

Szegedi Tudományegyetem
Informatikai Intézet
SZTE Informatika Doktori Iskola

Régió alapú pose és homográfia becslés centrális kamerákkal

Ph.D. értekezés tézisei

Fröhlich Róbert

Témavezető:

Prof. Kató Zoltán

Külső konzulens:

Dr. Tamás Levente

Szeged

2019

Bevezetés

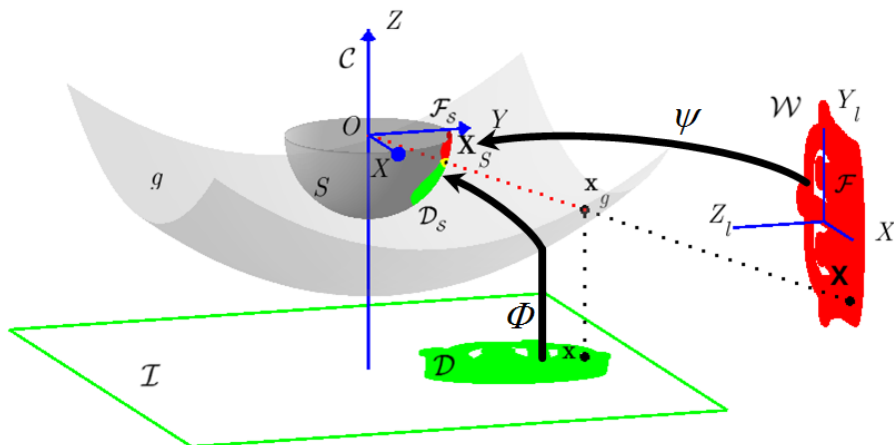
A számítógépes látás az a tudományos terület, amely által a számítógépek egy digitális kép vagy kép-szekvencia alapján magasabb rendű információra tehetnek szert. Gyakorlatilag azon műveletek automatizálása a cél, melyekre az emberi látószerv is képes. Fő feladatai közé tartozik a digitális képek rögzítése, feldolgozása, elemzése, értelmezése és információk kinyerése a valós világról. Különbféle alkalmazások léteznek már, melyek olyan alaposan kutatott témakörökre bonthatóak, mint például a 3D rekonstrukció, esemény detekció, videó követés, felismerés, 3D pose becslés, tanulás, szegmentálás, mozgásbecslés vagy a képhelyreállítás.

A disszertáció összefoglalja a Szerző kutatási eredményeit az abszolút pose becslés, síkhomográfia becslés és síkrekonstrukció területéről.

Régió alapú pose becslés

Ebben a szekcióban egy általános, nem-lineáris, explicit megfeleltetések nélkül működő pose becslő módszert javasunk. Az abszolút kamera pose becslés egy közös Lidar-kamera sík folt regisztrációja alapján történik. A javasolt módszer minimális információt használ fel (csak mélység adat a 3D és radiometria információ a 2D oldalon) és kellően általános ahhoz, hogy perspektív vagy omnidirekcionális centrális kamerákkal is használható legyen. Módszerünket [1] 2D regisztrációs keretrendszere inspirálta, mely alapján [2] fogalmazta meg először az abszolút pose becslés egyenleteit perspektív kamerák számára, majd [Tamas, Frohlich, Kato, 2014] terjesztette ki omnidirekcionális kamerákra. Az általános egységesített keretrendszer, amit [Frohlich, Tamas, Kato, 2019] javasolt, egy rekurzív megoldást használ az egyenletrendszer integráljainak számolására, és azt a szub-optimális számolási módhoz hasonlítja amit [2] mutatott be. Továbbá [Frohlich, Tamas, Kato, 2019] kibővítette az omni kamerák szferikus reprezentációját, olyan módon, hogy az a perspektív kamerákat is kezelje, majd összehasonlította a kétféle formalizmust, és egy új numerikus módszert javasolt a felszíni integrálok számolására, amely gömbfelszíni háromszögeket használ a korábbi pont-alapú megközelítéshez [Tamas, Frohlich, Kato, 2014] képest.

Először fogalmazzuk meg az abszolút pose becslés feladatát centrális, szferikus ka-



1. ábra. A $\mathcal{D}_S$ és $\mathcal{F}_S$ felszíni foltok vetítése a szferikus kamera modellben.

merák számára. A [3, 4] által javasolt általános szferikus kamera modellt megfigyelve láthatjuk, hogy egy 3D világ $\mathbf{X} = [X_1, X_2, X_3]^\top \in \mathbb{R}^3$ pontjának vetülete a kamerában lényegében az $\mathcal{S}$ gömbfelszínre történő centrális vetítés által meghatározható, az $(\mathbf{R}, \mathbf{t})$ pose paramétereit figyelembe véve. Tehát egy 3D világ $\mathbf{X}$ pontja és annak $\mathbf{x} \in \mathcal{I}$ képe között a következő kapcsolat áll fenn az $\mathcal{S}$ gömbön keresztül [Tamas, Frohlich, Kato, 2014]:

$$\Phi(\mathbf{x}) = \mathbf{X}_{\mathcal{S}} = \Psi(\mathbf{X}) = \frac{\mathbf{R}\mathbf{X} + \mathbf{t}}{\|\mathbf{R}\mathbf{X} + \mathbf{t}\|}. \quad (1)$$

A klasszikus, pont-megfeleltetés alapú megközelítésben (1) megoldása az lenne, hogy 2D-3D pontmegfeleltetéseket keresünk például speciális kalibrációs tárgyak segítségével [5, 6] vagy képi jellemzők alapján, majd megoldjuk az egyenletet az abszolút pose $(\mathbf{R}, \mathbf{t})$ paramétereire valamilyen hibafüggvény minimalizálásával. Viszont több valós alkalmazásban is előfordul, hogy a kalibrációs minták használata nem kivitelezhető, és a legtöbb 3D adat (például Lidar szkennelével rögzített pontfelhők) csak mélységi információt tartalmaznak, ami komoly kihívás elé állítja a képi leírókat párosító algoritmusokat. Az ilyen, kihívást jelentő helyzetekre mutatunk be egy megoldást.

Szférikus kamerák abszolút helyzete

Szférikus kamerák esetében az egységgömb felszínén dolgozhatunk, mivel az a kamera belső paramétereitől független reprezentációt biztosít. Mivel nem állnak rendelkezésünkre pont-megfeleltetések, ezért (1) nem használható direkt módon, viszont az egyéni pont-párokat kiintegrálhatjuk, a következő egyenletet kapva [Tamas, Frohlich, Kato, 2014]:

$$\iint_{\mathcal{D}_{\mathcal{S}}} \mathbf{X}_{\mathcal{S}} d\mathcal{D}_{\mathcal{S}} = \iint_{\mathcal{F}_{\mathcal{S}}} \mathbf{Z}_{\mathcal{S}} d\mathcal{F}_{\mathcal{S}}, \quad (2)$$

ahol $\mathcal{D}_{\mathcal{S}}$ a kamera $\mathcal{I}$ képsíkján látható $\mathcal{D}$ régiónak megfelelő $\mathcal{S}$ gömbfelszíni foltot jelöli, míg $\mathcal{F}_{\mathcal{S}}$ az $\mathcal{F}$ 3D sík régiónak megfelelő folt, amit Ψ vetít az $\mathcal{S}$ gömbre (1) szerint (lásd 1. ábrán).

A $\mathcal{D}_{\mathcal{S}}$ és $\mathcal{F}_{\mathcal{S}}$ foltokat Φ és Ψ függvények által a $\mathcal{D}$ és $\mathcal{F}$ régiók felett paraméterezve felírhatjuk a felszíni integrál explicit képletét [Tamas, Frohlich, Kato, 2014]. Mivel egy $\mathcal{S}$ felszínen levő pontnak csak 2 független komponense van, ez a felírás csak egy 2 egyenletről álló rendszert eredményez. A pose-nak viszont 6 ismeretlen paramétere van, ezért újabb egyenleteket konstruálunk az [1] által bemutatott általános mechanizmussal. Az egyenlet mindkét oldalán egy $\omega : \mathbb{R}^3 \rightarrow \mathbb{R}$ függvényt alkalmazva, a következő explicit alakot kapjuk:

$$\iint_{\mathcal{D}} \omega(\Phi(\mathbf{x})) \left\| \frac{\partial \Phi}{\partial x_1} \times \frac{\partial \Phi}{\partial x_2} \right\| dx_1 dx_2 = \iint_{\mathcal{F}} \omega(\Psi(\mathbf{X})) \left\| \frac{\partial \Psi}{\partial X_1} \times \frac{\partial \Psi}{\partial X_2} \right\| dX_1 dX_2, \quad (3)$$

ahol a parciális deriváltak keresztszorzatának magnitúdója a felszíni elemet jelöli. Nem-lineáris $\{\omega_i\}_{i=1}^{\ell}$ függvények egy ℓ halmazát alkalmazva, minden függvény egy új független egyenletet generál. A pose $(\mathbf{R}, \mathbf{t})$ paramétereit a nem-lineáris (3) egyenletrendszert standard *Levenberg-Marquardt* algoritmussal megoldva kapjuk.

Bár tetszőleges ω_i függvényeket használhatnánk, a hatvány függvények számításigény alacsonyabb [1, 2], mivel ezeket akár rekurzív módon is lehet számolni:

$$\omega_i(\mathbf{X}_{\mathcal{S}}) = X_{S1}^{l_i} X_{S2}^{m_i} X_{S3}^{n_i}, \text{ ahol } 0 \leq l_i, m_i, n_i \leq 2 \text{ és } l_i + m_i + n_i \leq 3. \quad (4)$$

Vegyük észre, hogy (3) bal oldala konstans, tehát elegendő csak egyszer kiszámolni, viszont a jobb oldal tartalmazza az ismeretlen pose paramétereit, ezért minden iterációban újra kell számolni ami költséges lehet nagyobb régiók esetében. Ezért a pont alapú reprezentáció [Tamas, Frohlich, Kato, 2014] helyett használjuk az $\mathcal{F}$ 3D sík régió $\mathcal{F}^\Delta$ háromszögelt reprezentációját. Ennek köszönhetően csak az $\mathcal{F}^\Delta$ háromszögeinek $\{\mathbf{V}_i\}_{i=1}^V$ csúcsaira kell alkalmaznunk Ψ -t, ami így megadja a szférikus folt $\mathcal{F}_S^\Delta$ gömbfelszíni háromszögek által kifejezett változatát [Frohlich, Tamas, Kato, 2019]:

$$\forall i = 1, \dots, V : \quad \mathbf{V}_{S,i} = \Psi(\mathbf{V}_i). \quad (5)$$

Az $\mathcal{F}_S$ folt új reprezentációjának köszönhetően átírhatjuk (3) integrál jobb oldalát, az ω_i függvényeket is alkalmazva, a következő formára [Frohlich, Tamas, Kato, 2019]:

$$\iint_{\mathcal{D}} \Phi_1^{l_i}(\mathbf{x}) \Phi_2^{m_i}(\mathbf{x}) \Phi_3^{n_i}(\mathbf{x}) \left\| \frac{\partial \Phi}{\partial x_1} \times \frac{\partial \Phi}{\partial x_2} \right\| dx_1 dx_2 \approx \sum_{\forall \Delta \in \mathcal{F}_S^\Delta} \iint_{\Delta} Z_{S1}^{l_i} Z_{S2}^{m_i} Z_{S3}^{n_i} d\mathbf{Z}_S, \quad (6)$$

ahol $\Phi = [\Phi_1, \Phi_2, \Phi_3]^\top$ jelöli $\Phi : \mathcal{I} \rightarrow \mathcal{S}$ koordináta függvényeit. Tehát csak a háromszög csúcspontokat kell $\mathcal{S}$ -re vetíteni, és az integrál a felszíni háromszögek felett a [7] által bemutatott módon számolható. A pose paramétereit pedig a (6) egyenletrendszer legkisebb négyzetek szerinti megoldásával kapjuk.

Az optimális becslési eredmény érdekében fontos a numerikus normalizálás és a megfelelő inicializálás. [1]-el ellentétben, a fenti egyenletben minden pont már az egységgömbön található, tehát az adat-normalizálás implicit módon adott. A pose paraméterek jó inicializálása érdekében több lépést végzünk. Először a 3D adatot hozzávetőlegesen a kamerával szembe forgatjuk, ezzel biztosítva, hogy az a felszín jó oldalát látja [Frohlich, Tamas, Kato, 2019], majd egy eltolást alkalmazunk, ami az $\mathcal{F}^\Delta$ centroidját viszi $[0, 0, -1]^\top$ -be, ezáltal a régiót a $Z = -1$ síkba helyezve. Ez biztosítja, hogy a sík nem metszi $\mathcal{S}$ -et amíg a pose paramétereit inicializáljuk. Ha a kamera vertikális irányáról van egy nagyságrendi becslésünk, ami akár szenzorból is jöhet, vagy egyszerűen az adathalmazból megállapítható lehet, akkor a pontfelhő vertikális irányát a kamera X tengelyéhez igazítjuk, ez által biztosítva a helyes vertikális irányát a vetületnek. Minél kevésbé szimmetrikusak a régiók annál kevésbé van szükség a vertikális irány nagyságrendi inicializálására. Végül a pose paraméterek inicializálásával biztosítjuk, hogy a felszíni $\mathcal{D}_S$ és $\mathcal{F}_S^\Delta$ foltok jól átfedjenek [Frohlich, Tamas, Kato, 2019], azáltal, hogy $\mathcal{D}_S$ és $\mathcal{F}_S^\Delta$ centroidjai között számolt forgatással inicializáljuk $\mathbf{R}$ -et, míg $\mathbf{t}$ -t egy olyan eltolással, amellyel $\mathcal{F}_S^\Delta$ területe megközelítőleg azonos lesz $\mathcal{D}_S$ területével.

Kettő vagy több nem koplanáris régió esetében az algoritmus hasonlóan indul, először egy régióval becsül pose-t, majd azt iniciális értéként felhasználva az összes régióra felírt egyenletrendszert oldja meg.

Perspektív kamerák abszolút helyzete

Lévén, hogy a perspektív kamera is centrális, alkalmazható rá a [3, 4] által bemutatott szférikus kamera modell. Mivel kalibrált kameránk van, felírhatjuk egy 3D $\mathbf{X}$ pont vetítését normalizált képi koordinátákba $\mathbf{x} = [x_1, x_2]^\top \in \mathbb{R}^2$:

$$\mathbf{x} \leftarrow \mathbf{K}^{-1} \tilde{\mathbf{x}} \cong \mathbf{K}^{-1} \mathbf{P} \mathbf{X} = [\mathbf{R} | \mathbf{t}] \mathbf{X}. \quad (7)$$

Lényegében a szférikus modell g felszíne és a normalizált $\mathcal{I}$ kép között $g \equiv \mathcal{I}$ áll fenn, tehát a $\Phi : \mathcal{I} \rightarrow \mathcal{S}$ bijektív függvény perspektív kamera esetében nem más lesz, mint $\mathbf{x}$ egységvektora:

$$\mathbf{X}_S = \Phi(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}. \quad (8)$$

A fenti szférikus reprezentációt alkalmazva perspektív kamerákra, az előző szekcióban bemutatott teljes módszer változtatás nélkül használható. Viszont számítási komplexitását tekintve kedvezőbb a normalizált képsíkon maradni, mivel így egyszerűbb dupla integrálokat használhatunk az $\mathcal{I}$ képsíkon, az $\mathcal{S}$ gömbfelszín helyett. Így (7) egyéni pontpárjait integrálva, és nem-lineáris $\omega : \mathbb{R}^2 \rightarrow \mathbb{R}$ függvényeket alkalmazva az egyenlet mindkét oldalán a következő egyenletet kapjuk [2]:

$$\int_{\mathcal{D}} \omega(\mathbf{x}) d\mathbf{x} = \int_{[\mathbf{R}|\mathbf{t}]\mathcal{F}} \omega(\mathbf{z}) d\mathbf{z}, \quad (9)$$

ahol $\mathcal{D}$ jelöli a normalizált kamera képen látható régiót, és $[\mathbf{R}|\mathbf{t}]\mathcal{F}$ a neki megfelelő 3D sík régió normalizált kamera mátrix $[\mathbf{R}|\mathbf{t}]$ által vetített képet. Hatvány $\omega_i(\mathbf{x}) = x_1^{n_i} x_2^{m_i}$ függvényeket választva, és az $\mathcal{F}$ 3D régió $\mathcal{F}^\Delta$ háromszögelt reprezentációját, egy hatékony számolási módot alkalmazhatunk, mivel az így kapott egyenletek az $[\mathbf{R}|\mathbf{t}]\mathcal{F}$ levetített régiónak a 2D geometriai momentumait tartalmazzák. Így az integrál egyenlet is átírható $[\mathbf{R}|\mathbf{t}]\mathcal{F}^\Delta$ elemei szerint [2]:

$$\int_{\mathcal{D}} x_1^{n_i} x_2^{m_i} d\mathbf{x} = \int_{[\mathbf{R}|\mathbf{t}]\mathcal{F}} z_1^{n_i} z_2^{m_i} d\mathbf{z} \approx \sum_{\forall \Delta \in [\mathbf{R}|\mathbf{t}]\mathcal{F}^\Delta} \int_{\Delta} z_1^{n_i} z_2^{m_i} d\mathbf{z}. \quad (10)$$

A jobb oldali közelítés az $\mathcal{F}$ régió $\mathcal{F}^\Delta$ általi diszkrét háromszöghálós közelítése miatt szükséges. A háromszögek feletti integrálok viszont különböző geometriai momentumok, amit rekurzívan, a lennebb tárgyalt módon hatékonyan lehet számolni.

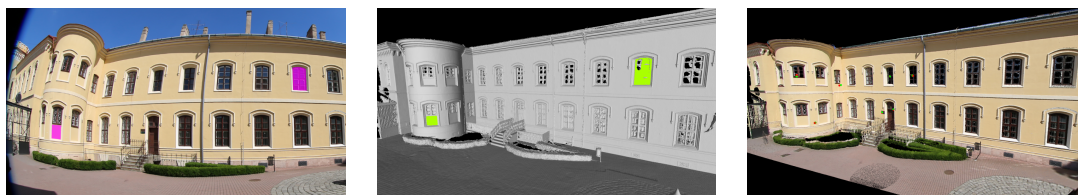
A pose paraméterek inicializálása a szférikus kamerához hasonló módon történik, azzal a különbséggel, hogy ez esetben az egységgömb helyett a normalizált képsíkon történnek a műveletek. Előbb egy Z tengely menti eltolást határozzunk meg, amely által a $\mathcal{D}$ képi régió és a levetített 3D régió nagyjából azonos méretű lesz, majd az $\mathbf{R}$ forgatást a két régió centroidja közötti forgatással inicializáljuk [Frohlich, Tamas, Kato, 2019].

2D geometriai momentumok számolása

A módszerünkben $\mathcal{D} \subset \mathcal{I}$ és $[\mathbf{R}|\mathbf{t}]\mathcal{F}^\Delta \subset \mathcal{I}$ 2D régiók felett kell integrált számolnunk, amire könnyedén adaptálhatjuk a 3D háromszögekre [8] által javasolt hatékony rekurzív formulákat. Mivel a normalizált $\mathcal{I}$ képsíkon $Z = 1$ -ben van, tekinthetjük úgy, hogy a vertex pontjaink Z koordinátája konstans 1, így az S felszín (i, j, k) geometriai momentumának általános 3D formulái [8] egyszerű 2D momentumokká vezethetők át a mi speciális 2D esetünkben [Frohlich, Tamas, Kato, 2019]:

$$M_{ijk} = \int_S x^i y^j z^k dS = \int_S x^i y^j dx dy, \quad (11)$$

ahol M_{ijk} utolsó tagja mindig 1 a k értékétől függetlenül. i és j egész számok, $i + j = N$ pedig a geometriai momentum rangja. Tehát a [Frohlich, Tamas, Kato, 2019] által javasolt egyenleteket használva hatékonyan kiszámolható minden háromszög kontribúciója a képi régiók minden geometriai momentumába.



2. ábra. Példa pose becslési eredmény: omni kép a baloldalon, sűrű Lidar pontfelhő közepén, a megfelelő szegmentált régiók magentával és zölddel jelölve. Jobbról pedig az eredmény: a becsült pose-al a képi információ a 3D pontfelhőre vetítve (markerekben mért előrevetítési hiba középértéke 7 cm).

A javasolt algoritmus MATLAB-ban lett implementálva és experimentálisan kiértékelve nagy méretű szintetikus perspektív és omnidirekcionális adathalmazokon. Szegmentálási zajokra 12% és 20%-ig bizonyult robusztusnak az omni és perspektív esetekben, 3 sík régiót használva. Kimutattuk, hogy a minimális 1 esethez képest, a régiók számának növelésével a módszer teljesítménye drasztikusan növekszik. A korábbi pont alapú reprezentáció helyett a felszíni háromszögekre való váltás egy nagyságrenddel csökkenti az algoritmus futásidejét. Különböző valós tesztesetek lettek vizsgálva, ritka és sűrű 3D lézeres pontfelhőn, katadioptrikus és halszem kamerákkal (például 2. ábrán egy omni teszt eset látható), perspektív DSLR kamerákkal, és drónos kamerával is. A közkezdelt publikus KITTI [9] adathalmazon pedig a javasolt módszer részben jobbnak bizonyult a kölcsönös információ-tartalom alapján működő módszerhez [10] képest, továbbá a CPU implementáció futásideje két nagyságrenddel kisebb a [10] GPU implementációjánál.

2D-3D vizuális adatok fúziója

Ebben a szekcióban két fúziós alkalmazást mutatunk be. Az első a [Frohlich, Tamas, Kato, 2019]-ben bemutatott perspektív régió alapú regisztrációs módszeren alapul, kiterjesztve azt nem-sík de kellően sima felszínekre, és egy ICP szerű lépést bevezetve a pose paraméterek finomítására 3D intenzitás adat felhasználásával [Frohlich *et al.*, 2016]. A második alkalmazás egy, a sok nézet használatakor felmerülő, kulcsfontosságú kérdésre, a 3D modell kiszínezéséhez megfelelő kamera kiválasztására fókuszál. Ezen kívül további technikai megoldást is javasol a fúziós eredmények textúrázott modellként történő vizualizációjára, nagy mennyiségű textúra képek esetén [Frohlich *et al.*, 2018].

Nem-sík régiók

Sok alkalmazás szerves részét képezi a kamera abszolút pose becslés, de sok esetben a sík régiók használata és ezáltal a régió alapú módszer alkalmazása nem lehetséges, egyszerűen a 3D adat adottságai miatt. Például a kulturális örökségvédelem területén egyre nagyobb az érdeklődés a fennmaradt hagyatékok, mint például tárgyi eszközök és az épített örökségek digitalizációjára. De legtöbbször ezek a tárgyi eszközök, romok vagy régi templomok finom megmunkálásuk miatt nem igazán rendelkeznek olyan sík felületekkel, mint amelyeneket például egy modern épített környezetben könnyedén találunk.

Tekintsük újra meg a (9)-ben bemutatott integrál egyenletet. Jól látható, hogy az egyenlet teljesül görbült, sima felszínekre is [Frohlich *et al.*, 2016], mindaddig, amíg öntakarás

nem történik, vagyis $\mathcal{D}$ és $\mathcal{F}$ régiók teljesítik azt, hogy:

$$\mathcal{D} = \mathbf{P}\mathcal{F}, \text{ ahol } \mathcal{D} = \cup_{i=1}^N \mathcal{D}_i \text{ és } \mathcal{F} = \cup_{j=1}^M \mathcal{F}_j, \quad (12)$$

$\{\mathcal{D}_i\}_{i=1}^N$ és $\{\mathcal{F}_j\}_{j=1}^M$ pedig tetszőleges elemszámú, de egymásnak összességében megfelelő 2D-3D régió halmazok.

ICP finomítás

Miután (10) algebrai hibájának minimalizálásával már becsültünk egy kamera pose-t, tovább pontosíthatjuk azt egy releváns geometriai hiba minimalizálásával. [Frohlich *et al.*, 2016]-ban bemutattuk, hogyan használható egy standard ICP (Iterative Closest Point [11]) algoritmus, ha rendelkezésünkre áll 3D szín információ. A javasolt munkafolyamatban ICP algoritmus segítségével igazítjuk a 3D élék vetületét a 2D kamera képről kinyert $\mathbf{x}_e$ élkép-hez. A 3D adatot előbb levetítjük az iniciális pose-zal, majd az így kapott képen éléket detektálunk, amelyek ilyen módon megfelelnek 3D $\mathbf{X}_e$ éléknek is. Az algoritmus aztán minden iterációban újra levetíti a 3D $\mathbf{X}_e$ pontokat az aktuális $\mathbf{K}[\mathbf{R}^n | \mathbf{t}^n]$ kamera mátrixsal, amelyben a kamera pose $(\mathbf{R}^n, \mathbf{t}^n)$ paraméterei változnak az iterációk között, így kapva a $\mathbf{z}_e^n$ vetített élpontokat az n -edik iterációban:

$$\mathbf{z}_e^n = \mathbf{K}[\mathbf{R}^n | \mathbf{t}^n] \mathbf{X}_e. \quad (13)$$

Tehát az ICP algoritmus a vetített $\mathbf{z}_e^n$ -t igazítja a 2D kép $\mathbf{x}_e$ éltérképéhez, lényegében a visszavetítési hibát minimalizálva ilyen módon.

Ha rendelkezésünkre áll egy becsült kamera pose, és ismertek a kamera kalibrációs paraméterei a 3D pontfelhő kiszínezhető a 2D képről. Ha viszont több 2D képünk van, akkor a több kamerából is látható 3D pontok esetében el kell döntenünk melyik kamera a megfelelőbb annak kiszínezésére. Ehhez elegendő minden $\mathbf{X}_i$ pontra kiszámolni annak $\mathbf{n}_i$ normálvektorának a kamerák $\mathbf{c}_j$ optikai tengelyének orientációjával bezárt szögét:

$$\cos \theta = \frac{\mathbf{c}_j \cdot \mathbf{n}_i}{\|\mathbf{c}_j\| \|\mathbf{n}_i\|}, \quad (14)$$

és azon j kamera képét kiválasztani amelyre $\cos \theta$ maximális [Frohlich *et al.*, 2016]. Így egy jó minőségű kiszínezett 3D modellt kapunk, ami kisebb kulturális örökségvédelmi tárgyak, kerámiák, edények esetében megfelelőnek bizonyult.

Szintetikus adatokon végzett kvantitatív kiértékeléssel kimutattuk, hogy a módszer görbült felszínekre is jól működik akár már a minimális 1 régiót használva is, és robusztus szegmentálási hibákkal szemben. Valós tesztek keretében többféle jó minőségű és kevésbé pontos 3D adattal rendelkező eset is lett vizsgálva, ami kimutatta mind a kézi 3D szkennert pontatlanságát, mind pedig a módszer robusztusságát az ilyen adatokra.

Kamera szelekció

Tekintsünk most azokra a kulturális örökségvédelmi alkalmazásokra, amelyek nagy méretű barlangok, romok vagy templomok dokumentálásával foglalkoznak. Mivel a régészek egyre gyakrabban használnak 3D szkennelési technikákat munkájuk során, ezért egyre nagyobb szükség van új megoldásokra. Bár a legtöbb kereskedelembe kapható eszköz megoldást kínál 2D kamerakép regisztrálására a 3D pontfelhőhöz, ezek azt feltételezik, hogy a kamera

és a szkennerek egy mereven összeszerelt egység, így nem kezelik a különböző nézőpontokból esetleg különböző időben rögzített képeket. A 3D modell fúziója így legtöbb esetben nem megoldható, ami felveti a megfelelő kamera szelekció kérdését. Egy teljes dokumentálási munkafolyamatot javasoltunk [Frohlich *et al.*, 2018], amely egy speciális algoritmus segítségével választja ki a megfelelő kamerát minden egyes 3D felszín elem számára.

Lényegében az algoritmus egy sor kritériumot vizsgál meg, mely során kiszűri a rossz kamerákat, és egy rangsorolt listát állít fel a jó kamerákból minden egyes vertex pont számára. Először megvizsgáljuk, hogy egy $\mathbf{X}$ pont látható-e egy adott kamerából vagy takarásban van. Erre a *Hidden Point Removal* operátort [12] alkalmaztuk, ami arra az észrevételre alapoz, hogy a pontfelhő egy kiválasztott nézőpont szerinti szférikus kifordítása után kapott konvex burkán található pontok a kiválasztott nézőpontból látható pontok.

Következő lépésként vizsgáljuk, hogy egy adott pont képe éles lehet-e a kamera képen, így csak azok a pontok színezhetőek ki egy kamerából, amelyek annak a mélységélességi tartományába esnek. Csupán a kamera képe alapján nem egyszerű megállapítani annak a valós fókuszpont távolságát, de ehelyett direkt módon mérhetjük a mélységélességi tartomány alsó és felső határait, mivel minden pixelhez rendelkezésünkre áll a neki megfelelő $\mathbf{X}$ 3D pont, így könnyedén számolható a valós kamera-pont távolság, nincs más dolgunk, minthogy éles képrészleteket keressünk és azok alapján meghatározzuk az élességi tartományt. Erre a célra a [13] által bevezetett élességi mértéket használtuk, ami a wavelet transzformáció koefficienseinek különböző magas frekvenciájú sávok béli statisztikáját alkalmazza. Egy csúszó ablak technikát alkalmazva kiválasztjuk azokat a véletlenszerűen mintavételezett w_s ablakokat, amelyekben az élességi mérték egy experimentálisan meghatározott határérték felett van, majd kiszámoljuk a kamera és a w_s ablak pontjai közötti $dist(w_s)$ átlagos távolságot, ez által meghatározhatjuk a minimális és maximális távolságot a szélsőértékek leszűrése után. Az így kapott határértékekkel szűrjük ki azokat a kamerákat, amelyekben egy adott pontnak nem lehet éles a vetülete.

Ezen a ponton már minden $\mathbf{X}$ 3D ponthoz csatolva van azon kamerák listája, amelyek rálátnak a pontra, és élesen láthatják azt. Következő lépésként kiválasztjuk azt, amelyik optimális szögből lát rá a felszínre, és a lehető legjobb felbontásban. Számoljuk előbb ki $\mathbf{X}$ pontban a felszín $\mathbf{n}_\mathbf{X}$ normálját, és annak az $\mathbf{o}_{\mathbf{X}_i}$ vetítő sugárral bezárt szögét, amely $\mathbf{X}$ pontból C_i kamera optika középpontjába mutat:

$$\theta = \arccos\left(\frac{\mathbf{n}_\mathbf{X} \cdot \mathbf{o}_{\mathbf{X}_i}}{\|\mathbf{n}_\mathbf{X}\| \cdot \|\mathbf{o}_{\mathbf{X}_i}\|}\right), \quad (15)$$

ahol $\mathbf{o}_{\mathbf{X}_i} = \mathbf{X} - \mathbf{c}_i$ a vetítő sugár az i -edik kamerába. A $|\theta| \in [0 \dots \pi/2)$ szögek a geometriailag helyesek, mivel bármilyen más szög azt jelentené, hogy a kamera a felszín hátsó felét látja. Természetesen egy közel merőleges rálátás, alacsony $|\theta|$ értékkel a legkedvezőbb.

Vizsgáljuk még a régió vetületének felbontását is, mivel egy nagyobb fókusz-távolságú kamera akár távolabbról is képes részletgazdagabb képet adni, de akár egy kisebb fókusz-távolságú, nagy látószögű kamera is lehet előnyösebb, ha kellően közelről rögzít egy felszínt. Egy $\mathbf{X}_m$ pont i -edik kamerában levő vetületének felbontását $res_{mi} = f_i/D_{mi}$ által jellemezzük, ahol f_i a kamera fókusz-távolsága és D_{mi} pedig az i kamera távolsága $\mathbf{X}_m$ ponttól.

A végső döntés pedig a legmagasabb dc értékkel rendelkező kamera kiválasztásával történik meg, ahol

$$dc_{mi} = res_{mi}/\theta' \quad (16)$$

és θ' egy $\theta' \in [0 \dots 1)$ tartományba skálázott változata a θ szögnek, ahol a 0 felel meg a

merőleges rálátásnak és 1 pedig a $\pi/2$ szögnek. dc_{mi} jelöli az i . kamera döntési értékét az $\mathbf{X}_m$ 3D pontra való tekintettel.

Textúrázás

Sok alkalmazásban előnyös a kisebb méretű 3D modell, ha az nagyjából azonos vizuális részletességgel bír. Ezt tipikusan a pontfelhő helyett háromszögelt felszínek használatával lehet elérni, ami nagyban leegyszerűsíti a modellünket lecsökkentve annak vertex pontjai számát, például az ún. tizedelő algoritmusokkal [14]. Másik fontos előnye, hogy a vertex pontok színezése helyett lehetőség nyílik textúrák illesztésére a háló minden háromszögéhez.

Alkalmazva ezt a javasolt munkafolyamatunkban [Frohlich *et al.*, 2018], a pontok helyett az F háromszögeket járjuk végig, így lehetőségünk nyílik különböző kamerákat kiválasztani szomszédos háromszögekhez, amelyeknek közös vertex csúcsai vannak, tehát nem vagyunk vertex pontonként 1 kamerára korlátozva. A kamera rangsorolási lépés továbbra is szükséges marad, csak az algoritmus utolsó lépését szükséges adaptálni, mivel ebben az esetben az F háromszögeken iterálunk végig, és vizsgáljuk a csúcspontjaihoz tartozó $3 C^{v_k}$ kamera rangsorolási listát, amelyek az előzőekben definiált dc döntési értékeket tartalmazzák minden C_i kameráról:

$$C^{v_k} = dc_{ki}, \text{ ahol } k \in (a, b, c) \text{ és } i \in (1..n). \quad (17)$$

Végül azt a C_i kamerát választjuk ki és rendeljük F_j -hez, amelyik mindhárom C^{v_k} listában szerepelt, és a lehető legmagasabb dc értékekkel bír:

$$C^{F_j} \in (C^{v_a} \cap C^{v_b} \cap C^{v_c}) \text{ ahol } dc = \max dc_{ki} \quad (18)$$

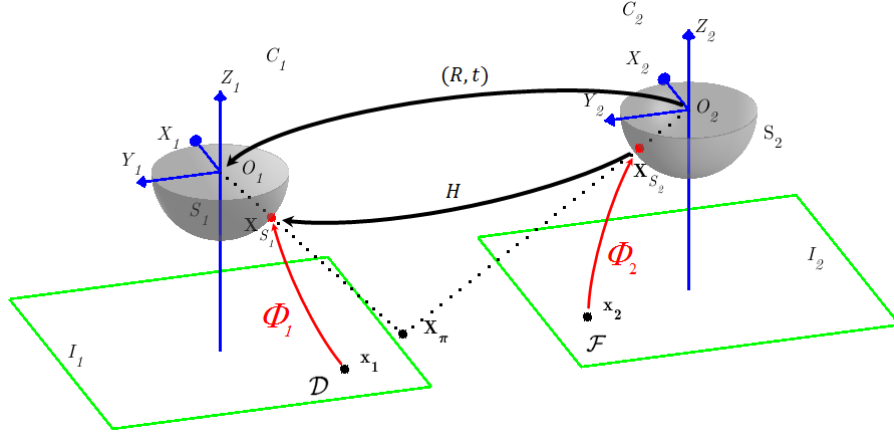
Az így előálló adatstruktúra egyszerűen exportálható egy ASCII formátumú Wavefront OBJ fileba, annak standard specifikációi [15] szerint.

A javasolt módszer hatékonyságát két nagyméretű esettanulmányon bizonyítottuk, a szlovákiai Somorja és Kolozsnéma református templomairól készült részletes adathalmazokon.

Homográfia becslés

Ebben a szekcióban a homográfia becsléssel kapcsolatos eredményeinket mutatjuk be. Az [1] 2D regisztrációs módszere által inspirálva kiterjesztettük azt szférikus kamerák közötti homográfiákra [Frohlich, Tamas, Kato, 2016]. Lényegében a homográfia ebben az esetben egyazon régió két gömbfelszíni projekciója között hat. Általánosságban elmondható, hogy mind a kamerák relatív pose-ja [Frohlich, Tamas, Kato, 2016], mind pedig az indukáló sík normál vektora [Molnár *et al.*, 2014] és távolsága kifaktorizálható egy ilyen síkhomográfiából. De a homográfia alapvető paraméterezésének köszönhetően lehetőség nyílik direkt megoldók kifejlesztésére is. Az általunk javasolt módszer szimultán módon direkt becsüli a relatív kamera pose-t, a síkhomográfiát és az azt indukáló sík paramétereit is [Frohlich, Kato, 2018].

Ha adott egy π sík, az $\mathbf{X}_\pi \in \pi$ pontok vetítését az $\mathcal{S}_i, i = 1, 2$ kamerák gömbfelszínére $\Phi(\mathbf{x}) = \mathbf{X}_\mathcal{S} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ írja le, tehát ez egy bijektív függvény. Feltételezve, hogy a referencia koordináta-rendszer az első kamerában van, jelöljük a π sík normálvektorát és távolságát az



3. ábra. Egységgömbök által reprezentált omni kamerák között ható homográfia.

origótól $\mathbf{n} = (n_1, n_2, n_3)^T$ és d -vel, és a második kamera relatív pose-ját $\mathbf{R}$ forgatással és $\mathbf{t} = (t_1, t_2, t_3)^T$ eltolással (lásd 3. ábrán). Így az $\mathcal{S}_2$ gömbről az $\mathcal{S}_1$ -re való vetítés egyszerűen a fenti pose transzformáció alkalmazásával majd a gömbfelszínre való normalizálással történik:

$$\mathbf{x}_{\mathcal{S}_1} = \frac{\mathbf{R}\mathbf{X}_{\mathcal{S}_2} + \mathbf{t}}{\|\mathbf{R}\mathbf{X}_{\mathcal{S}_2} + \mathbf{t}\|}.$$

A centrális kamerák egy nézőpontos feltételének teljesülése miatt a síkhomográfiák érvényesek maradnak omni kamerák esetében is [16].

Lényegében mondhatjuk, hogy Φ egy ekvivalens szférikus képet ad és a $\mathbf{H}$ síkhomográfia ezek a szférikus képek között hat [Frohlich, Tamas, Kato, 2016], ahogy azt a 3. ábrán is szemléltetjük. Tehát a homográfia az $\mathbf{x}_{\mathcal{S}_1} \propto \mathbf{H}\mathbf{x}_{\mathcal{S}_2}$ vetítősugarakat transzformálja, vagyis a szférikus pontok közötti transzformáció is bijektív lesz. Így az $\mathbf{X}_\pi$ pont és annak szférikus képei $\mathbf{x}_{\mathcal{S}_1}$, $\mathbf{x}_{\mathcal{S}_2}$ és a megfelelő omni pixelek $\mathbf{x}_1$, $\mathbf{x}_2$ között a következő kapcsolat:

$$\Phi_1(\mathbf{x}_1) = \mathbf{X}_{\mathcal{S}_1} = \frac{\mathbf{H}\mathbf{x}_{\mathcal{S}_2}}{\|\mathbf{H}\mathbf{x}_{\mathcal{S}_2}\|} = \Psi(\Phi_2(\mathbf{x}_2)). \quad (19)$$

A fenti egyenletet bármelyik $(\mathbf{x}_1, \mathbf{x}_2)$ megfelelő pontpár teljesíti, így a klasszikus megoldás legalább 4 ilyen pont-megfeleltetést keresne standard intenzitás alapú módszerekkel és azokkal oldaná meg az egyenletet $\mathbf{H}$ -ra. Viszont az omni kamerák nem-lineáris torzítása komoly kihívás elé állítja a hagyományos kulcspont detektorokat és invariáns leíró kinyerőket. Ezért mi egy pontmegfeleltetések nélküli, régió alapú módszert javasunk. Integráljuk ki (19) mindkét oldalát, így $\mathcal{S}_1$ -en a $\mathcal{D}_\mathcal{S} = \Phi_1(\mathcal{D})$ és $\mathcal{F}_\mathcal{S} = \Psi(\Phi_2(\mathcal{F}))$ foltok feletti felszíni integrálokat kapunk. A $\mathcal{D}_\mathcal{S}$ és $\mathcal{F}_\mathcal{S}$ foltokat Φ_1 és $\Psi \circ \Phi_2$ függvények által a $\mathcal{D} \subset \mathbb{R}^2$ és $\mathcal{F} \subset \mathbb{R}^2$ régiók felett paraméterezhetjük:

$$\begin{aligned} \forall \mathbf{X}_{\mathcal{S}_1} \in \mathcal{D}_\mathcal{S} & : \mathbf{X}_{\mathcal{S}_1} = \Phi_1(\mathbf{x}_1), \mathbf{x}_1 \in \mathcal{D} \\ \forall \mathbf{Z}_{\mathcal{S}_1} \in \mathcal{F}_\mathcal{S} & : \mathbf{Z}_{\mathcal{S}_1} = \Psi(\Phi_2(\mathbf{x}_2)), \mathbf{x}_2 \in \mathcal{F}, \end{aligned}$$

így felírhatjuk a felszíni integrálok explicit képletét:

$$\iint_{\mathcal{D}} \omega_i(\Phi_1(\mathbf{x}_1)) \left\| \frac{\partial \Phi_1}{\partial x_{11}} \times \frac{\partial \Phi_1}{\partial x_{12}} \right\| dx_{11} dx_{12} = \iint_{\mathcal{F}} \omega_i(\Psi(\Phi_2(\mathbf{x}_2))) \left\| \frac{\partial(\Psi \circ \Phi_2)}{\partial x_{21}} \times \frac{\partial(\Psi \circ \Phi_2)}{\partial x_{22}} \right\| dx_{21} dx_{22} \quad (20)$$

Mivel az egyenletek száma nem elegendő, újabb egyenleteket konstruálunk az [1] által bemutatott technikával. Valóban, egy megfelelően megválasztott ω esetében

$$\omega(\mathbf{x}_{S_1}) = \omega(\Psi(\Phi_2(\mathbf{x}_2))). \quad (21)$$

Így kellően sok független egyenletet generálhatunk nem-lineáris $\{\omega_i\}_{i=1}^\ell$ függvények használatával. Megjegyezzük viszont, hogy az új egyenletek nem tartalmaznak új információt, csak egyszerűen új lineárisan független megkötéseket hoznak be. Az egyenletrendszer megoldása direkt módon megadja a $\mathbf{H}$ paramétereit.

A számítási komplexitás nagyban redukálható, ha észrevevesszük, hogy a (20) integrál bal oldala konstans. Viszont az ismeretlen $\mathbf{H}$ homográfia részt vesz a jobb oldal számolásában a Ψ függvény által, így ezeket minden iterációban újra kell számolni. Természetesen a $\mathbf{X}_{S_2} = \Phi_2(\mathbf{x}_2)$ szférikus pontok is előre kiszámolhatóak, de a felszíni elem számolása jóval komplexebb. Írjuk át az összetett $\Psi \circ \Phi_2$ függvény deriváltját a Ψ függvény $\mathbf{J}_\Psi$ Jacobijának és a Φ_2 gradienseinek függvényében:

$$\left\| \frac{\partial(\Psi \circ \Phi_2)}{\partial x_{21}} \times \frac{\partial(\Psi \circ \Phi_2)}{\partial x_{22}} \right\| = \left\| \mathbf{J}_\Psi \frac{\partial \Phi_2}{\partial x_{21}} \times \mathbf{J}_\Psi \frac{\partial \Phi_2}{\partial x_{22}} \right\|$$

Mivel Φ_2 gradiensei függetlenek $\mathbf{H}$ -tól, előre kiszámolhatóak. Így csak $\Psi(\Phi_2(\mathbf{x}_2))$ -t és $\mathbf{J}_\Psi(\Phi_2(\mathbf{x}_2))$ -t kell iterációnként újra kiszámolni, ami egy hatékonyabb algoritmust eredményez [Frohlich, Tamas, Kato, 2016].

Mivel az egyenletrendszert az algebrai hiba minimalizálásával oldjuk meg, fontos a megfelelő adat-normalizálás a numerikus stabilitás érdekében [1]. Viszont [1]-el ellentétben, esetünkben a szférikus koordináták már eleven a $[-1, 1]$ tartományban vannak, így nincs szükség további normalizálásra. Viszont az ω_i függvényeket is normalizálni kell $[-1, 1]$ -be, hogy minden egyenlet kiegyensúlyozott kontribúcióval járjon hozzá az algebrai hiba számolásához. Esetünkben ez úgy érhető el, hogy az integrálokat leosztjuk a félgömb feletti felszíni integrál maximális magnitúdójával. Az optimális eredmény érdekében fontos a megfelelő inicializálás is, ami biztosítja, hogy a $\mathcal{D}_S$ és $\mathcal{F}_S$ felszíni régiók jól átfednek. Ez egyszerűen elérhető ha a $\mathcal{D}_S$ és $\mathcal{F}_S$ régiók centroidja közötti forgatással inicializáljuk a $\mathbf{H}$ homográfiát.

Ebben a szekcióban bemutattunk egy régió alapú homográfia becslő algoritmust, ami a kamerák belső Φ_1 és Φ_2 vetítésétől független, viszont ezen függvények, és gradienseik ismerete szükséges a (20) egyenletek számolásához. A módszer robusztusságát szintetikus adatokon validáltuk, és a két legelterjedtebb omni kamera modell használatát is összehasonlítottuk.

Homográfia faktorizálás

Figyelembe véve, hogy egy $\mathbf{H}$ síkhomográfia az $\mathbf{R}$ forgatásból, a $\mathbf{t}/d$ eltolás és sík távolság arányából és az $\mathbf{n}$ normálvektorból a következőképpen szerelhető össze:

$$\mathbf{H} \propto \mathbf{R} - \mathbf{t}\mathbf{n}^T/d, \quad (22)$$

a pose paramétereit kifejezhetjük a [17] által javasolt módon, $\mathbf{H}$ SVD felbontásával. Természetesen mivel a d sík távolság ismeretlen, ezért a $\mathbf{t}$ eltolás is csak egy skála erejéig határozható meg. Ezt a szabad skála faktort $\mathbf{H}$ utolsó elemének h_{33} 1-re állításával fogtuk le.

Másféle megközelítést tesz lehetővé, ha például egy ember-építette környezetet feltételezhetünk, ahol az ún. *weak Manhattan world* [18] feltételezés könnyen teljesül, mivel tetszőleges orientációjú vertikális síkok vannak, amelyek párhuzamosak a gravitációs vektorral, és merőlegesek a talaj síkjára. [18] nyomán felhasználhatjuk a vertikális irány ismeretét is, amit akár egy a kamerához csatolt inerciális mérőeszköz (IMU) is szolgáltathat. Míg [18] perspektív kamerákkal dolgozott, [Frohlich, Tamas, Kato, 2016]-ban bemutattuk, hogy az omni kamerák között ható homográfiák úgyszintén alkalmazhatóak a pose kifaktorizálására.

Tételezzük fe, hogy egy π vertikális síkunk van, melynek $\mathbf{n} = (n_x, n_y, 0)^T$ a normál vektora (z a vertikális tengely, lásd 3. ábrán). A sík d távolsága 1-re állítható, mivel $\mathbf{H}$ csak egy szabad skála erejéig meghatározott. A vertikális irányt ismerve, az $\mathbf{R}$ forgató mátrix (22)-ben egy $\mathbf{R}_z$ z tengely körüli forgatássá egyszerűsödik:

$$\mathbf{H} = \mathbf{R}_z - (t_x, t_y, t_z)(n_x, n_y, 0)^T. \quad (23)$$

Egy ilyen *weak Manhattan* homográfia becslése az előzőekben leírt módon történik, csak $\mathbf{H}$ utolsó oszlopa $(0, 0, 1)^T$, így csak 6 szabad paramétere van [Frohlich, Tamas, Kato, 2016]. A fenti paraméterezés alapján $\mathbf{H}$ -ből könnyedén kifaktorizálható az α forgatási szög, és a $\mathbf{t} = (t_x, t_y, t_z)^T$ relatív eltolás a kamerák között [Frohlich, Tamas, Kato, 2016], például olyan információkat felhasználva, mint $n_x^2 + n_y^2 = 1$, $t_z = \pm \sqrt{h_{31}^2 + h_{32}^2}$ (lásd [18]-t több részletért).

Szintetikus generált *weak Manhattan* típusú adathalmazon bizonyítottuk, hogy mind a homográfia becslés, mind pedig a relatív pose faktorizálás a becsült homográfiákból robusztusan működik, és a standard SVD alapú módszerhez [19] mérhető eredményeket produkál.

Sík rekonstrukció

Omnidirekcionális kamerák közötti síkhomográfia és a pose paraméterek ismeretében egy zárt alakú megoldást javasoltunk [Molnár *et al.*, 2014]-ben a 3D sík régió normálvektorának rekonstrukciójára, melyet az előző szekcióban bemutatott homográfia becslő eredményein validáltuk. Mikor már rendelkezésünkre áll az $\mathbf{n}$ normálvektor, d egyszerűen számolható (22) alapján, például [20]-által leírt módon. A [Molnár *et al.*, 2014] féle differenciál geometriai megoldásban kamera-független egyenleteket konstruálunk a képrégiók közötti lineáris transzformáció Jacobi mátrixa segítségével, melynek elemeit a megfigyelt régió

normálvektora és a vetítési függvények gradiensei által fejezzük ki [21]:

$$[\mathbf{J}_{ij}] = \frac{1}{|\nabla x_i^1 \mathbf{n} \nabla x_i^2|} \begin{bmatrix} |\nabla x_j^1 \mathbf{n} \nabla x_i^2| & |\nabla x_i^1 \mathbf{n} \nabla x_j^1| \\ |\nabla x_j^2 \mathbf{n} \nabla x_i^2| & |\nabla x_i^1 \mathbf{n} \nabla x_j^2| \end{bmatrix}. \quad (24)$$

A fenti mennyiségek mind invariáns elsőrendű differenciák. Észrevehetjük, hogy (24) egy általános képlet: se a projekció alakjáról, se a felszínről nem feltételez semmit, így tetszőleges kamera típussal és bármilyen kellően sima felszínnel alkalmazható. A képlet alapján az $\mathbf{n}$ normál vektor kiszámolható, ha a vetítési függvények és a $\mathbf{J}_{ij}$ Jacobi ismertek. Írjuk fel egy becsült síkhomográfia deriváltjainak mátrix elemeit [Molnár *et al.*, 2014]:

$$[\mathbf{J}_{ij}]_{est} = \begin{bmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{bmatrix}. \quad (25)$$

Sor-, oszlop- vagy kereszt-arányokat használva a közös nevező megszüntethető [Molnár *et al.*, 2014], így a 3D felszíni normál egyenlete felírható a $\frac{a_1^1}{a_2^2}$ és $\frac{a_2^1}{a_1^2}$ kereszt-arányokkal. A $[\mathbf{J}_{ij}]_{est} = [\mathbf{J}_{ij}]$ átrendezése után a következőt kapjuk:

$$\begin{aligned} \mathbf{n} \cdot \left[a_2^2 (\nabla x_i^2 \times \nabla x_j^1) - a_1^1 (\nabla x_j^2 \times \nabla x_i^1) \right] &= 0 \\ \mathbf{n} \cdot \left[a_1^2 (\nabla x_j^1 \times \nabla x_i^1) - a_2^1 (\nabla x_i^2 \times \nabla x_j^2) \right] &= 0 \end{aligned} \quad (26)$$

ahol két ismert vektorunk van, mindkettő a normálra merőleges:

$$\begin{aligned} \mathbf{p} &= \mathbf{n} \cdot \left[a_2^2 (\nabla x_i^2 \times \nabla x_j^1) - a_1^1 (\nabla x_j^2 \times \nabla x_i^1) \right] \\ \mathbf{q} &= \mathbf{n} \cdot \left[a_1^2 (\nabla x_j^1 \times \nabla x_i^1) - a_2^1 (\nabla x_i^2 \times \nabla x_j^2) \right] \end{aligned} \quad (27)$$

Így a normál könnyedén kifejezhető a vetítési függvények gradiensei által:

$$\mathbf{n} = \frac{\mathbf{p} \times \mathbf{q}}{|\mathbf{p} \times \mathbf{q}|}. \quad (28)$$

A $\nabla x_k^l, k = i, j; l = 1, 2$ koordináta gradiensek térbeli koordináták és $\mathbf{J}_{ij}$ szerinti számolása részletesen be lett mutatva [Molnár *et al.*, 2014]-ben egy omni kamera-pár esetében.

Összefoglalva, ha adott egy pár egymásnak megfelelő F és D régió egy omnidirekcionális kamera-párban, melyek Φ_i, Φ_j vetítési függvénye ismert, a π 3D sík rekonstruálható a következő lépések által:

1. $\mathbf{H}$ homográfia becslése a megfelelő szférikus régiók között (például használva [Frohlich, Tamas, Kato, 2016]-ot), ami megadja a Ψ függvényt.
2. $(\mathbf{R}, \mathbf{t})$ relatív pose becslése a kamerák között. Ha már $\mathbf{H}$ megvan, használhatjuk például [19] vagy [Frohlich, Tamas, Kato, 2016] homográfia faktorizáló módszereket.
3. $\mathbf{n}$ normál vektor számolása (28) direkt képlettel, majd d távolság számolása standard módszerrel (22) alapján (például [20]).

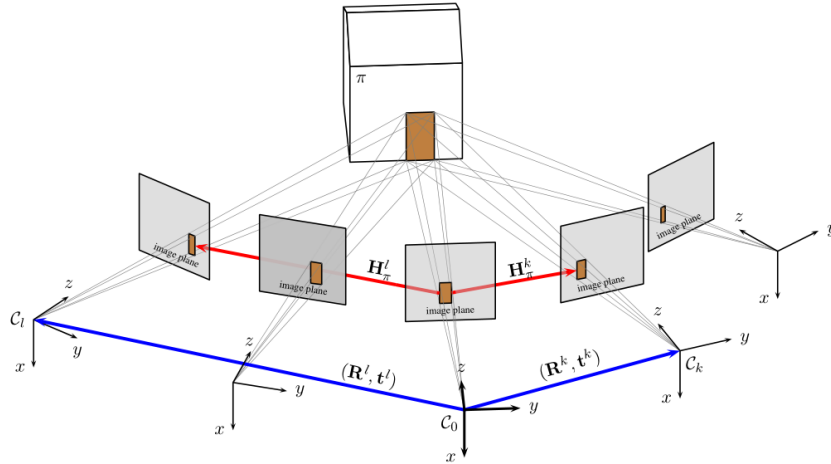
Az algoritmust szintetikus adathalmazon validáltuk, és a klasszikus Hartley és Zisserman [20] féle homográfia faktorizáló módszerhez hasonló teljesítményt produkált.

Szimultán relatív pose becslés és sík rekonstrukció

Az előző szekcióban bemutatott módszerekkel ellentétben, ahol lépésenként, függetlenül becsültünk síkhomográfiát képi régiók között, majd faktorizáltunk relatív pose-t a homográfiából, végül az indukáló sík paramétereit számoltuk ki, ebben a szekcióban egy olyan szimultán módszert [Frohlich, Kato, 2018] mutatunk be perspektív kamerákra, ami egyszerre oldja meg a fentieket.

Visszatérve a perspektív kamerák abszolút helyzetének képletére (7), az első szekcióban leírt módon dolgozhatunk ismét egyenesen a normalizált képeken. Fogalmazzuk meg egy π sík és annak két normalizált kamerában vett $\mathcal{D}^0$ és $\mathcal{D}^1$ vetített képe közötti kapcsolatot (lásd a 4. ábrán). $\mathcal{C}_0$ kamerát megválasztva a referencia koordináta rendszernek, a π síkot annak $\mathbf{n} = (n_1, n_2, n_3)^\top$ egység normálvektorával és d origótól vett távolságával fejezhetjük ki ebben a koordináta rendszerben. A $\mathcal{C}_1$ kamera relatív pose-ja pedig az a 3D merevtest transzformáció $(\mathbf{R}^1, \mathbf{t}^1)$ ami $\mathcal{C}_0 \rightarrow \mathcal{C}_1$ hat. Tehát bármely homogén $\mathbf{X}$ 3D pont képe a két normalizált kamerában:

$$\mathbf{x}_{\mathcal{C}_0} \cong [\mathbf{I}|\mathbf{0}]\mathbf{X} \quad \text{és} \quad \mathbf{x}_{\mathcal{C}_1} \cong [\mathbf{R}^1|\mathbf{t}^1]\mathbf{X}. \quad (29)$$



4. ábra. Egy 3D π sík vetítése egy több-nézetes kamera rendszerben.

Természetesen a 3D sík $\mathbf{X}_\pi \in \pi$ pontjainak $\mathcal{C}_i, i = 0, 1$ kamerákba való vetítését ugyanazek az egyenletek írják le, így felírható a $\mathbf{H}_\pi^1 : \mathcal{D}^0 \rightarrow \mathcal{D}^1$ síkhomográfia, amit $\pi = (\mathbf{n}, d)$ sík indukál a $\mathcal{D}^0$ és $\mathcal{D}^1$ képi régiók között. $\mathbf{H}_\pi^1$ egy szabad skálafaktor erejéig (22) szerint szerelhető össze. Így bármely $\mathbf{X}_\pi \in \pi$ pont esetében a következő kapcsolat áll fenn a megfelelő $\mathbf{x}_{\mathcal{C}_0}$ és $\mathbf{x}_{\mathcal{C}_1}$ normalizált képi pontok között:

$$\mathbf{x}_{\mathcal{C}_1} \cong \mathbf{H}_\pi^1 \mathbf{x}_{\mathcal{C}_0} \cong (\mathbf{R}^1 - \frac{1}{d} \mathbf{t}^1 \mathbf{n}^\top) \mathbf{x}_{\mathcal{C}_0}. \quad (30)$$

A klasszikus megoldás szerint elegendő lenne legalább 4 ilyen pontpárt találnunk és megoldanunk az egyenletet $\mathbf{H}_\pi^1$ -re, majd kifaktorizálhatjuk $\mathbf{R}^1, \mathbf{t}^1$ és $\mathbf{n}$ paramétereiket [22]. Viszont a javasolt régió alapú megközelítésünk [Frohlich, Kato, 2018] egy speciális egyenletrendszer megoldásával, pont-megfeleltetések nélkül, robusztusan határozza meg a keresett paramétereiket.

[1] felvetését követve, (30) mindkét oldalát kiintegrálva kikerüljünk a pontmegfeleltetések használatát. Megfelelő $\omega : \mathbb{R}^2 \rightarrow \mathbb{R}$ függvényeket alkalmazva az egyenlet mindkét

oldalán az egyenlőség igaz marad, és a következő integrál egyenletek kapjuk:

$$\int_{\mathcal{D}^1} \omega(\mathbf{x}_{\mathcal{C}_1}) d\mathbf{x}_{\mathcal{C}_1} = \int_{\mathcal{D}^0} \omega(\mathbf{H}_{\pi}^1 \mathbf{x}_{\mathcal{C}_0}) |\mathbf{J}_{\mathbf{H}_{\pi}^1}(\mathbf{x}_{\mathcal{C}_0})| d\mathbf{x}_{\mathcal{C}_0}, \quad (31)$$

ahol az $\mathbf{x}_{\mathcal{C}_1} = \mathbf{H}_{\pi}^1 \mathbf{x}_{\mathcal{C}_0}$, $d\mathbf{x}_{\mathcal{C}_1} = |\mathbf{J}_{\mathbf{H}_{\pi}^1}(\mathbf{x}_{\mathcal{C}_0})| d\mathbf{x}_{\mathcal{C}_0}$ integrál transzformációk lettek alkalmazva. Mivel $\mathbf{H}_{\pi}^1$ egy 3×3 -as homogén mátrix aminek 8 szabadsági foka van, az utolsó elemét 1-re állítjuk. Vegyük észre, hogy a fenti egyenlet inhomogén $\mathbf{x}_{\mathcal{C}_i}$ pont koordinátákra igaz, melyeket projektív osztással kaphatunk. A Jacobi determinánsa $|\mathbf{J}_{\mathbf{H}_{\pi}^1}| : \mathbb{R}^2 \rightarrow \mathbb{R}$ megadja a transzformáció mértékét minden egyes pontban [1].

Az ismeretlen $(\mathbf{R}^1, \mathbf{t}^1)$ és $(\mathbf{n}, d)$ paraméterek a (31) nemlineáris egyenletrendszer megoldásaként kaphatóak meg. Praktikusán egy túlhatározott egyenletrendszert konstruálunk és legkisebb négyzetek szerint oldjuk meg az algebrai hiba minimalizálásával standard *Levenberg-Marquardt* algoritmussal.

Több régió és több nézet

A javasolt módszer a minimális konfigurációban (2 kamera és 1 régió) a klasszikus homográfia becselő és faktorizáló megközelítéshez hasonló eredményeket produkál, kulcsfontosságú előnye viszont abban rejlik, hogy több kamera és/vagy több régió használata esetén képes azokat egy rendszerben kezelni. Praktikusán, mivel minden π_i sík eltérő $\mathbf{H}_{\pi_i}^1$ homográfiát generál a megfelelő $\mathcal{D}_i^0$ és $\mathcal{D}_i^1$ képrégiók között, (30) és (31) igaz marad ezekre a homográfiákra, azzal az észrevétellel, hogy az $(\mathbf{R}^1, \mathbf{t}^1)$ relatív pose közös az összes $\mathbf{H}_{\pi_i}^1$ homográfiában, azok csak a 3D sík $(\mathbf{n}_i, d_i)$ paramétereiben különböznek. Így minden $\{\pi_i\}_{i=1}^N$ síkra felírható:

$$\mathbf{x}_{\mathcal{C}_1} \cong \mathbf{H}_{\pi_i}^1 \mathbf{x}_{\mathcal{C}_0} \cong (\mathbf{R}^1 - \frac{1}{d_i} \mathbf{t}^1 \mathbf{n}_i^\top) \mathbf{x}_{\mathcal{C}_0}, \text{ ahol } \mathbf{x}_{\mathcal{C}_0} \in \mathcal{D}_i^0 \text{ és } \mathbf{x}_{\mathcal{C}_1} \in \mathcal{D}_i^1 \quad (32)$$

és így (31) egy N egyenletből álló rendszer lesz [Frohlich, Kato, 2018] a közös kamera pose $(\mathbf{R}^1, \mathbf{t}^1)$ és a $\{\pi_i\}_{i=1}^N$ síkok $(\mathbf{n}_i, d_i)$ paraméterei szerint:

$$\int_{\mathcal{D}_i^1} \omega(\mathbf{x}_{\mathcal{C}_1}) d\mathbf{x}_{\mathcal{C}_1} = \int_{\mathcal{D}_i^0} \omega(\mathbf{H}_{\pi_i}^1 \mathbf{x}_{\mathcal{C}_0}) |\mathbf{J}_{\mathbf{H}_{\pi_i}^1}(\mathbf{x}_{\mathcal{C}_0})| d\mathbf{x}_{\mathcal{C}_0}, \quad 1 \leq i \leq N \quad (33)$$

Egy ω függvényre így a fenti egyenletek N megkötést adnak a relatív pose paramétereire, de csak 1 megkötést az egyes π_i síkok paramétereire, összesen N egyenletben. Ne feledjük, hogy a teljes (33) rendszernek továbbra is egy szabad skálája van, mivel az egyes d_i távolságok relatív értékét meghatározzuk, csak egyik fogható le szabadon.

Hasonlóan, amikor több kamera látja ugyanazokat a sík régiókat, egy olyan egyenletrendszert konstruálhatunk, amely több megkötést tartalmaz nem csak a kamera pose-ra de minden 3D síkra is. Először megválasztjuk a $\mathcal{C}_0$ kamerát a referencia koordináta rendszernek, így minden kamera relatív pose-ja $\mathcal{C}_0$ -hoz képest lesz kifejezve, és minden sík rekonstrukciós paraméterei is $\mathcal{C}_0$ -ban lesznek adottak. Feltételezve, hogy minden $\{\pi_i\}_{i=1}^N$ sík látható minden $\{\mathcal{C}_k\}_{k=0}^{M-1}$ kamerában, az egyes π_i síkok $\mathbf{H}_{\pi_i}^k$ homográfiákat generálnak a megfelelő $\mathcal{D}_i^0$ régió és a k -adik kamera $\mathcal{D}_i^k$ régiója között:

$$\forall 1 \leq k \leq M-1 : \quad \mathbf{x}_{\mathcal{C}_k} \cong \mathbf{H}_{\pi_i}^k \mathbf{x}_{\mathcal{C}_0} \cong (\mathbf{R}^k - \frac{1}{d_i} \mathbf{t}^k \mathbf{n}_i^\top) \mathbf{x}_{\mathcal{C}_0}. \quad (34)$$

Tehát minden kamera egy újabb megkötést hoz be az $(\mathbf{n}_i, d_i)$ sík paraméterekre, így össze-

sen $M - 1$ megkötést eredményezve π_i rekonstrukciójára [Frohlich, Kato, 2018]. Ha egy sík nem látható minden kamerából, akkor a megkötések száma értelemszerűen csökken.

Több kamera esetében azt mondhatjuk, hogy egy $(\mathcal{C}_0, \mathcal{C}_k)$ kamerapár N egyenletet határoz meg a közös $(\mathbf{R}^k, \mathbf{t}^k)$ pose paraméterek és a $\{\pi_i\}_{i=1}^N$ síkok $(\mathbf{n}_i, d_i)$ paraméterei szerint, így (31)-hez hasonlóan N egyenletből álló rendszert eredményezve. Tehát

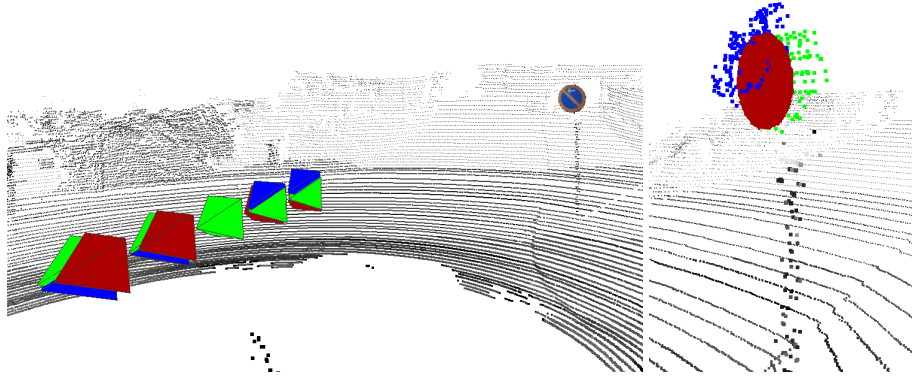
$$\int_{\mathcal{D}_i^k} \omega(\mathbf{x}_{\mathcal{C}_k}) d\mathbf{x}_{\mathcal{C}_k} = \int_{\mathcal{D}_i^0} \omega(\mathbf{H}_{\pi_i}^k \mathbf{x}_{\mathcal{C}_0}) |\mathbf{J}_{\mathbf{H}_{\pi_i}^k}(\mathbf{x}_{\mathcal{C}_0})| d\mathbf{x}_{\mathcal{C}_0},$$

$$1 \leq i \leq N \text{ és } 1 \leq k \leq M - 1 \quad (35)$$

Egy ω függvényre így a fenti egyenletek N megkötést nyújtanak az $(\mathbf{R}^k, \mathbf{t}^k)$ relatív pose-ra és $M - 1$ megkötést minden π_i síkra, összesen $N(M - 1)$ egyenletben. Az egyenletek minimális száma ami $M \geq 2$ kamera és $N \geq 1$ sík régió esetében a megoldáshoz szükséges: $E = 6(M - 1) + 3N - 1$.

A javasolt algoritmus [Frohlich, Kato, 2018] több kamera és több régió esetében egy kétlépcsős megoldást alkalmaz. Először minden szomszédos kamerapárra oldjuk meg az egyenleteket egy páronkénti megoldásban, ahol a pose és rekonstrukciós paraméterek semmiféle specifikus inicializálást nem igényelnek. Majd második lépésként egy kötegetelt behangolást végzünk, amelyhez minden paramétert a kiválasztott referencia kamera koordináta rendszerébe transzformálunk, felírjuk a referenciához képesti relatív pose-okat, és a rekonstrukciós paramétereket a páronkénti megoldások iniciális eredményének szűrt átlagával inicializáljuk. Ennek az egyenletrendszernek a globális megoldása adja a végső eredményeket, egy szabad skálafaktor erejéig.

A javasolt módszert alapos szintetikus kiértékeléssel validáltuk többféle konfigurációban, és valós adatokon is teszteltük. A KITTI [9] adathalmazon végzett összehasonlító kiértékelésben a State-of-the-Art COLMAP [23] általános rekonstrukciós módszernél jobb eredményt ért el (egy példa összehasonlító eredmény az 5. ábrán látható).



5. ábra. Példa összehasonlító eredmény a relatív kamera pose-ok (balról) és egy forgalmi tábla rekonstrukcióját (jobbról) illusztrálva. Zöld a referencia, piros a javasolt módszer, kék pedig COLMAP[23].

A szerző hozzájárulásának összefoglalása

A dolgozat eredményeit két fő téziscsoportban foglaltam össze, ahol az elsőben abszolút pose becslésével és vizuális adatok fúziójával foglalkozom, míg a másodikban kamerák közötti síkhomográfia becsléssel és 3D síkrekonstrukcióval. A téziscsoportok és az elfogadott publikációim közötti kapcsolatot a B.1 táblázatban prezentálom.

I. Abszolút pose becslés és adatfúzió

Az [1] által bemutatott 2D regisztrációs módszer által inspirálva, [2] egy újféle régió alapú abszolút pose becslő megoldást javasolt, amely perspektív kamerák egy 3D tér-adathoz képesti helyzetét képes meghatározni egy általános 2D-3D regisztrációs megoldással, mindenféle kalibrációs minta vagy explicit pontmegfeleltetések használata nélkül. Ez az alap ötlet kibővíthető egy általános abszolút pose becslési keretrendszerre centrális szférikus kamerák számára, amely különböző, vizuális adatfúziós feladatokra alkalmazható. Az alap ötlet egy nem-lineáris egyenletrendszer konstruálása, melynek a megoldása direkt módon adja meg a keresett transzformáció paramétereit. Ez a tézis csoport az abszolút pose becslési és adat-fúziós témákban elért eredményeimet foglalja össze.

- (a) Kísérleti úton kimutattam az omnidirekcionális kamerák számára [Tamas, Frohlich, Kato, 2014] által bevezetett régió alapú abszolút pose becslő algoritmus teljesítményét. A centrális kamerák számára bemutatott általános regisztrációs keretrendszerhez validáltam a gömbfelszíni háromszöghálón dolgozó integrál számolót. A perspektív kamerák felszíni integráljainak számolására egy hatékony, 2D geometriai momentumok rekurzív felírásán alapuló számolási módszer vezettem be. A forgatási és eltolási paraméterek inicializálására egy automatikus megoldást javasoltam perspektív és omni kamerák számára. A módszer kvantitatív kiértékeltem szintetikus adathalmazokon, összehasonlítva a korábbi pont-alapú integrál közelítéssel megoldással, és vizsgálva a perspektív kamerák esetében a szférikus és klasszikus modell használatát. A módszer használhatóságát különféle kamerákkal és 3D szenzorokkal rögzített valós adatokon is igazoltam.
- (b) A kulturális örökségvédelmi objektumok vizuális-adat fúziójához adaptáltam a régió-alapú regisztrációs módszerünket [Tamas, Frohlich, Kato, 2014], kiterjesztve azt nem-sík, de síma régiókra. A javasolt munkafolyamat részeként egy intenzitás információból kinyerhető élekre támaszkodó ICP alapú finomítási lépést javasoltam, míg a több kamerából történő fúzióra egy egyszerű megoldást a kamerák orientációja alapján. Kvantitatív kiértékelés alapján bizonyítottam, hogy a módszer nem-sík felületekre kiterjesztve is robusztus marad, míg valós, kulturális örökségvédelmi szempontból érdekes tárgyakon is helyes eredményeket kaptam. A második alkalmazás egy nagyméretű kulturális örökségvédelmi objektumok (például templomok) dokumentálására szolgáló fúziós munkafolyamat, melyhez megoldást javasoltam a nagy mennyiségű nézőpont esetében felmerülő kamera szelekciós kérdésre, ami figyelembe veszi az egyes kamerák rálátását, élességét, betekintési szögét és felbontását. Az eredmények vizualizációjára javasoltam egy olyan technikai megoldást, amely képes nagy mennyiségű

gű, különálló textúra-kép kezelésére. A munkafolyamatot két református templomról rögzített 2D-3D nagyméretű adathalmazon validáltam.

II. Síkhomográfia becslés és 3D rekonstrukció

Az [1] 2D regisztrációs megoldás kiterjeszthető szférikus kamerák között ható síkhomográfiák becslésére is. Lényegében ez esetben a homográfiák az azonos sík régióknak megfelelő gömbfelszíni vetületek között értelmezhetők. Általánosságban elmondható, hogy a relatív pose paraméterek és az indukáló sík paraméterei különböző módszerekkel faktorizálhatóak ki az így meghatározott homográfiából, de a síkhomográfia eredendő paraméterezésének köszönhetően direkt megoldásokra is lehetőség nyílik, ezáltal kikerülhető a faktorizálás, annak minden velejáró bizonytalanságával. Ez a tézis csoport a síkhomográfia becslés és 3D síkrekonstrukció témákban elért eredményeimet foglalja össze.

- (a) Kísérleti úton igazoltam az omnidirekcionális kamerák között ható síkhomográfiák becslésére javasolt módszert, többféle szférikus modellt felhasználva. [18] által inspirálva bemutattam egy megoldást a relatív pose faktorizálására síkhomográfiából, amely vertikális irány ismeretében és egy *Manhattan világ* feltételezés mellett a [17] standard módszerhez mérhető pontosságot produkált szintetikus adatokon. Ha már a relatív pose paraméterei rendelkezésre állnak, az indukáló sík paramétereit is meghatározhatjuk a homográfiából. A bemutatott módszerünk [Frohlich, Tamas, Kato, 2016] által becsült homográfiákkal igazoltam a sík normálvektorának a kiszámolására javasolt differenciál geometriai megoldást. Szintetikus adathalmazon végeztem összehasonlító kiértékelést, melyben a klasszikus [20] módszernél jobb teljesítményt értünk el, és a forgatási és eltolási paraméterekben levő hibákra is kellően robusztusnak bizonyult a módszerünk.
- (b) A homográfia becslés az egyenletek megfelelő átparaméterezésével akár direkt módon is megadhatja a kamerák relatív pose-át és a sík paramétereit, ezzel kiküszöbölhető a homográfia faktorizálás és az azzal járó bizonytalanságok. Mivel minden kamera pár és minden régió egy újabb homográfiát határoz meg, ezek egyenleteit felírtam a közös paraméterek függvényében egy több kamerás több régiós rendszerben, és validáltam az algoritmust úgy a minimális megoldási esetben, mint több különböző konfigurációban is. A több kamerás esetre egy köteget behangolási megoldást is javasoltam, mely szimultán módon egyszerre finomítja az összes keresett paramétert. Az algoritmus teljesítményét kiértékeltem szintetikus és többféle valós adaton is: pontos Lidar pontfelhővel és marker alapú referencia mérésekkel rendelkező saját adathalmazon, továbbá a KITTI publikus adathalmazon is, ahol a módszerünk a legjobb pont-alapú általános rekonstrukciós módszernél [23] jobban teljesített.

	I		II	
	a	b	a	b
[Tamas, Frohlich, Kato, 2014]	•			
[Frohlich, Tamas, Kato, 2019]	•			
[Frohlich <i>et al.</i> , 2016]		•		
[Frohlich <i>et al.</i> , 2018]		•		
[Frohlich, Tamas, Kato, 2016]			•	
[Molnár <i>et al.</i> , 2014]			•	
[Frohlich, Kato, 2018]				•

1. táblázat. A tézispontokhoz kapcsolódó publikációk.

Publikációs lista

Folyóirat cikkek

[Frohlich, Tamas, Kato, 2019] R. Frohlich, L. Tamas, and Z. Kato. “Absolute Pose Estimation of Central Cameras Using Planar Regions”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019). accepted subject to minor revision, under review, pp. 1–16.

Könyv fejezetek

[Frohlich *et al.*, 2018] R. Frohlich, S. Gubo, A. Lévai, and Z. Kato. “3D-2D Data Fusion in Cultural Heritage Applications”. In: *Heritage Preservation: A Computational Approach*. Ed. by B. Chanda, S. Chaudhuri, and S. Chaudhury. Springer Singapore, 2018, pp. 111–130. ISBN: 978-981-10-7221-5. DOI: 10.1007/978-981-10-7221-5_6.

[Frohlich, Tamas, Kato, 2016] R. Frohlich, L. Tamas, and Z. Kato. “Handling Uncertainty and Networked Structure in Robot Control”. In: vol. 42. *Studies in Systems, Decision and Control*. Chapter 6. Springer, Feb. 2016. Chap. Homography Estimation Between Omnidirectional Cameras Without Point Correspondences, pp. 129–151.

Konferencia-kiadványokban megjelenő cikkek

[Frohlich, Kato, 2018] R. Frohlich and Z. Kato. “Simultaneous Multi-View Relative Pose Estimation and 3D Reconstruction from Planar Regions”. In: *Proceedings of ACCV Workshop on Advanced Machine Vision for Real-life and Industrially Relevant Applications*. Ed. by G. Carneiro. Vol. 11367. Lecture Notes in Computer Science. Springer, Dec. 2018. ISBN: ISBN 978-3-030-21074-8. DOI: 10.1007/978-3-030-21074-8.

[Frohlich *et al.*, 2016] R. Frohlich, Z. Kato, A. Tremeau, L. Tamas, S. Shabo, and Y. Waksman. “Region Based Fusion of 3D and 2D Visual Data for Cultural Heritage Objects”. In: *Proceedings of International Conference on Pattern Recognition*. IEEE. Cancun, Mexico: IEEE, Dec. 2016, pp. 2404–2409.

[Molnár *et al.*, 2014] J. Molnár, R. Frohlich, C. Dmitry, and Z. Kato. “3D Reconstruction of Planar Patches Seen by Omnidirectional Cameras”. In: *Proceedings of International Conference on Digital Image Computing: Techniques and Applications*. Wollongong, Australia: IEEE, Nov. 2014, pp. 1–8. ISBN: ISBN 978-1-4799-5409-4.

[Tamas, Frohlich, Kato, 2014] L. Tamas, R. Frohlich, and Z. Kato. “Relative Pose Estimation and Fusion of Omnidirectional and Lidar Cameras”. In: *Proceedings of the ECCV Workshop on Computer Vision for Road Scene Understanding and Autonomous Driving*. Ed. by L. de Agapito, M. M. Bronstein, and C. Rother. Vol. 8926. Lecture Notes in Computer Science. Zurich, Switzerland: Springer, Sept. 2014, pp. 640–651. ISBN: ISBN 978-3-319-16180-8.

Referenciák

- [1] C. Domokos, J. Nemeth, and Z. Kato. “Nonlinear Shape Registration without Correspondences”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.5 (May 2012), pp. 943–958. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.200.
- [2] L. Tamas and Z. Kato. “Targetless Calibration of a Lidar - Perspective Camera Pair”. In: *Proceedings of ICCV Workshop on Big Data in 3D Computer Vision*. IEEE. Sydney, Australia: IEEE, Dec. 2013, pp. 668–675.
- [3] D. Scaramuzza, A. Martinelli, and R. Siegwart. “A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion”. In: *Proceedings of the Fourth IEEE International Conference on Computer Vision Systems*. ICVS-06. Washington, USA: IEEE Computer Society, 2006, pp. 45–51.
- [4] D. Scaramuzza, A. Martinelli, and R. Siegwart. “A Toolbox for Easily Calibrating Omnidirectional Cameras.” In: *IEEE/RSJ International Conference on Intelligent Robots*. Beijing: IEEE, Oct. 2006, pp. 5695–5701.
- [5] D. Herrera C, J. Kannala, and J. Heikkila. “Joint Depth and Color Camera Calibration with Distortion Correction.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.10 (2012), pp. 1–8.
- [6] F. M. Mirzaei, D. G. Kottas, and S. I. Roumeliotis. “3D LIDAR-camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization”. In: *The International Journal of Robotics Research* 31.4 (2012), pp. 452–467.
- [7] J. E. G. Joseph O’Rourke, ed. *Handbook of Discrete and Computational Geometry, Second Edition (Discrete Mathematics and Its Applications)*. Chapman and Hall/CRC, 2004. ISBN: 9781584883012.
- [8] J. M. Pozo, M. C. Villa-Uriol, and A. F. Frangi. “Efficient 3D Geometric and Zernike Moments Computation from Unstructured Surface Meshes”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.3 (Mar. 2011), pp. 471–484. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.139.
- [9] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. “Vision meets Robotics: The KITTI Dataset”. In: *International Journal of Robotics Research (IJRR)* (2013), pp. 1231–1237.
- [10] Z. Taylor and J. Nieto. “A Mutual Information Approach to Automatic Calibration of Camera and Lidar in Natural Environments”. In: *Australian Conference on Robotics and Automation*. Wellington, Australia, Dec. 2012, pp. 3–5.
- [11] P. J. Besl and N. D. McKay. “Method for registration of 3-D shapes”. In: *Robotics-DL tentative*. International Society for Optics and Photonics. 1992, pp. 586–606.

- [12] S. Katz, A. Tal, and R. Basri. “Direct Visibility of Point Sets”. In: *ACM SIGGRAPH 2007 Papers*. SIGGRAPH '07. San Diego, California: ACM, 2007. DOI: 10.1145/1275808.1276407.
- [13] G. Yang and B. J. Nelson. “Wavelet Based Autofocusing and Unsupervised Segmentation of Microscopic Images”. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2003, Las Vegas, Nevada, October*. Vol. 3. IEEE, 2003, 2143–2148 vol.3. DOI: 10.1109/IROS.2003.1249188.
- [14] M. Garland and P. S. Heckbert. “Surface simplification using quadric error metrics”. In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH '97* (1997), pp. 209–216. ISSN: 00978930. DOI: 10.1145/258734.258849.
- [15] K. McHenry and P. Bajcsy. *An overview of 3D data content, file formats and viewers*. Tech. rep. 2008, p. 21.
- [16] C. Mei, S. Benhimane, E. Malis, and P. Rives. “Efficient Homography-Based Tracking and 3-D Reconstruction for Single-Viewpoint Sensors”. In: *Robotics, IEEE Transactions on* 24.6 (Dec. 2008), pp. 1352–1364. ISSN: 1552-3098. DOI: 10.1109/TRO.2008.2007941.
- [17] O. Faugeras and F. Lustman. *Motion and structure from motion in a piecewise planar environment*. Tech. rep. RR-0856. June 1988.
- [18] O. Saurer, F. Fraundorfer, and M. Pollefeys. “Homography based visual odometry with known vertical direction and weak Manhattan world assumption”. In: *IEEE/IROS Workshop on Visual Control of Mobile Robots (ViCoMoR)*. 2012.
- [19] P. Sturm. “Algorithms for plane-based pose estimation”. In: *Proceedings of International Conference on Computer Vision and Pattern Recognition*. Vol. 1. June 2000, pp. 706–711. DOI: 10.1109/CVPR.2000.855889.
- [20] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge, UK: Cambridge University Press, 2004.
- [21] J. Molnár and D. Chetverikov. “Quadratic Transformation for Planar Mapping of Implicit Surfaces”. In: *Journal of Mathematical Imaging and Vision* 48.1 (2014), pp. 176–184.
- [22] O. Faugeras and F. Lustman. *Motion and structure from motion in a piecewise planar environment*. Tech. rep. RR-0856. Sophia Antipolis, France: INRIA, June 1988.
- [23] J. L. Schonberger and J.-M. Frahm. “Structure-from-Motion Revisited”. In: *Proceedings of International Conference on Computer Vision and Pattern Recognition*. IEEE, June 2016. DOI: 10.1109/cvpr.2016.445.