak. Amennyiben alacsony felbontású vetületi térkép készül, akkor a módszer tömörítésként viselkedik.

A párhuzamos megoldás nagy méretű képek esetén is gyors választást biztosít.

A kidolgozott módszer jól működik *one-shot* klasszifikációs problémákra is: olyan objektumok felismerésére is alkalmas, amelyekkel korábban, tanítási fázisban még nem találkozott.

A gépi tanulás alapú módszer vizsgálata során kidolgozott neurális hálózati architektúrákat generáló metódus bármilyen bemeneti mátrixra képes adott számú konvolúciós és pooling rétegpárokból álló struktúra generálására egy előre meghatározott memória limit betartása mellett. Ennélfogva a megoldás más esetekben is alkalmazható, ahol esetleg több, memóriakritikus neurális architektúrára van szükség.

Irodalmi hivatkozások listája

- Bay, Herbert és tsai. (2008). “Speeded-up robust features (SURF)”. *Computer vision and image understanding* 110.3, 346–359. old.
- Bengio, Yoshua (2012a). “Deep learning of representations for unsupervised and transfer learning”. *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 17–36. old.
- (2012b). “Practical Recommendations for Gradient-Based Training of Deep Architectures”. *Neural networks: Tricks of the trade*. Springer, 437–478. old.
- Betke, Margrit, Esin Haritaoglu és Larry S Davis (2000). “Real-time multiple vehicle detection and tracking from a moving vehicle”. *Machine vision and applications* 12.2, 69–83. old.
- Bromley, Jane és tsai. (1994). “Signature verification using a " siamese" time delay neural network”. *Advances in neural information processing systems*, 737–744. old.
- Brunelli, R. (2009). *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley.
- Choe, Tae Eun, Mun Wai Lee és Niels Haering (2010). “Traffic Analysis with Low Frame Rate Camera Networks”. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 9–16. old. ISBN: 978-1-4244-7030-3. DOI: 10.1109/CVPRW.2010.5543801.
- Chopra, Sumit, Raia Hadsell és Yann LeCun (2005). “Learning a similarity metric discriminatively, with application to face verification”. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 1. köt. IEEE, 539–546. old.
- Deans, Stanley R. (1983). *The Radon Transform and Some of Its Applications*. New York: John Wiley és Sons.
- Ginkel, M. van, C.L. Luengo Hendriks és L.J. van Vliet (2004). *A short introduction to the Radon and Hough transforms and how they relate to each other*. Techn. jel. QI-01-2004.
- Goodfellow, Ian és tsai. (2016). *Deep learning*. 1. köt. MIT press Cambridge.
- Guo, Yanlin, Steve Hsu és tsai. (2007). “Robust Object Matching for Persistent Tracking with Heterogeneous Features”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.5, 824–839. old. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2007.1052.
- Guo, Yanlin, Cen Rao és tsai. (2008). “Matching vehicles under large pose transformations using approximate 3D models and piecewise MRF model”. *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 8. old.
- Hou, Tingbo, Sen Wang és Hong Qin (2009). “Vehicle Matching and Recognition under Large Variations of Pose and Illumination”. *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*. IEEE, 24–29. old. ISBN: 978-1-4244-3994-2. DOI: 10.1109/CVPRW.2009.5204071.
- Hough, Paul VC (1962. dec.). *Method and means for recognizing complex patterns*. US Patent 3,069,654.

- Jelača, Vedran és tsai. (2013). “Vehicle matching in smart camera networks using image projection profiles at multiple instances”. *Image and Vision Computing* 31, 673–685. old.
- Kadyrov, Alexander és Maria Petrou (2001). “The trace transform and its applications”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.8, 811–828. old.
- Karpathy, Andrej, FF Li és J Johnson (2017). “CS231n: Convolutional Neural Networks for Visual Recognition, 2016”. URL <http://cs231n.github.io>.
- Koch, Gregory, Richard Zemel és Ruslan Salakhutdinov (2015). “Siamese Neural Networks for One-shot Image Recognition”. *ICML 2015 Deep Learning Workshop*.
- Krizhevsky, Alex, Ilya Sutskever és Geoffrey E. Hinton (2012). “ImageNet Classification with Deep Convolutional Neural Networks”. *Proceedings of the 25th International Conference on Neural Information Processing Systems. NIPS’12*. Lake Tahoe, Nevada: Curran Associates Inc., 1097–1105. old.
- LeCun, Yann A és tsai. (2012). “Efficient BackProp”. *Neural networks: Tricks of the trade*. Springer, 9–48. old.
- LeCun, Yann, Yoshua Bengio és tsai. (1995). “Convolutional networks for images, speech, and time series”. *The handbook of brain theory and neural networks* 3361.10.
- LeCun, Yann, Yoshua Bengio és Geoffrey Hinton (2015). “Deep learning”. *Nature* 521.7553, 436. old.
- LeCun, Yann, Bernhard Boser és tsai. (1989). “Backpropagation applied to handwritten zip code recognition”. *Neural computation* 1.4, 541–551. old.
- Lee, Seungkyu, Yanxi Liu és Robert Collins (2007). “Shape variation-based frieze pattern for robust gait recognition”. English (US). *2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR’07*. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. ISBN: 1424411807. DOI: 10.1109/CVPR.2007.383138.
- Lindeberg, Tony (1998). “Feature detection with automatic scale selection”. *International journal of computer vision* 30.2, 79–116. old.
- Liu, Yanxi, Robert Collins és Yanghai Tsin (2002). “Gait sequence analysis using frieze patterns”. *European Conference on Computer Vision*. Springer, 657–671. old.
- Lowe, David G és tsai. (1999). “Object recognition from local scale-invariant features”. *Proceedings of the International Conference on Computer Vision*. 2, 1150–1157. old.
- Oren, Michael és tsai. (1997). “Pedestrian detection using wavelet templates”. 193–199. old.
- Oriol Vinyals and Charles Blundell and Timothy P. Lillicrap and Koray Kavukcuoglu and Daan Wierstra (2016). “Matching Networks for One Shot Learning”. Szerk. D. D. Lee és tsai., 3630–3638. old.
- Papageorgiou, Constantine P, Michael Oren és Tomaso Poggio (1998). “A general framework for object detection”. *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*. IEEE, 555–562. old.
- Radon, Johann (1917). “Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten”. *Berichte über die Verhandlungen der Königlich-Sächsischen Akademie der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, 262–277. old.

- Raina, Rajat, Anand Madhavan és Andrew Y. Ng (2009). “Large-scale Deep Unsupervised Learning Using Graphics Processors”. *Proceedings of the 26th Annual International Conference on Machine Learning*. ICML '09. Montreal, Quebec, Canada: ACM, 873–880. old. ISBN: 978-1-60558-516-1. DOI: 10.1145/1553374.1553486.
- Rinner, Bernhard és Wayne Wolf (2008). “An Introduction to Distributed Smart Cameras”. *Proceedings of the IEEE* 96.10, 1565–1575. old.
- Rios-Cabrera, Reyes, Tinne Tuytelaars és Luc Van Gool (2012). “Efficient Multi-camera Vehicle Detection, Tracking, and Identification in a Tunnel Surveillance Application”. *Comput. Vis. Image Underst.* 116.6, 742–753. old. ISSN: 1077-3142. DOI: 10.1016/j.cviu.2012.02.006.
- Sanchez, Angel és tsai. (2011). “Video-Based Distance Traffic Analysis: Application to Vehicle Tracking and Counting”. *Computing in Science and Engg.* 13.3, 38–45. old. ISSN: 1521-9615. DOI: 10.1109/MCSE.2010.143.
- Schmid, Cordelia, Roger Mohr és Christian Bauckhage (2000). “Evaluation of interest point detectors”. *International Journal of computer vision* 37.2, 151–172. old.
- Schmidhuber, Jürgen (2015). “Deep learning in neural networks: An overview”. *Neural networks* 61, 85–117. old.
- Schroff, Florian, Dmitry Kalenichenko és James Philbin (2015). “FaceNet: A Unified Embedding for Face Recognition and Clustering”. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015* abs/1503.03832.
- Shan, Ying, Harpreet S. Sawhney és Rakesh Kumar (2005). “Vehicle Identification between Non-Overlapping Cameras without Direct Feature Matching”. *10th IEEE International Conference on Computer Vision (ICCV'05)* 1, 378–385. old.
- Shi, Jianbo és Carlo Tomasi (1993). *Good features to track*. Techn. jel. Cornell University.
- Szegedy, Christian és tsai. (2015). “Going Deeper with Convolutions”. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9. old.
- Szeliski, Richard (2010). *Computer Vision: Algorithms and Applications*. 1st. New York, NY, USA: Springer-Verlag New York, Inc. ISBN: 1848829345, 9781848829343.
- Taigman, Yaniv and Yang, Ming and Ranzato, Marc’Aurelio and Wolf, Lior (2014). “DeepFace: Closing the Gap to Human-Level Performance in Face Verification”. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. CVPR '14. Washington, DC, USA: IEEE Computer Society, 1701–1708. old. ISBN: 978-1-4799-5118-5. DOI: 10.1109/CVPR.2014.220.
- Viola, Paul és Michael Jones (2001a). “Rapid Object Detection using a Boosted Cascade of Simple Features”. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. 1. köt. IEEE, 511–518. old.
- (2001b). “Robust real-time object detection”. *International journal of computer vision* 4.34-47, 4. old.
- Yuan, Quan és Stan Sclaroff (2009). “Is a detector only good for detection?": *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 1066–1073. old.

A tézispontokhoz kapcsolódó tudományos közlemények

- [1] Kertész, Gábor, Sándor Szénási és Zoltán Vámosy (2017). “Application and properties of the Radon transform for object image matching”. *Proceedings of SAMI 2017. IEEE 14th International Symposium on Applied Machine Intelligence and Informatics* (Herlany, Slovakia, 2017. jan. 26–28.). IEEE, 353–358. old.
- [2] — (2016). “A Novel Method for Robust Multi-Directional Image Projection Computation”. *Proceedings of INES 2016. 20th IEEE International Conference on Intelligent Engineering Systems* (Budapest, Hungary, 2016. jún. 30. – júl. 2.). IEEE, 239–243. old.
- [3] — (2018). “Multi-Directional Image Projections with Fixed Resolution for Object Matching”. *Acta Polytechnica Hungarica* 15.2, 211–229. old.
- [4] — (2019a). “A novel method for Convolutional Neural Architecture Generation with memory limitation”. *Proceedings of SAMI2019. IEEE 17th World Symposium on Applied Machine Intelligence and Informatics* (Herlany, Slovakia, 2019. jan. 24–26.). IEEE, 229–234. old.
- [5] — (2019b). “Distributed training and evaluation of projection-based descriptors in Siamese Neural Networks”. *Proceedings of the Sixth International Conference on Parallel, Distributed, GPU and Cloud Computing for Engineering*, Paper 25., 1–12. DOI: 10.4203/ccp.112.25.
- [6] — (2019c). “Multi-Directional Projection Transformations for Machine Learning based Object Matching”. *SACI 2019 : IEEE 13th International Symposium on Applied Computational Intelligence and Informatics*, 269–274. old.

További tudományos közlemények

- [7] Kertész, Gábor, Sándor Szénási és Zoltán Vámosy (2015a). “Performance Measurement of a General Multi-Scale Template Matching Method”. *Proceedings of INES 2015. 19th IEEE International Conference on Intelligent Engineering Systems* (Bratislava, Slovakia, 2015. szept. 3–5.). IEEE, 153–158. old.
- [8] — (2015b). “Parallelization Methods of the Template Matching Method on Graphics Accelerators”. *Proceedings of CINTI 2015. 16th IEEE International Symposium on Computational Intelligence and Informatics* (Budapest, Hungary, 2015. nov. 19–21.). IEEE, 161–164. old.
- [9] Kertész, Gábor, Szabolcs Sergyán’ és tsai. (2016). “Implementation of Object Recognition based on Image Projection Signatures using Matlab”. *Proceedings of CINTI 2016. 17th IEEE International Symposium on Computational Intelligence and Informatics* (Budapest, Hungary, 2016. nov. 17–19.). IEEE, 123–128. old.
- [10] Kertész, Gábor, Sándor Szénási és Zoltán Vámosy (2018b). “Vehicle Image Matching Using Siamese Neural Networks with Multi-Directional Image Projections”. *12th IEEE International Symposium on Applied Computational Intelligence and Informatics, SACI 2018, Timisoara, Romania, May 17-19, 2018*, 491–496. old.
- [11] Kertész, Gábor és Eva Hajnal (2013). “Irisz Project: A Web Application for the Introduction of University Students to the Labor Market”. *International Symposium on Applied Informatics and Related Areas : AIS 2013 Szekesfehervar, Magyarország : Óbudai Egyetem, (2013)*, 125–129. old.
- [12] — (2014). “Special Issues in the Development of a Large User Based Web Application”. *Proceedings of the IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMI 2014) Budapest, Magyarország : IEEE Hungary Section, (2014)*, 141–145. old.
- [13] Kertész, Gábor és Zoltán Vámosy (2015). “Current challenges in multi-view computer vision”. *10th IEEE Jubilee International Symposium on Applied Computational Intelligence and Informatics, SACI 2015, Timisoara, Romania, May 21-23, 2015*, 237–241. old.
- [14] — (2016). “A Brief Review of Recent Advances in Multi-View Computer Vision”. *Scientific Bulletin of Politechnica University of Timisoara - Transactions on Automatic Control and Computer Science* 61(75), 73–78. old. ISSN: 1224-600X.
- [15] Kertész, Gábor, Daniel Kiss és tsai. (2016). “Multiprocessing of an individual-cell based model for parameter testing”. *11th IEEE International Symposium on Applied Computational Intelligence and Informatics, SACI 2016, Timisoara, Romania, May 12-14, 2016*, 491–496. old.
- [16] Szénási, Sándor és tsai. (2018a). “Comparison of Road Accident Black Spot Searching Methods”. *IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI 2018) Budapest, Magyarország : IEEE Hungary Section, (2018)*, 247–250. old.

- [17] Szenasi, Sandor és tsai. (2018b). “Road Accident Black Spot Localisation using Morphological Image Processing Methods on Heatmap”. *IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI 2018) Budapest, Magyarország : IEEE Hungary Section, (2018)*, 251–256. old.
- [18] Danko, Bence és Gabor Kertesz (2018a). “Recognition of the Hungarian Fingerspelling Alphabet using Convolutional Neural Network based on Depth Data”. *IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI 2018) Budapest, Magyarország : IEEE Hungary Section, (2018)*, 41–46. old.
- [19] — (2018b). “Recognition of the Hungarian fingerspelling alphabet using Recurrent Neural Network”. *SAMI 2019 : IEEE 17th World Symposium on Applied Machine Intelligence and Informatics*, 251–256. old.

