

Hajder Levente „Geometriai alapú háromdimenziós számítógépes látás digitális kamerák és LiDAR érzékelők segítségével” című MTA doktori disszertációjának bírálata.

A disszertáció témája a gépi látás, elsősorban a 3D objektumrekonstrukció képek alapján.

A dolgozat a választott téma alapjainak bemutatásával kezdődik.

A 3. Fejezet az I. téziscsoportot részletezi, amely az affin megfeleltetésekre épül. Ennek lényege, hogy egy 3D térbeli pontnak a két képre vetítésének nem csupán a lokális pontokban vett értékét használjuk, hanem a kétváltozós, 2D vektor értékű függvény elsőrendű Taylor sorát, azaz a deriváltakat is felhasználjuk. Világos, hogy egy függvénykapcsolatról sokkal több információt kapunk, ha nem csak egy-egy pontban felvett értéket ismerjük, hanem a deriváltakat is. Geometriai oldalról ez azt jelenti, hogy nem csak a 2D pontok közötti kapcsolatot adjuk meg, hanem az ottani felület érintősíkjait is összerendeljük. Az extra információ lehetőséget ad pontosabb algoritmusok kidolgozására, amit a dolgozatban láthatunk. Ide kapcsolódik az első kérdés. Miért állunk meg az első deriválnál, miért nem lépünk tovább a görbületet meghatározó második deriváltakra, amelyek még tovább pontosíthatnák a kapcsolatokat. Ebben a részben érdemes lett volna az affin transzformációk geometriai definícióját is megadni (párhuzamostartó, egyenest-egyenesbe képző transzformáció), illetve megemlíteni, hogy az affin tulajdonságból következik a linearitás és fordítva. Az, hogy egy affin transzformáció mindig felbontható forgatás, skálázás és eltolás transzformációkra, ismert tény.

A dolgozat Molnár József és Csetverikov Dmitrij eredményére épít, akik egyenletet írtak fel lokális affin transzformációk (azaz a deriváltak) meghatározására a térbeli pont helyének, a kameraképnek és a felületre merőleges normálisnak függvényében.

A disszertáció „törvények” között sorolja fel az általános összefüggést egyszerűsítését homográfia (azaz a két kép közötti homogén lineáris transzformáció) esetére, illetve a fundamentális mátrix (amely a két képre kerülő vetületre határoz meg kényszert) és az affin paramétereket összekapcsolja, amiből az utóbbit (II.2 altézis) tartom tézisértékűnek.

A második téziscsoport az affin transzformációt előállító összefüggés alkalmazásait mutatja be. Az első alkalmazásban (II.1. altézis) a cél a normálvektor meghatározása, azaz a korábban sugallt megközelítéssel, miszerint a látható pont felületi normálisa ad lehetőséget az affin paraméterek meghatározására, most feltételezzük, hogy az affin paraméterek ismertek és abból határozzuk meg a normálvektort. Itt a gyakorlati alkalmazhatóság alátámasztásához jó lett volna kitérni arra, hogy hogyan teszünk szert ebben az esetben az affin paraméterekre. Sajnos a dolgozatban ezen módszerhez nem tartozik numerikus vagy valós demonstráció, amiből a részletek kiderülhetek volna. A feladat maga egy inverz probléma, amit nem analitikusan old meg a Jelölt, hanem a direkt egyenlet két oldalának eltérését minimalizálja. Kérdésem, volt-e kísérlet az analitikus megoldásra, és ha igen, az miért hiúsult meg.

A II.2. altézis a II.1-hez hasonlóan az affin transzformációt meghatározó összefüggés inverzét vizsgálja, de most nem a normálvektort, hanem a kameraparamétereket tekinti ismeretlennek. A kameraparaméterek meghatározását kalibrációnak nevezi a gépi látás irodalma. A feladat megoldására három módszert is tárgyal a dolgozat, amelyeket gömböt és dobozt (síkokat) tartalmazó virtuális és valós

környezetben kipróbál. A feladat speciális, a normálvektorok és az affin transzformáció ismertek. Világos, hogy a tárgyalt kalibrációs alkalmazásokban ezek a feltételek teljesülnek, de jó lenne kitérni arra, hogy milyen más gyakorlati feladatoknál lehet ezeket készen megkapni, azaz hol vannak a felhasználás korlátai. Szerencsétlennek érzem azt a kijelentést, hogy „felületi pontokat igyekeztünk nagyjából egyenletesen mintavételezni”. A gömb és kocka felületén az „egyenletes” mintavételezés értelmezésének komoly háttere van (diszkrepancia, kvázi-Monte Carlo, Poisson-diszk mintavételezés, stb.), ha pedig az nem lényeges, akkor nem is érdemes nem precíz kijelentéseket tenni.

A II.3. altézis a két kép közötti homogén lineáris transzformáció paramétereit becsüli ugyancsak az affin paraméterekből. Homográfia akkor áll fenn, ha a látott pontok egy síkon vannak. Ebben az esetben a lokális affin transzformációt kell kiterjeszteni a síkra, így ígéretes annak felhasználása. Itt megemlíthető volna a projektív síkgeometriabeli dualitás fogalom, mert ez világosabbá tenné az érvelést.

A disszertáció a hagyományos DLT OpenCV implementációját veti össze a klasszikus fundamentális mátrixos és a két új megoldással, a HA: normalizált affin és a HAF: affin transzformációt és a fundamentális mátrixot egyaránt alkalmazó módszerrel, és megmutatja, hogy az új módszerek, különösen a HAF lényegesen kisebb visszavetítési hibát vétenek. A tesztek, különösen az említett publikációkkal együtt, alátámasztják a javasolt módszer jóságát a hagyományos eljárásokkal szemben.

A II.4. tézisben megfogalmazott síkok klaszterezése (szegmentálása) lokális affin transzformációk segítségével, kalibrált kamerák esetén értékes algoritmus síkok detektálására, amelynek nagy gyakorlati jelentősége van épületek 3D geometriai rekonstrukciójában. Az ötlet a RANSAC módszerre emlékeztet. Az egy síkon lévő pontok homográfiával transzformálhatók, ami lehetőséget ad arra, hogy véletlenül kiválasztott pontokról eldöntsük, hogy azok egy síkon vannak el.

A II.5. tézis módszert ad a fundamentális és esszenciális mátrixok becslésére ugyancsak a lokális affin transzformációból. A II.6. tézis ennek a tézisnek egy speciális esete, amikor csak a fókusz-távolság meghatározása a cél, a többi paraméter ismert. A meggyőző teszteredmények a 4.7. fejezetben találhatók.

A II.7. tézis a síkban mozgás elmozdulásának irányát és az elfordulási szöget határozza meg lokális affin transzformációból. A megoldás speciális, de gyakorlatban fontos elrendezést feltételez: a gépkocsira rögzített kamera optikai tengelye vízszintes, az út vízszintes sík. A II. 8. altézis egy speciális esetet vizsgál, amikor ismeretlen a fókusz-távolság és a kameraparaméterek a mozgás során rögzítettek.

A II.9. tézis magának az affin transzformációnak a meghatározásával foglalkozik, amit egymásnak megfeleltetett egyenesek irányából, illetve a fundamentális mátrixból vagy optikai áramlásból nyer ki.

A III. téziscsoport a kötegelt behangolás módszertanát új problémákra alkalmazza. A kötegelt behangolás (bundle adjustment) egy nemlineáris optimalizációs eljárás a visszavetítési hiba minimalizálására, amihez az összes képet felhasználjuk. A kötegelt behangolást tipikusan a SLAM feladat megoldásánál használják. A javasolt adaptáció azon alapul, hogy a kötegelt behangolás felfogható úgy, mint speciális típusú ritkamátrixok invertálása.

III.1. tézisben a cél a kötegelt behangolás alkalmazása fotometrikus sztereóra. A megoldást eredetinek, értékesnek és kellően alátámasztottnak tartom. A tárgyalás egyes pontjaival kapcsolatban azonban vannak kritikai megjegyzéseim. A disszertáció megemlíti, hogy a számítógépes grafikában hogyan kezelik az érdességet (mezostruktúra). A helyzet a leírásnál bonyolultabb, létezik csak az árnyalásnál felhasznált bump-mapping, vagy a geometriát is módosító displacement-mapping, sőt távolságfüggő folyamatos átmenet is a mezostruktúra és a BRDF-fel leírt mikrostruktúra között. A „más rekonstrukciós technikák” között ma nem megkerülhetők a mélységkamera alapú megoldások sem. A „megvilágítási egyenlet” nehezen érthető, abban skalárok, vektorok és spektrumok (szín) szerepelnek, amit jó volna a jelölésekben is visszatükrözni (a különböző típusokra más műveletek érvényesek). Úgy tűnik, hogy az 52. egyenlet / változója csak egyetlen hullámhosszhoz kapcsolódik. Szerencsétlen a **b** változó alkalmazása is, mert ha az csak a normálvektor, akkor, ahogy korábban, **n**-nel kellene jelölni, ha pedig nem csak az, akkor pl. az 57. egyenlet más formát ölt. Az 57. egyenletben szereplő λ -nak nincs fizikai megalapozása, ezért itt sem kellene használni. Ezt eredetileg az OpenGL használta arra, hogy a számítások numerikus stabilitását javítsa akkor, ha a fényforrás nagyon közel van a megvilágított felülethez.

A fotometrikus sztereó általában feltételezi a felületek diffúz voltát. Hogyan lehetne továbblépni csillanó (spekuláris) vagy tükörszerű felületek esetére?

A III.2. tézis a kötegelt behangolás alkalmazása irányított pontok rekonstrukciójára, affin transzformációk alkalmazásával, amelynek során Structure from Motion módszerek pontosíthatók.

A IV. téziscsoport LiDAR és perspektív kamerák közös kalibrációjával foglalkozik. A LiDAR az elmúlt években vált a gépi látás és önvezető autók egyik kedvenc megoldásává. A LiDAR mélységadatokat szolgáltat a mért pontokról, amelyek sűrűsége a távolsággal csökken. Ezt a hiányosságot lehet kompenzálni normál kamerák felhasználásával, ami pedig feltételezi, hogy a LiDAR és a kamera „kalibrált” legyen, azaz a mért információ közös koordináta-rendszerbe hozható.

A IV.1. tézis a kamera-LiDAR kalibrációt téglatest (doboz), azaz annak 7 látható sarkának felismerésével végzi el, amit eredeti ötletnek tartok. Az első lépés a II.4. tézisre emlékeztet, de itt RANSAC-ot alkalmazza és nincsenek affin transzformációk. A megoldás ismét egy nemlineáris hibafüggvény megfogalmazása és minimalizálása. A IV.2. tézis lényegében ugyanazt a feladatot oldja meg gömböt feltételezve. A 71. egyenlet az ismeretlen gömbegyütthetőkra vonatkozó nemlineáris kényszereket, lényegében egy változcserével, lineárisaként adja meg, ami szellemes megoldás. Megjegyzem, hogy az a tény, hogy a gömb középpontos vetülete ellipszis, közismert. A IV.3. tézis az előző tézis feladatának egy speciális esete, amikor egy forgatási paraméterre redukált a kamera-LiDAR kalibráció.

A disszertáció magyar nyelven íródott, stílusa jó, nagyon kevés elütést tartalmaz. Az ábrákat lehetett volna még javítani, szebb volna, ha az ábrán szereplő feliratok betűnagysága közel állandó és a szöveghez hasonló lenne.

A tézisek kimondása eltér a megszokottól, amikor is a tudományos eredmény elérését első szám, első személyben fogalmazzák meg. Világos, hogy a kutatás csapatomunka, de a tudományos fokozat személyre szóló, így talán jobb lett volna csak a nagyon szorosan személyhez köthető eredményeket kiemelni. A dolgozat bőven tartalmaz értékes és új felismeréseket, így annak súlya nem csökkent volna a rövidítéstől.

A dolgozat 9 folyóirat és 31 fontosabb konferencia cikket sorol fel, amelyek a tézisekhez kapcsolódnak, ezekben a Jelölt jellemzően utolsó szerző, első vagy pedig egyetlen szerző. A folyóiratok között a téma legrangosabb folyóiratai is megjelennek (IEEE Transactions on Image Processing).

Összefoglalva, a Jelölt nagyon nagyszámú értékes és előremutató tudományos eredményt foglalt össze az MTA doktori disszertációjában. A kapcsolódó publikációk bizonyítják, hogy az eredmények a nemzetközi tudósközösség szigorú szűrőjén is átmentek. A téziseket elfogadom. A nyilvános vita kitérését javaslom a fokozat odaítélését támogatom.

Budapest, 2025. október 1.

.....
Dr. Szirmay-Kalos László
MTA levelező tagja