

Hajder Levente

***Geometriai alapú háromdimenziós számítógépes látás
digitális kamerák és LiDAR érzékelők segítségével***

c. MTA doktori értekezésének bírálata

A Jelölt az NJSZT Képfeldolgozók és Alakfelismerők Társasága Szakosztálya (KÉPAF) aktív tagja. Több évtizede kutatja a számítógépes látás területét, eredményeit nemzetközi szinten ismerik, elismerik és hivatkozzák, továbbá kiemelkedően eredményes a kutatói utánpótlás-nevelés terén is. Értekezésében a doktori (PhD) fokozata megszerzése óta eltelt másfél évtized kutatómunkájának eredményeit összegezte és rendezte négy téziscsoportba. A Jelölt doktori művét magyar nyelven készítette el, ami dicséretesen egybeesik az MTA törekvésével: „A Magyar Tudományos Akadémiának alapításától fogva fontos küldetése a magyar nyelv művelése, védelme, a magyar tudományos nyelv ápolása, megtartása.”¹

Bírálatomban formai észrevételeimet markánsan szétválasztottam a szakmai értékeléstől, mivel szinte az összes kritikai megjegyzésem formai, míg az elért eredményeket méltatásra érdemesnek tartom.

Kezdem a formai észrevételekkel:

A doktori mű logikusan felépített, arányosan tagolt, a szűk terjedelmi korláton — amely szerint az értekezés érdemi fejezeteinek terjedelme nem haladhatja meg a 100 oldalt — nem terjeszkedik túl, számos fontos részlet megértését jól szolgálja a 31 oldalas, 14 részből (A-N) álló függelék. Az értekezés jól olvasható, mentes a bántó nyelvhelyességi hibáktól, csak minimális számú elütéssel terhelt. A fenti erények mellett, sajnos, számos kifogásolható elembe ütköztem. Az alábbiakban néhányat fel is sorolok közülük:

- Az értekezés elején (a 2. oldalon) a Jelölt részletesen taglalja az általa követett „magyarosítási” stratégiát, táblázatba is foglalja a hét legfontosabb angol kulcsszó „magyarosított” változatát. Ugyanakkor szerencsésnek tartottam volna, ha a mű további részében következetesen a szakszavak első előfordulásakor zárójelben megadta volna azok angol megfelelőit. Követendő példának tartom a 11. oldalról az alábbi mondatot: „Ez a feladat a számítógépes grafikában [106,41] a sugárvetés (ray-casting), amely a sugárkövetés (ray-tracing) alapproblémája.”
- A 2. lábjegyzet (az értekezés 10. oldalán) hibás: „Angolul: rektifikáció.”
- A többsoros egyenletek igazítása számos esetben nem csupán esztétikailag kifogásolható, hanem a megértést is nehezíti.
- Zavaró a mondatokba ékelt egyenleteknél az egyenleteket követő szövegsorok behúzása (ami az új bekezdések kezdetének illúzióját kelti).
- Tucatnyi oldalon a jobb oldali margón jócskán túllógó egyenlet vagy szövegrész található.

¹ https://mta.hu/data/dokumentumok/egyeb_dokumentumok/2024/197_KGY_magyar_nyelv_vedelme_javaslatok.pdf

- A 30. oldal alján a fattyú sor (özvegy sor) tördelési hiba.
- A DLT mozaikszó először a 29. oldalon jelenik meg, irodalmi hivatkozása ([62]) a 30. oldalon található, feloldására (Direct Linear Transformation) csak a 35. oldalon a 11. lábjegyzetben kerül sor.
- A G., az I., valamint az M. függelék-elemekre nem találtam hivatkozást az értekezésben.
- Több mű egyidejű hivatkozásakor a források sorrendje sokszor esetleges (pl. [115, 103, 21, 142] a 34. oldalon).
- A Blender rendszer először a 71. oldalon a 30. lábjegyzetben kerül elő, majd két további említését követően, csak a 87. oldalon kap egy irodalmi hivatkozást és magyarázatot az olvasó.
- Zavaró, hogy a Jelölt ugyanazt a módszert háromféleképpen is említi: Levenberg-Marquardt optimalizálás, a Levenberg-optimalizálás és a Levenberg-Marquardt algoritmus.
- Háromszor fordul elő a Jakobi-mátrix, kétszer pedig a Jacobi-mátrix, holott ugyanarról (a személyről és mátrixról) van szó.
- Az értekezés összegzésébe (105. oldal) tévesen került be a következő kitétel: „*Amiről nem szól, nem szólhat ez a tézisfüzet (kiemelés tőlem – P.K.), az a jövő.*”
- Az Irodalomjegyzékben kifogásolható, hogy pl. hiányoznak a [2] szerzői; nincs hiperhivatkozással ellátva a [3] linkje (ami ráadásul nem is működik); a szerzőlistán és a címen túl minden adat hiányzik a [138] forrásműnél; a konferenciák leírása következetlen; hiányoznak az oldalszám-intervallumok számos műnél; az első nevek (keresztnevek) hol részletesen kiírva, hol pedig csak kezdőbetűikkel rövidítve szerepelnek; megbicsaklik az első szerző családneve szerinti alfabetikus sorrend.

Az elmaradt gondos korrektúrázás azért is fájó, mivel az értekezés megírásába — különös tekintettel az angolul közölt eredmények magyar nyelvű bemutatásába — fektetett tiszteletre méltóan nagy munkához képest elhanyagolható időt, csak egy-két órát vett volna igénybe egy formai szempontokból is kifogástalan és esztétikus mű elérhetővé tétele.

A formai kritika után rátérek az értekezés tartalmi, szakmai szempontok szerinti értékelésre:

A tömör (egyoldalas 1. fejezet) bevezetést követő 2. fejezet (az A. függelékkel együtt) az elméleti alapok jó áttekintését adja. A négy téziscsoport részletes bemutatására rendre a következő négy (3-6.) fejezet szolgál. A művet az egyoldalas összegzés, a 191 elemű irodalomjegyzék és a 14 részből (A-N) álló függelék zárja.

Pozitívumnak tartom azt, hogy a téziscsoportokhoz rendelt 3-6. fejezetek szakirodalmi áttekintéssel kezdődnek. Ugyanakkor feltűnő, hogy idegen szerzőktől és az elmúlt négy év közleményeiből csupán kettőt (!) tartalmaz az Irodalomjegyzék. Dacára annak, hogy a doktori művekben a jelöltek eredményeinek bemutatása a hangsúlyos, szerencsés lett volna több „friss” publikációval pozicionálni a saját munkásságot, különös tekintettel az intenzíven kutatott LiDAR-os területre.

A Jelölt az I. téziscsoport eredményeit két tézispontba foglalta, az első a lokális affin transzformációk és a kétképes (sztereó) látás kapcsolatát leíró három „törvény” lefektetésével, a második pedig az epipoláris geometria és az affin transzformációk között lineáris kapcsolat megteremtésével foglalkozik. A téziscsoportához kapcsolódó közlemények: egy (SJR: D1 kategóriájú) folyóiratcikk és kettő, rangos nemzetközi konferenciák kiadványaiban megjelent konferenciatickk.

1. kérdés: A doktori mű 3. fejezetében, az affin transzformáció dekompozíciós formulájában a nyírás transzformációs mátrixa csak az

$$\begin{pmatrix} 1 & b_x \\ 0 & 1 \end{pmatrix}$$

alakban szerepel, mellőzve a függőleges tengely mentén deformáló

$$\begin{pmatrix} 1 & 0 \\ b_y & 1 \end{pmatrix}$$

esetet. Hatással lenne-e az I. téziscsoport eredményeire az, ha a nyírás második alakját vagy még inkább a kétféle nyírás kompozícióját venné figyelembe?

A II. téziscsoport kilenc tézispontot tartalmaz, így nem meglepő, hogy a doktori mű 4. fejezete a leghosszabb, meghaiadja a rendelkezésre álló terjedelem harmadát. Az itt bemutatott kilenc algoritmus az I. téziscsoport elméleti eredményein alapul. A II. téziscsoportot meggyőzően támasztja alá három folyóiratban megjelent cikk (kettő közülük SJR:D1, a harmadik pedig SJR:Q1 minősítésű), továbbá a kilenc rangos konferenciaközlemény.

2. kérdés: A Jelölt a CLOSED és az ALTER eljárásait a klasszikus DLT algoritmussal gömbön és kockán mintavételezett felszínpontok és a hozzájuk tartozó normálvektorok segítségével hasonlította össze. Érdekes lenne-e a jövőben a kockánál bonyolultabb konvex poliéderekkel is tesztelni? A híres magyar Gömböcnek (ami egy stabil és egy instabil egyensúlyi helyzettel bír) lehetnének-e tesztobjektumként is előnyös tulajdonságai?

3. kérdés: Meggyőző és ötletes a 12. ábrán bemutatott kalibrációs eszköz. Pontosabb kalibrációt eredményezne-e a kocka helyett pl. a csonkolt oktaéder (az FCC rács Voronoi cellája) vagy a rombikus dodekaéder (a BCC rács Voronoi cellája)?

4. kérdés: a 4.3.5. pont utolsó bekezdésében a Jelölt az alábbiakat szögezi le: „A futási idő tekintetében is fontos eredmény, hogy csak pontmegfelelést használva három pontpárra van szükségünk a homográfia megbecsléséhez a fundamentális mátrix ismerete esetén, míg az affin transzformáció megléte egyetlen egy megfelelésre szorítja le a minimális modell elkészítéséhez szükséges bemeneteket.” Szükséges-e az, hogy a pontok ne essenek egy egyenesbe? Függ-e a becslés pontossága a pontok konvex burkának területétől? A pont-kijelölések pontossága hogyan befolyásolja a becslés pontosságát? (Ezek a kérdések a pont-alapú képregisztrációnál már vizsgálat tárgyát képezték.)

A III. téziscsoport a köteget behangolás alkalmazásaival foglalkozik. Az első tézis a fotometrikus sztereóra, a második pedig az irányított pontok rekonstrukciójára (affin transzformációkkal) vonatkozik. Az eredményeket egy (SJR: Q2) folyóiratcikk és egy konferenciaközlemény tartalmazza.

5. kérdés: A 3.1. tézishoz tartozó BA módszer a Levenberg-Marquardt optimalizálást alkalmazza. A hatékony és számos kutató által előnyben részesített módszer mellett milyen további optimalizálók lehetnének még ígéretesek?

A IV. téziscsoport a LiDAR mérőeszközök és a perspektív kamerák kalibrációjával foglalkozik. A csoport első két téziséhez a doboz és a gömb segítségével történő kalibráció eredményei tartoznak, a harmadik tézis pedig az egyetlen forgatási paraméterre redukált kamera-LiDAR kalibráció problémájára ad megoldást. Ennek a téziscsoportnak a publikáltsága is meggyőző, az eredményeket a jelölt két folyóiratcikkben (egy SJR:D1 és egy SJR:Q2 kategóriájában), valamint három konferenciaközleményben közölte. Ezeknek a publikációknak a visszhangja bizonyult a legnagyobbak: eddig több mint 200 független mű hivatkozik rájuk.

6. kérdés: A doboz és a gömb mellett milyen további poliéderekkel és parametrikus felületekkel lenne megoldható és célszerű a kalibráció? (A 2. és a 3. kérdéshez hasonló ez a felvetés is.)

7. kérdés: A Jelölt az utolsó három téziscsoport problémáira javasolt megoldásaiban a robusztus RANSAC algoritmus három (standard, szekvenciális, gráf-vágás) variánsát is alkalmazta. A 6.3.4. pontban megemlíti azt is, hogy a LiDAR mérőeszközök és a perspektív kamerák kalibrációja néhány percet igényel, ahol a futási idő oroszlánrészét a RANSAC-alapú gömb- és ellipszisillesztés emészti fel. Célszerű lenne-e – a pontosabb becslés elérése és a futási idő rövidülése érdekében – további RANSAC-változatokat² is megvizsgálni?

8. kérdés: Végezetül napjaink megkerülhetetlen és szorongatóan súlyos kérdése: A Jelölt által művelt területekbe milyen mértékben és eredményességgel türemkedett be a mélytanulás és a mesterséges intelligencia? Azok (létező) modelljeibe beépíthetők-e a Jelölt eredményei?

² További RANSAC-variánsokon például az alábbi közleményekben javasoltakat értem:

- R. Raguram, J.-M. Frahm, and M. Pollefeys: *A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus*, Proc. European Conf. Computer Vision, 500-513. 2008.
- O. Chum and J. Matas: *Optimal Randomized RANSAC*, IEEE Trans. Pattern Anal. Mach. Intell. 30, 1472-1482, 2008.
- R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm: *USAC: A Universal Framework for Random Sample Consensus*, IEEE Trans. Pattern Anal. Mach. Intell. 35, 2022-2038, 2013.

Összegzés:

A doktori mű tézisei és az azokat kibontó közlemények alapján megállapítható, hogy a Jelölt jártas és nagy biztonsággal mozog számos tudományterületen (pl. számítógépes grafika, geometria, lineáris algebra, operációkutatás, számítógépes látás). Publikációs aktivitása meggyőző és — dacára annak, hogy csak társszerzős közleményeket sorakoztatott fel tézisei alátámasztására — az elért eredményekhez meghatározó a hozzájárulása. Az elméleti eredmények biztos alapul szolgálnak a felmerült gyakorlati problémák megoldásához, amelyekhez ésszerű tesztelési környezetek társulnak. Mind a négy téziscsoport összes tézisétfelfogadom új és jelentős tudományos eredménynek.

Az értekezésbe foglalt és hitelesnek talált eredmények mennyiségét és minőségét elegendőnek tartom az MTA doktora cím megszerzéséhez, javasolom a nyilvános vita kitézését.

Szeged, 2025.11.25.


Palágyi Kálmán
MTA doktora