

# Válaszok Dr. Szirmay-Kalos László bírálataira

## 1. Általános köszönetnyilvánítás

Válaszaim elején mindhárom bírálónak közösen szeretném megköszönni az alapos munkájukat. Miután a számítógépes látás területe hatalmas, az én saját témáimmal rajtam és tanítványaimon kívül nem sokan foglalkoznak, pláne nem MTA doktora címmel rendelkező tudósok. Ennek ellenére a bírálatok nemcsak arról tanúskodnak, hogy megértették a meglehetősen tömény leírást, hanem mindhárman adtak olyan ötleteket, amelyek további vizsgálatokra sarkalnak, érdekes kérdéseket vetnek fel, amelyre érdemes lenne megtalálni a választ.

A tisztelt Bírálók motiváló megjegyzéseinek kézzelfogható eredménye is volt, mivel az Arxiv preprint szerverre két munkát is feltettem [Haj26a, Haj26b], melyeket később hivatalos fórumokra is szeretnék beküldeni.

## 2. Válaszok a kérdésekre/megjegyzésekre

**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.**

A magasabb rendű deriváltak alkalmazása kézenfekvő ötletnek tűnik, jómagam először a Kató Zoltán - Molnár József kutatópárostól hallottam először, hogy érdemes volna foglalkozni vele. Viszont praktikus okokból egyelőre még nem vettem napirendre: már az affin paraméterek pontos kinyerése is nehéz, hiszen az olyan alap-algoritmusok, mint a SIFT [Low04], csak durván képesek kinyerni. Például a forgatásra 36 darab “vödrot” (angolul: bin), használ, emiatt a kvantálás miatt  $360/36=10$  fok szerint növekvő értékeket lehet alkalmazni.

Saját magam egyébként először 2025-ben fejlesztettem ki affin kinyerő algoritmust, aminek a lépéseit lentebb részletezem is. Egyelőre van annyi gond vele, hogy a magasabb rendű tagok kinyerése reménytelennek látszik.

**Az, hogy egy affin transzformáció mindig felbontható forgatás, skálázás és eltolás transzformációkra, ismert tény.**

A függelékben az affin transzformáció két lehetséges felbontását adtam meg, mert ezek segíthetnek a transzformáció értelmezésében. Nem szerettem volna állítani, hogy ezt senki nem csinálta meg előttem, sajnálom, ha félreérthető voltam.

Illetve itt bátorokodnék pontosítani, mert a nyírás kimaradt a Bíráló felsorolásából. Ahogyan azt az F függelékben leírom, a forgatásból egy, a két skálából további kettő, az eltolásból szintén kettő szabadságfok adódik, ami összesen öt. A hatodik szabadságfokot a nyírás adja meg.

Az F függelékben bemutatott másik felbontás szerint két forgatás, két skála és a kétdimenziós eltolás adják ki a hat szabadságfokot, tehát igaza van a Bírálónak, hogy a nyírás nem feltétlenül

szükséges.

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 kiderülhetnek 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.

**Affin paraméterek kinyerése** A bíráló jól sejti, hogy az affin paraméterek kinyerése nem triviális feladat. Több módszer is létezik, az évek alatt kollégáimmal mi is sokat próbáltunk ki.

- Jómagam egyébként előszeretettel alkalmazom az ősrégi Lucas-Kanade algoritmust [LK81], mert mintaillesztésre kiválóan alkalmasak, cserébe a futási idejük sajnos jelentősnek mondható, még a mai gépeken sem lehet valós idejű feladatokhoz alkalmazni. Viszont tesztelési célra kiváló. A módszer iteratív, és kezdeti paramétereket igényel.
- A szintén idősebb SIFT jellegzetes-pont detektáló algoritmus [Low04] már maga is eredményez forgatást és skálázást. Párosítás (matching) után a kettőből a minták közötti relatív skála és forgatás meghatározható. Hátránya a módszernek, hogy a nyírás nem nyerhető ki. Viszont a kinyert transzformációkat az előző bekezdésben említett Lucas-Kanade algoritmusnak [LK81] lehet átadni kezdeti értéként.
- Létezik egy variációja is a SIFT algoritmusnak, az ASIFT [MY09], amikor cél a nagyobb perspektív korrekciót szoftveresen csökkenteni. Ezért a képeket eltorzítják affin transzformációval, a perspektívaváltást közelítve, és a torzított képekre futtatják a SIFT algoritmust. Az eredeti képrészek közötti affin transzformáció így bonyolultabb lesz, de lehet nyírást is bevinni a rendszerbe.
- Manapság a gépi tanulást szinte minden képfeldolgozási feladathoz szokás alkalmazni, így az affin transzformáció kinyeréséhez is [SGY+25].

Végezetül beletennék egy eredményt, amely az 1. ábrán látható. Az egyetemünk közelében levő Infoparkban készült két képet tettem össze. Az affin megfeleltetéseket a SIFT+Lucas-Kanade módszerrel készítettem el. A képre nagyítva látszanak az egymásnak megfeleltetett négyszögek, affin transzformációval módosítva.

Ez egy jelenleg publikálatlan eredmény, de könnyen belátható, hogy amennyiben a kameráink belső paraméterei ismertek, akkor a két kép között csak a forgatás ismertelen, mert a panorámaképek esetén az eltolás zérus, más szóval ugyanott vannak a fókuszpontok. A térbeli forgatás három paraméterére a mintaközéppontok két egyenletet adnak, a harmadik ismeretlen paraméter pedig számítható az affin transzformációból, sőt, túl is lesz a feladat határozva.

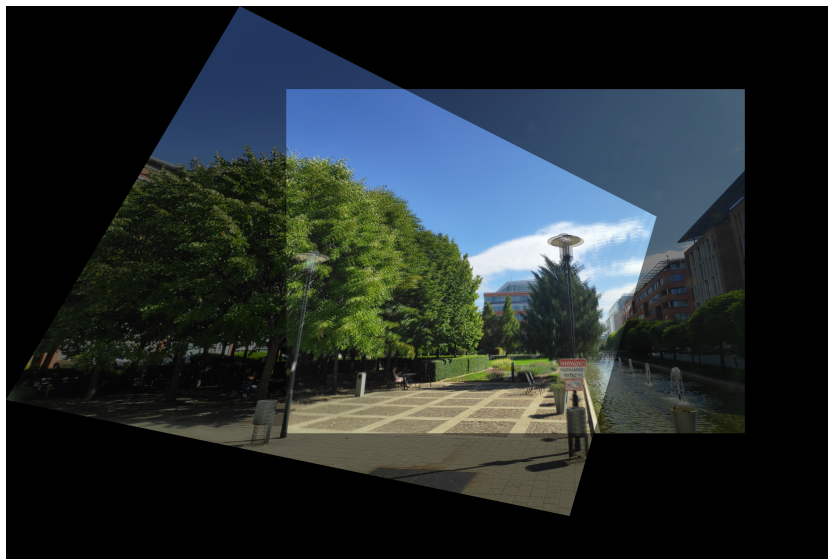
Erre az elvre építettem egy villámgyors robusztus becslőt. A RANSAC algoritmus elvét annyiban követtem, hogy egy minimális modell előállítható egy affin megfeleltetés segítségével, a robusztus szűrés ennek alapján elvégezhető. Néhány affin paraméter segítségével meg lehet kapni a jó transzformációt, de nagy megbízhatóság esetén az összes transzformációt is ellenőrizni lehet, értelemszerűen ekkor kicsit lassabb a módszer.

A kapott eredmény az összeillesztésre a 2. ábrán látható. Ugyan nem kapcsolódik ez az eredmény szorosan a disszertációhoz, de talán képes segíteni, hogyan lehet elképzelni az illesztést, illetve egy gyakorlati eredményt is meg lehet tekinteni.

**Normálvektorokra analitikus megoldás.** Itt vitába szállnék a bírálóval, miszerint nem volt analitikus megoldás. Ezt már Molnár Józseffel közös cikkünkben megmutattuk [TMH+14], hogy keresztszorzat segítségével hogyan lehet zárt alakú megoldást adni.



1. ábra. 100 darab SIFT [Low04] jellegzetes pont. A bal oldalon levő képeken zöld négyszögek a Lukas-Kanade [BM04] algoritmussal finomítva adják a jobb oldali képeken a piros négyzeteket. (Érdeemes a képekre ránagyítani.) Alaposabb vizsgálat után látható, hogy az algoritmus veszít pontokat, például a közeli lámpaoszlopokon. Ezért a becsléseket mindenképpen robusztifikálni kell.



2. ábra. Két kép összeillesztése panorámaképekhez alkalmazott, saját algoritmussal. A publikálatlan módszer kalibrált kamerák esetén képes egyetlen affin megfeleltetésből az illesztés homográfiáját kiszámítani.

Viszont ez a feladat nem optimális, mert négy affin paramétert (adatot, rajtuk keresztül egyenleteket) használ a normálvektor becslésére, pedig a normálvektor szabadságfoka csak kettő. Ezért nem analitikus megoldást kell adni, hanem egy becslést. Viszont magára a becslésre, legkisebb négyzetes hibát minimalizálva, megadtam analitikus megoldást. a II.1-es tézishez kapcsolódva.

...(II.2. tézishez kapcsolódóan) ... **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.**

Saját, őszinte véleményem szerint ez a tézis áll a leggyengébb lábakon a gyakorlati alkalmazhatóság szempontjából, mivel a kamera kalibráció a felületi normálvektorok ismeretében számítható ki. Ez pedig a gyakorlatban ritkán igaz, hacsaknem nagyon szabályos kalibrációs objektumot használunk.

Ugyanakkor, fontosnak a kimondását, hogy lehet kalibrációs eljárásokat készíteni affin paraméterek felhasználásával, és ez az elv megágyaz a III.2. tézispontnak, ahol a kötegelt behangolás segítségével lehet irányított pontfelhőt előállítani úgy, hogy közben a kamera paraméterei is újrarahangoljuk, azaz valamifajta kalibrációs eljárást végzünk.

**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 tisztelt Bírálónak teljesen igaza van, hogy az én mintavételezésem nem egyenletes, ezért megjegyzése után készítettem egy új tesztet, ahol három különböző mintavételezést valósítottam meg:

- **Hajder-pontok.** Az eredeti dolgozatban említett módon, gömbi koordináták alapján készítettem térbeli vektorokat:

$$\mathbf{p} = \begin{bmatrix} \cos \alpha \cos \beta \\ \cos \alpha \sin \beta \\ \sin \alpha \end{bmatrix},$$

ahol a két szöveget egyenletesen mintavételeztem az értelmezési tartományon. Mindez két 'for' ciklus segítségével, majd ezekből a pontokból véletlenszerűen választottam ki  $N$  darabot.

- **Random-pontok.** Nulla várható értékű 100.0 szórású Gauss-eloszlással készítettem három koordinátát, majd abból egy 3D vektort, aminek normalizáltam a hosszát, hogy egységvektor legyen. Így készítettem  $N$  darab normálirányt. Ezt a módszert naivnak is fogom nevezni a későbbiekben, hiszen egyszerű megfontolásra épül.
- **Egyenletes pontok.**  $N$  darab pont egyenletes mintavételezésére a Thomson-módszert [Tho04] használtam, ami egy fizikai szimuláción alapszik<sup>1</sup>, a pontok taszítják egymást, és megkeressük a nyugalmi állapotot, amikor a pontok távolsága maximális.

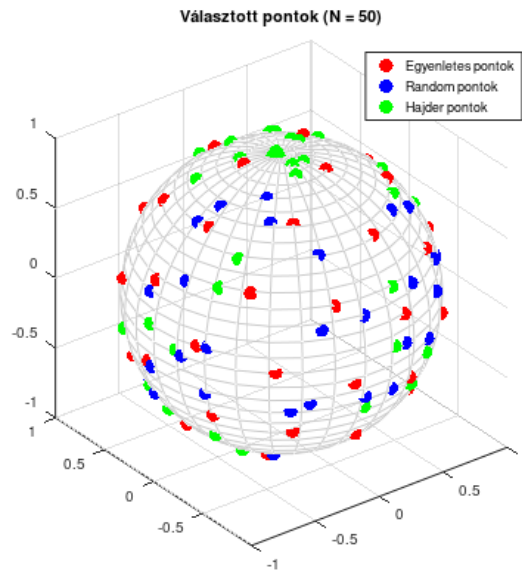
Egy  $N = 50$  vektort tartalmazó mintavételt a 2. ábrán láthatunk. A pirossal jelzett egyenletes mintavételezés szembeötlő, a másik két mintavételezésben nagyobb felületek maradnak üresen a véletlen törvényszerűségei alapján.

Az eredményül kapott normálvektorokat a legújabb és leggyorsabb, a disszertációból már kimaradt optimális normálvektor becslőnkre [HLB23] futtattam<sup>2</sup>, és a kapott normálvektorokat az alapigazsággal (ground truth) összevetve szöghibát lehet meghatározni. A pontokat és az affin transzformációkat is zajosítottam. Az eredmények a 2. képen láthatóak.

A szimuláció igyekezett valós méretű (megapixel tartomány) képre vetíteni a gömböt úgy, hogy kitöltse a teret. A hibákat fokban adtam meg, nulla várható értékű normális eloszlás adta a zajt, adott  $N$  zajszint esetén a a pixel koordinátákra  $N/10$ , az affin paraméterekre  $N/200$  szórást alkalmazva.

<sup>1</sup>Az implementációhoz mesterséges intelligenciát alkalmaztam: A MATLAB szkriptet a Gemini Pro 3 állította elő.

<sup>2</sup>Azt igazoltuk, hogy a legújabb módszer minden esetben ugyanazt adja, mint a disszertációban is ismertetett lassabb, de szintén optimális elődei.



3. ábra. A gömbfelületen háromféle mintavételezést próbáltam ki. A "Hajder pontok" a doktori téziszben említett módszernek felel meg, a 'random pontok' naiv számítás, míg az 'egyenletes pontok' esetén a pontok távolsága azonos. Szemmel is jól látszik, hogy jelentős különbség van a mintavételezésekben.

Összesen 50 mintavételi pontot vettem a gömbfelületen. A normálvektorokra az algoritmussal [HLB23] kiszámítottam, a kapott irány eltérését az alapigazságtól vettem hibaértéknek. Az ötven darab hibának az átlagát ( 2. ábra bal oldali grafikon) és a mediánját (jobb oldali grafikon) is kiszámítottam.

Az ábrákon jól látszik, hogy a bírálónak igaza volt, és ezt a kérdés valóban érdemes forszírozni, mert a három véletlenszerű kiválasztás tendenciózusan különböző értéket ad. A disszertáció során alkalmazott módszer egyértelműen a legkisebb hibát adja, de ez sajnos nem azt jelenti, hogy ez a legjobb, hiszen a teljesen egyenletes mintavételezés az ideális, hanem éppen azt mutatja, hogy számít a mintavételezés az eredmény szempontjából.

A fő ok, hogy extrém eseteket a saját mintavételezési módszer ritkábban ad, amikor valamelyik affín paraméter közel van a zérushoz. Ugyanakkor az is igaz, hogy valós feladatnál a két kép közötti mintákat akkor lehet megfeleltetni, és affín transzformációt kinyerni, ha mindkét képen jól látszanak, és ebben az esetben az affín transzformációk kevésbé vesznek fel szélsőségesnek tekinthető értékeket.

Érdekes még megemlíteni, hogy a medián hiba észrevehetően alacsonyabb az átlagos hibánál, ami azt sugallja, hogy szélsőséges esetek valóban előfordulnak a módszerekben.

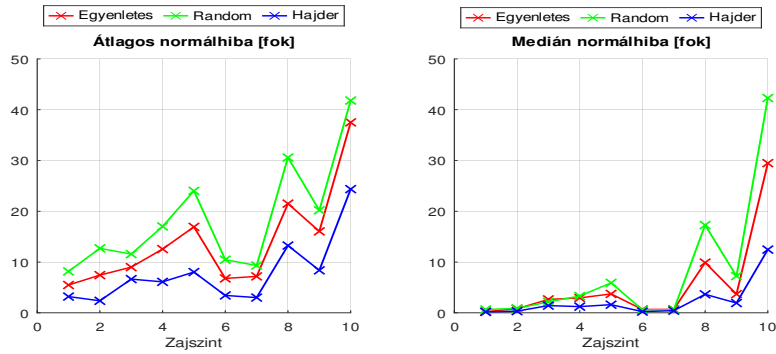
Ez a teszt rávilágít arra, hogy vannak ezen a területen is nyitott kérdések, helye lenne egy alaposabb vizsgálatnak a jövőben. Azt is szomorúan megállapíthatom, hogy valóban nem jártam el kellő alaposággal a korábbi teszteléseknél.

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

Ahogy a tisztelt Bíráló írja, a homográfia síkfelületre, konkrétan a látott felület érintősíkja-ra vonatkozik.

A dualitást a geometriában kétféleképpen is lehet értelmezni:

- Három dimenzióban a síkok és a pontok dualitását. Ennek következménye, hogy három pont meghatároz egy síkot, vagy három sík metszete meghatároz egy pontot.
- Két dimenzióban (valószínűleg erre gondolt a Bíráló) a pontok és az egyenesek duálisak. Itt két



4. ábra. Becslési hiba normálvektorok pontosságára. Bal oldalon az átlagos, jobb oldalon a medián hibát mutatják a grafikonok fokban kifejezve. Az előző képen mutatott három mintavételezőt alkalmaztam. Sajnos jól látszik, hogy következetesen eltérés van a mintavételezések eredménye között.

pont határoz meg egy egyenest, vagy két egyenes metszete a pontot.

Az közismert tény [HZ03], hogy ha két egymásnak megfelelő pontot a két képen egy  $\mathbf{H}$  homográfia köt össze:  $\mathbf{x}_2 = \mathbf{H}\mathbf{x}_1$ , akkor egymásnak megfelelő  $\mathbf{l}_1$  és  $\mathbf{l}_2$  egyenesekre pedig az alábbi összefüggés igaz:

$$\mathbf{l}_2 = \mathbf{H}^{-T}\mathbf{l}_1.$$

Mivel az egyeneseket reprezentáló  $\mathbf{l}_1$  és  $\mathbf{l}_2$  vektorok háromdimenziósak, de a skálájuk nem tartalmaz információt, két szabadságfokot lehet minden egyes egyenesből meghatározni.

Miután az affin transzformációk képesek a pontmegfelelések körüli irányokat transzformálni, ahogyan azt a II.9-es tétel is taglalja, a  $2 \times 2$ -es  $\mathbf{A}$  affin transzformációt úgy is fel lehetne használni a homográfia becslésére, hogy két egymásnak megfelelő egyenes határozzuk meg, és behelyettesítve kiszámoljuk a homográfia inverz transzponáltját, ha megfelelő számú egyenes-párral rendelkezünk.

Ez is egy érdekes kutatási irány lehet, mert ha így készítünk becslést, a II.3. tételben ismertetett algoritmusban használt módszert kellene megkapnunk. Ugyanakkor az algebrailag nehézséget jelent, hogy a pontmegfeleltetésből a homográfira, az egyenesekből pedig annak inverz transzponáltjára lehet egyenleteket felírni. A válasz elküldéséig még ne sikerült belátni, hogy ez a sejtés igaz, de a jövőben mindenképpen szeretnék vele foglalkozni. Egyúttal megköszönöm a Bírálónak, hogy felvetette a dualitás fogalmát, ezzel is egy érdekes elméleti kutatási irányra mutatott rá.

(Egy zárójeles bekezdés erejéig megemlíteném, hogy mostanában ellipszis illesztéssel is foglalkozunk, ahol a legelterjedtebb módszerben [FPF96] pontokra illesztnek ellipszist. Ugyanakkor a duális meghatározása lehetséges kizárólag érintő egyenesek segítségével [OH09] is. Kollégáimmal kidolgoztunk egy olyan algoritmust, ami pontokat, normálirányokat, sőt görbületeket is képes kinyerni a képekből, és azt ellipszis illesztésre felhasználni. Referenciát nem tudok adni, mert jelenleg elbírálás alatt van a munka.)

**III.1. tételben 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 érdekséget (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 I 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 lambdának 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.**

Sajnos jogosak a kritikák, az alkalmazott jelölések valóban félreérthetőek, elnézést kérek értük! Ez a terület csak egy esettanulmánya annak, hogy a kötegelt behangolás módszerét más problémára is sikerrel lehet alkalmazni. Több, mint tíz éve nem foglalkoztam a területekkel, amikor a doktori munka beadása előtt az eredeti angol szöveget fordítottam, a témával kapcsolatos ismereteim már eléggé megkoptak, sok mindent fel kellett frissíteni.

A megjegyzésekre térve:

A  $\mathbf{b}$  vektor valóban a normálvektort jelöli, pontosabban az azzal párhuzamos vektort, mert skálázva van az albedóval. Mi a fények viselkedésének hullámhossztól való függését egyáltalán nem vizsgáltuk, bár ez egyszerűsítése a tényleges fizikai törvényszerűségnek. A felületeket érdekességét normáltérképpel (melyet grafikában bump mapnek neveznek) ábrázoltuk. Az eltolási/elmozdulási térkép (displacement map) használata sokkal nehezebb eljárást eredményezett volna, hiszen a felületre merőlege irány és a beeső fény irányának viszonyától függ az adott fény iránya, a normálvektor így az elmozdulási képen a szomszédok segítségével számítható ki, ami további matematikai műveleteket vitt volna be az optimalizálásba.

A  $\lambda$  alkalmazását annyiban védeném meg, hogy numerikus módszereknél a szingularitás (nullával való osztás) elkerülése szintén fontos szempont. Aki például forgatások ábrázolásával és numerikus módszerekben alkalmazásával foglalkozott, az tudja, hogy Euler szögekkel leírni a forgatás három paraméterét numerikusan nem olyan kedvező, mint egységkvaternióval. Valami hasonlót lehet itt is elképzelni.

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

A kötegelt behangolás egy szofisztikált numerikus optimalizálás, amihez az elsőrendű deriváltakat kell kiszámolni. A hibafüggvényben az intenzitás eltérés nézzük, az intenzitás értéke a csillogó felületek szintén megadhatóak analitikus formában (pl. Phong-moddal), a hiba deriváltjai ezért kiszámíthatóak, és emiatt a numerikus módszerekkel lehet számolni.

Azonban például a fent megemlített Phong modellnél is az a probléma felmerül, hogy az okozott intenzitás változása nagy lehet, ha a csillanás (tükröződés) irányához közel van a paraméter. Ez szerencsétlen esetben a paraméterterben lokális optimumot okozhat. A kötegelt behangolást sokat vizsgálták, és a hagyományos (3D rekonstrukciós) feladatnál is sok esetben előfordulnak lokális optimumok, feltételezhetően a becsillanás kezelése esetén sokkal nagyobb lenne.

Ráadásul a kötegelt behangolás, mint minden numerikus algoritmus, egy jó kezdeti paraméterezést igényel, aminek előállítása csillogó felületek esetén szintén nehézkes.

**A tézisek kimondása eltér a megszokottól, amikor is a tudományos eredmény elérését első szám, első személyben fogalmazzák meg. Világos, hogy a kutatás csapatmunka, 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.**

Nagyon szerencsésnek érzem magamat abból a szempontból, hogy annak ellenére, hogy az utóbbi két évtizedben mindig dolgozhattam hallgatókkal (fennköltebben: tanítványokkal), mégis tudtam/tudok időt szakítani saját, egyszemélyes kutatómunkára is. A kidolgozott algoritmusok sokszor úgy készültek, hogy levezetéseket magam is végeztem, vagy akár közösen a doktorandusszal. A dolgozatba csak olyan munkát tettem bele, amelyekben érdemi hozzájárulásom van. Azt alapesetben el tudom mondani, hogy ahol én vagyok az első szerző, ott a hallgatók maximum a tesztelésben segítettek. Az is megjegyezném ezzel kapcsolatban, hogy gyengén perspektív kameramodellekkel kapcsolatban készítettem teljesen saját munkákat [Haj17a, Haj17b] is, azonban ezeknek a témája annyira eltér a többi munkától, hogy helyhiány és a dolgozat koherenciája miatt nem tudtam/akartam betenni.

Érdekesképpen leírnám, hogy legfontosabb egyéni eredményemnek az I.2 tételben kimondott összefüggést tartom, ez saját levezetés eredménye. A két oldal között először egy  $\beta$  [BMH16a] paraméter adott kapcsolatot, ezt Baráth Dániel jelezte a tesztek során, hogy  $-1$ -nek kell választani. Ezt sikerült [BTH17] szintén nekem bebizonyítani. Az ebből készült megoldó algoritmus viszont már

Baráth Dániel érdeme.

Például vannak vegyes tézisek is, a HA algoritmus (II.3. tézis) társítását a 3PT algoritmus-sal például én dolgoztam ki. Erre triviálisan épül a multi-model szegmentáló MultiH algoritmus a (II.4. tézis), bár erre cseh kollégánk, Jiri Matas mutatott rá, ezért is került bele a 2016-os publikációnkba [BMH16b] mint szerző. Ez a kis tézis-történelem azt mutatja, hogyan tudják motiválni egymást kutatók, miért olyan hasznos a saját "elefántcsonttorony"-ból kilépni, és másokkal együttműködni.

Kelt: Budapest, 2026. március 12.

---

Hajder Levente

## Hivatkozások

- [BM04] Simon Baker and Iain A. Matthews. Lucas-kanade 20 years on: A unifying framework. *Int. J. Comput. Vis.*, 56(3):221–255, 2004.
- [BMH16a] Daniel Barath, Jiri Matas, and Levente Hajder. Accurate closed-form estimation of local affine transformations consistent with the epipolar geometry. In Richard C. Wilson, Edwin R. Hancock, and William A. P. Smith, editors, *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016*. BMVA Press, 2016.
- [BMH16b] Daniel Barath, Jiri Matas, and Levente Hajder. Multi-h: Efficient recovery of tangent planes in stereo images. In Richard C. Wilson, Edwin R. Hancock, and William A. P. Smith, editors, *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016*. BMVA Press, 2016.
- [BTH17] Daniel Barath, Tekla Toth, and Levente Hajder. A minimal solution for two-view focal-length estimation using two affine correspondences. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 2557–2565. IEEE Computer Society, 2017.
- [FPF96] Andrew W Fitzgibbon, Maurizio Pilu, and Robert B Fisher. Direct least squares fitting of ellipses. In *Proceedings of 13th international conference on pattern recognition*, volume 1, pages 253–257. IEEE, 1996.
- [Haj17a] Levente Hajder. W-pnp method: Optimal solution for the weak-perspective n-point problem and its application to structure from motion. In Francisco H. Imai, Alain Trémeau, and José Braz, editors, *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017) - Volume 6: VISAPP, Porto, Portugal, February 27 - March 1, 2017*, pages 265–276. SciTePress, 2017.
- [Haj17b] Levente Hajder. Weak-perspective and scaled-orthographic structure from motion with missing data. In Ana Paula Cláudio, Dominique Bechmann, Paul Richard, Takehiko Yamaguchi, Lars Linsen, Alexandru C. Telea, Francisco H. Imai, and Alain Trémeau, editors, *Computer Vision, Imaging and Computer Graphics - Theory and Applications - 12th International Joint Conference, VISIGRAPP 2017, Porto, Portugal, February 27 - March 1, 2017, Revised Selected Papers*, volume 983 of *Communications in Computer and Information Science*, pages 128–153. Springer, 2017.
- [Haj26a] Levente Hajder. Affine correspondences in stereo vision: Theory, practice, and limitations, 2026. <https://arxiv.org/abs/2603.01836>.

- [Haj26b] Levente Hajder. A unified formula for affine transformations between calibrated cameras, 2026. <https://arxiv.org/abs/2602.06805>.
- [HLB23] Levente Hajder, Lajos Lóczi, and Daniel Barath. Fast globally optimal surface normal from an affine correspondence. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pages 3367–3378. IEEE, 2023.
- [HZ03] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [LK81] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In Patrick J. Hayes, editor, *Proceedings of the 7th International Joint Conference on Artificial Intelligence, IJCAI '81, Vancouver, BC, Canada, August 24-28, 1981*, pages 674–679, 1981.
- [Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60(2):91–110, 2004.
- [MY09] Jean-Michel Morel and Guoshen Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM J. Imaging Sci.*, 2(2):438–469, 2009.
- [OH09] Jean-Nicolas Ouellet and Patrick Hébert. Precise ellipse estimation without contour point extraction. *Mach. Vis. Appl.*, 21(1):59–67, 2009.
- [SGY<sup>+</sup>25] Pengju Sun, Banglei Guan, Zhenbao Yu, Yang Shang, Qifeng Yu, and Daniel Barath. Learning affine correspondences by integrating geometric constraints. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pages 27038–27048. Computer Vision Foundation / IEEE, 2025.
- [Tho04] J.J. Thomson. Xxiv. on the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 7(39):237–265, 1904.
- [TMH<sup>+</sup>14] Attila Tanács, Andras Majdik, Levente Hajder, József Molnár, Zsolt Sánta, and Zoltan Kato. Collaborative mobile 3d reconstruction of urban scenes. In *Computer Vision - ACCV 2014 Workshops - Singapore, Singapore, November 1-2, 2014, Revised Selected Papers, Part III*, volume 9010 of *Lecture Notes in Computer Science*, pages 486–501. Springer, 2014.