

Omnidirekcionális kamerák közötti homográfia becslése pontmegfeleltetések nélkül

Fröhlich Róbert¹, Tamás Levente², Kató Zoltán¹

¹ Szegedi Tudományegyetem, Képfeldolgozás és Számítógépes Grafika Tanszék

frohlich@inf.u-szeged.hu, kato@inf.u-szeged.hu

² Kolozsvári Műszaki Egyetem, Automatizálás Kar

Levente.Tamas@aut.utcluj.ro

Kivonat Ebben a cikkben egy újszerű megközelítést mutatunk be az omnidirekcionális kamerák közötti homográfia becslésére. A megoldást egyenletrendszerként írjuk fel, melyben minden egyenlet az omni kamerát modellező egységnyi gömbfelszín két egymásnak megfelelő régiója között ható nemlineáris függvény integráljaként áll elő. A módszerünk nem használ pontmegfeleltetéseket vagy egyéb hasonlósági metrikákat, csak egy pár, az omnidirekcionális képekről kinyert, egymásnak megfelelő sík régiót. A megbecsült homográfiából a kamerák közötti relatív helyzet is kifaktorizálható ún. gyenge Manhattan világ feltételek mellett. A bemutatott módszer hatékonyságát és robusztusságát szintetikus és valós adatokon is bizonyítottuk.¹

1. Bevezető

A homográfia becslés esszenciális része sok alkalmazásnak, így például a helyzet becslés [1], követés [2,3], struktúra mozgásból [4] és egyéb navigációra [5], látásra és észlelésre kielezett robotikai alkalmazásban is használatos. Klasszikus perspektív kamerák esetében léteznek hatékony módszerek a homográfia becslésére [6], de ezek általában nem megbízhatóak omnidirekcionális kamerákkal használva. Az omnidirekcionális szenzorok esetében a homográfia becslésének nehézségeit elsősorban a nemlineáris vetítési modell okozza, ami révén az alakzatok torzulva jelennek meg a képen, így a klasszikus módszerek direkt használata nem lehetséges.

Bár ezeknek a nem konvencionális (katadioptrikus és dioptrikus) centrális panoráma kameráknak a geometriai modellje sokkal komplexebb, kalibrálásuk ugyancsak a belső és külső paraméterek meghatározásán alapul. A belső paraméterek lényegében a kamera leképezését leíró modell paraméterei (a kamera fix gyári paraméterei). Ezzel ellentétben a külső paraméterek egy adott kép elkészítésekor a kamera térbeli helyét és orientációját írják le, amelyek értelemszerűen

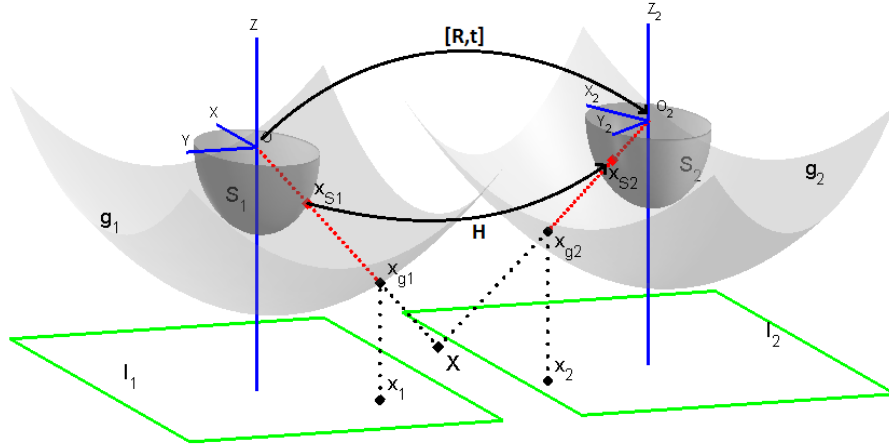
¹ A cikk publikálásra került az ICRA nemzetközi konferencia OMNIVIS workshopján, 2014. júniusában. Robert Frohlich, Levente Tamas and Zoltan Kato: Homography Estimation between Omnidirectional Cameras without Point Correspondences. In Proc. of ICRA Workshop on Non-classical Cameras, Camera Networks and Omnidirectional Vision (OMNIVIS), HongKong, June 2014. IEEE

minden elkészült képnél változnak. Az elmúlt időszakban többen is áthatóan tanulmányozták ezen kamera-rendszerek geometriai leírását [7,8,9,10,11,12]. Az így kapott geometriai modelleket felhasználva a kamerák belső kalibrálása megoldható ellenőrzött környezetben speciális kalibrációs mintákat használva [11,13,14,12]. Ha a kameránk kalibrálva van, ami általában a valós alkalmazásokban adott, akkor a kép-pontokat az egységnyi gömb felszínére vetítve egy egységes modellt kapunk, ami így már független a kamerák belső nemlineáris vetítésétől. A perspektív esettel ellentétben a homográfiát ezen gömbfelszíni pontok segítségével becsülhetjük meg [2,3]. Természetesen a helyzetbecslésnek az aktuális képek alapján kell történnie, melyeket valós környezetben rögzítenek a kamerák, így nem feltételezhetjük semmiféle speciális kalibrációs célpontok meglétét. A klasszikus megoldás pontmegfeleltetéseket keres, és a homográfiát az így kapott pont-párok alapján becsüli, viszont a standard kulcspont detektálók (mint például a SIFT [15]), amelyek használata eléggé elterjedt [4,2], nem túl hatékonyak nemlineáris torzítás jelenlétében. Mivel az omnidirekcionális képeken a formák felbontása nagyban változhat az invariáns jellemzők vagy kulcspontok kinyerése problémákat okozhat, ezek nélkül viszont nem kivitelezhető a megbízható pontmegfeleltetés. Skála invariáns jellemzők kinyerése is megoldható omnidirekcionális szenzorokra, de ez különös megfontolást igényel, ami viszont egy matematikailag igen komplex algoritmushoz vezet [16]. Létezik megfeleltetések nélküli módszer is [4] a kamera relatív mozgásának meghatározására, viszont ez is használ SIFT jellemzőket, mivel a kamera mozgást az epipoláris megkötést teljesítő összes jellemző pár felett integrálva számolja ki. Néhányan foglalkoztak már jellemző nélküli kép illesztéssel és felismeréssel (a legfontosabb [17]), jelentősebb sikerek nélkül.

Ebben a cikkben bemutatunk egy olyan homográfia becselő algoritmust, ami direkt módon kiszegmentált sík régiókkal dolgozik, így a módszerünk nem igényli kulcspontok és egyéb jellemzők kinyerését. Tulajdonképpen nem használunk semmilyen fotometriai információt a képről, így módszerünk akár multimodális szenzorokkal is használható. Mivel a legtöbb valós képfeldolgozó alkalmazásnak tartalmaznia kell egy adott ponton egy szegmentálási lépést, így ezek a szükséges sík régiók előfordulhat, hogy már eleve elérhetőek számunkra, vagy ha nem, akkor nagyon egyszerűen detektálhatóak. Érdekes továbbá azt is megjegyeznünk, hogy a szegmentálás kevésbé érzékeny a nemlineáris torzulásokra, nagyobb alakzatok így is könnyen kiszegmentálhatóak. Módszerünk legfőbb előnyei tehát a régiók használata pontmegfeleltetések helyett, és a probléma általános megfogalmazása, ami lehetővé teszi különböző kamera típusok kezelését egy keretrendszerben. A homográfia becslést egy alakillesztési feladatként fogalmazzuk meg, ami [18]-hoz hasonló módon hatékonyan megoldható. Bár a módszer alap elképzelése ugyanaz, mint a [18]-ban bemutatott, mégis az algoritmusunk lényegesen különbözik, lévén ez egy specifikus eset, omnidirekcionális kamerákra alkalmazott változat. Bár a bemeneti adataink 2D képekről (nemlineáris torzítást tartalmazó omni képekről) kiszegmentált régiók, de az omnidirekcionális kamera modell használata által a 2D-2D regisztráció helyett egy 3D gömbfelszínre megszorított regisztrá-

ciós feladatot kapunk. Ebből adódóan, másképp írjuk fel az integrálokat, eltérő normalizálás és inicializálás szükséges.

A bemutatott módszer szintetikus adathalmazon kiértékelve hatékonynak és robusztusnak bizonyult. Az [5]-ben bemutatottak alapján azt is kimutattuk, hogy a megbecsült homográfia akár a kamerák közötti relatív helyzet kinyerésére is felhasználható, bizonyos *Manhattan világ* feltételezés mellett.



1. ábra. Omnidirekcionális kamera modell

2. Omnidirekcionális kamera modell

A középpontos omnidirekcionális kamerákra egy egységes modellt mutatott be Geyer és Daniilidis [9] amely a panoráma kamerát mint egy egységnyi gömb felszínére való vetítést ír le. Ez a formalizmus többek által is elfogadott és alkalmazott modellt eredményezett, így például Micusik [10] is ezt használva a kamerák belső vetítő függvényét határozta meg, majd Scaramuzza [19] egy általános polinom függvény formában fejezte ki ezeket a belső vetítési paramétereket, ami így bármilyen omnidirekcionális kamerára alkalmazható.

Vizsgáljuk meg először az \mathcal{I} omnidirekcionális képsík \mathbf{x} pontjának a kapcsolatát az \mathcal{S} egységnyi gömb felszínén vett képével (lásd 1. ábrán). Megfigyelhető, hogy csak a gömbfelszín felét használjuk, mivel a másik fele nem látható a képsíkból. Feltételezzük, hogy a kamera koordináta rendszere az \mathcal{S} gömbben található, úgy hogy a gömb középpontja, a koordináta rendszer origója és a kamera vetítési középpontja is egybe esnek [11,19], továbbá a z tengely a kamera optikai tengelye, ami a képsíkot a főpontban metszi.

A középpontos omnidirekcionális optikák nemlineáris (viszont szimmetrikus) torzítását [11,19] úgy fejezi ki, hogy egy g felszínt helyez a képsík és az \mathcal{S}

gömb közé, ami forgatás-szimmetrikus a z tengely mentén. Ez a g felszín többféle módon leírható, mi egy negyedrangú polinomiális [11] formát használtunk $g(\|\mathbf{x}\|) = a_0 + a_2\|\mathbf{x}\|^2 + a_3\|\mathbf{x}\|^3 + a_4\|\mathbf{x}\|^4$ aminek 4 paramétere (a_0, a_2, a_3, a_4) leírja a kamera belső leképezését (csak 4 paramétert használunk, mivel a_1 mindig 0, [11] alapján). A $\Phi : \mathcal{I} \rightarrow \mathcal{S}$ bijektív leképezés 2 lépésből áll: 1) az $\mathbf{x} \in \mathcal{I}$ képi pontok felemelése a g felszínre egy ortografikus vetítés által

$$\mathbf{x}_g = \begin{bmatrix} \mathbf{x} \\ a_0 + a_2\|\mathbf{x}\|^2 + a_3\|\mathbf{x}\|^3 + a_4\|\mathbf{x}\|^4 \end{bmatrix} \quad (1)$$

majd 2) az így kapott \mathbf{x}_g pontok középpontos vetítése az \mathcal{S} egységnyi gömb felszínére.

$$\mathbf{x}_S = \Phi(\mathbf{x}) = \frac{\mathbf{x}_g}{\|\mathbf{x}_g\|} \quad (2)$$

Tehát az omnidirekcionális kamera vetítését leírhatjuk az \mathbf{x}_S egységvektorok által, amik lényegében a vetítő sugarak egységvektorainak is megfelelnek.

2.1. Sík homográfia

A kamera koordináta rendszerében felvett $\mathbf{X} \in \mathbb{R}^3$ 3D pont középpontos leképezéssel vetül az \mathcal{S} felszínre, így az \mathbf{X} pont és a képe \mathbf{x} az omni kamerában a következő kapcsolatban áll:

$$\Phi(\mathbf{x}) = \mathbf{x}_S = \frac{\mathbf{X}}{\|\mathbf{X}\|} \quad (3)$$

Vegyünk egy π térbeli síkot, majd írjuk fel a kapcsolatot a sík \mathcal{D} és \mathcal{F} képe között, amelyek az \mathcal{S}_1 és \mathcal{S}_2 gömbfelszín által modellezett kamerákban keletkeznek. A sík $\mathbf{X}_\pi \in \pi$ pontjai a (3) egyenlet szerint vetülnek az $\mathcal{S}_i, i = 1, 2$ kamerák gömbfelszínére, tehát egy bijektív leképezést kapunk, kivéve ha a π sík áthalad a kamera középpontján, mivel ebben az esetben π láthatatlan a kamera számára. Feltételezzük, hogy az első kamera koordináta rendszere a világ koordináta rendszer, jelöljük a π normálját és távolságát az origótól $\mathbf{n} = (n_1, n_2, n_3)^T$ és d -vel. A második kamera relatív pozíciója egy \mathbf{R} forgatásból és $\mathbf{t} = (t_1, t_2, t_3)^T$ eltolásból tevődik össze, ami a két kamera \mathcal{S}_1 és \mathcal{S}_2 gömbje között hat. Tehát bármely \mathbf{X} 3D pont képét a második kamerában felírhatjuk mint

$$\mathbf{x}_{S_2} = \frac{\mathbf{R}\mathbf{X} + \mathbf{t}}{\|\mathbf{R}\mathbf{X} + \mathbf{t}\|} \quad (4)$$

Mivel egy közös nézőpont van, a sík homográfia felírható az omni kamerákra is [2]. A \mathbf{H} standard sík homográfia egy skála faktor erejéig meghatározható mint:

$$\mathbf{H} \propto \mathbf{R} + \frac{1}{d}\mathbf{t}\mathbf{n}^T \quad (5)$$

Lényegében a homográfia az $\mathbf{x}_{S_1} \propto \mathbf{H}\mathbf{x}_{S_2}$ sugarak közötti transzformáció, ezért a gömbfelszíni pontok között ható sík homográfia is bijektív. A \mathbf{H} egy skála faktor erejéig van meghatározva, amit megválaszthatunk ha $h_{33} = 1$, vagyis ha

\mathbf{H} -t leosztjuk az utolsó elemével, feltéve hogy az nem zéró. Viszont $h_{33} = 0$ csak akkor teljesül ha $\mathbf{H}(0, 0, 1)^T = (h_{13}, h_{23}, 0)^T$, tehát ha az első kamera koordináta rendszerének origója az ideális vonalra vetül a második kamera képén. Ez csak extrém helyzetekben fordulhat elő, amit a valós alkalmazásokban általában kizárnak bizonyos fizikai megkötések. Tehát az \mathbf{X}_π síkbeli pont és a gömbfelszíni képei \mathbf{x}_{S_1} , \mathbf{x}_{S_2} közötti kapcsolat

$$\mathbf{X}_\pi = \lambda_1 \mathbf{x}_{S_1} = \lambda_2 \mathbf{H} \mathbf{x}_{S_2} \Rightarrow \mathbf{x}_{S_1} = \frac{\lambda_2}{\lambda_1} \mathbf{H} \mathbf{x}_{S_2} \quad (6)$$

amiből látható hogy \mathbf{x}_{S_1} és $\mathbf{H} \mathbf{x}_{S_2}$ egy sugáron van, így felírható:

$$\mathbf{x}_{S_1} = \frac{\mathbf{H} \mathbf{x}_{S_2}}{\|\mathbf{H} \mathbf{x}_{S_2}\|} = \Psi(\mathbf{x}_{S_2}) \quad (7)$$

3. Homográfia becslés

Tegyük fel, hogy van egy omnidirekcionális kamera párunk melyek egy sík régiót figyelnek meg, hogyan számolhatunk homográfiát a két kamera képe között? Először fogalmazzuk meg a kapcsolatot két egymásnak megfelelő omni \mathbf{x}_1 , \mathbf{x}_2 képpont között. A (2) egyenlet alapján ezek gömbfelszínre vetített koordinátáit a kamerák Φ_1 , Φ_2 belső vetítési függvényeit alkalmazva kapjuk, majd ezek a felszíni pontok között a (7) egyenlet hat, így felírható:

$$\Phi_1(\mathbf{x}_1) = \mathbf{x}_{S_1} = \frac{\mathbf{H} \mathbf{x}_{S_2}}{\|\mathbf{H} \mathbf{x}_{S_2}\|} = \Psi(\Phi_2(\mathbf{x}_2)) \quad (8)$$

Ezt az egyenletet bármely $(\mathbf{x}_1, \mathbf{x}_2)$ pontpár kielégíti, ezért ha a klasszikus megoldást követnénk, egyszerűen intenzitás alapú pont-párosítással összegyűjtenénk a megfelelő pontokat, és megoldanánk az egyenletet \mathbf{H} -ra. Viszont minket egy olyan megoldás érdekel, ami nem használ pontmegfeleltetéseket.

Bemutatjuk, hogy a \mathbf{H} homográfia könnyedén megbecsülhető minden további információ nélkül, csak a két képen beazonosított, \mathcal{D} és \mathcal{F} -el jelölt, sík régió használatával. Mivel a kameránk belső paraméterei már ismertek, tehát egy kalibrált kamerával dolgozunk, ezért könnyedén használhatjuk a pontok gömbfelszíni megfelelőjét. Hogy megszabaduljunk az egyéni pont-pároktól, integráljuk a (8) egyenlet mindkét oldalát. Így egy felszíni integrált kapunk az \mathcal{S}_1 gömbön, \mathcal{D}_S és \mathcal{F}_S felszíni foltok felett, amiket az első kép \mathcal{D} régiójának gömbfelszínre vetítéséből és a második kép \mathcal{F} régiójának gömbfelszínre vetítéséből és transzformálásából kapunk. A felszíni integrálok explicit kifejezését úgy kapjuk, hogy a felszíni \mathcal{D}_S és \mathcal{F}_S foltokat paraméterezzük Φ_1 és $\Psi \circ \Phi_2$ függvények által a síkbeli \mathcal{D} és \mathcal{F} régiók felett, úgy hogy: $\mathcal{D} \subset \mathbb{R}^2$, $\mathcal{F} \subset \mathbb{R}^2$; és $\forall \mathbf{x}_{S_1} \in \mathcal{D}_S : \mathbf{x}_{S_1} = \Phi_1(\mathbf{x}_1)$, $\mathbf{x}_1 \in \mathcal{D}$ valamint $\forall \mathbf{z}_{S_1} \in \mathcal{F}_S : \mathbf{z}_{S_1} = \Psi(\Phi_2(\mathbf{x}_2))$, $\mathbf{x}_2 \in \mathcal{F}$, ami a következő integrál egyenletet eredményezi:

$$\iint_{\mathcal{D}} \Phi_1(\mathbf{x}_1) \left\| \frac{\partial \Phi_1}{\partial x_{11}} \times \frac{\partial \Phi_1}{\partial x_{12}} \right\| dx_{11} dx_{12} = \iint_{\mathcal{F}} \Psi(\Phi_2(\mathbf{x}_2)) \left\| \frac{\partial(\Psi \circ \Phi_2)}{\partial x_{21}} \times \frac{\partial(\Psi \circ \Phi_2)}{\partial x_{22}} \right\| dx_{21} dx_{22} \quad (9)$$

ahol a parciális deriváltak keresztszorzatának magnitúdója a felszíni elemként ismert. A fenti integrálok lényegében elemenkénti felszíni integráljai skaláris mézőknek, így lényegében 2 egyenletből állnak. Mivel a felszíni integrálok értéke független a paraméterezéstől, a fenti egyenlőség igaz, mert mindkét oldal egy integrált tartalmaz az \mathcal{S}_1 gömbön, Φ_1 illetve $\Psi \circ \Phi_2$ függvényeken keresztül paraméterezve.

3.1. Az egyenletrendszer felépítése

Természetesen ez a 2 egyenlet nem elegendő a homográfia 8 paraméterének a meghatározására, ezért újabb egyenleteket konstruálunk egy általános mechanizmust [18] használva, egy $\omega : \mathbb{R}^3 \rightarrow \mathbb{R}$ függvényt alkalmazva a (8) egyenletünk mindkét oldalán:

$$\iint_{\mathcal{D}} \omega_i(\Phi_1(\mathbf{x}_1)) \left\| \frac{\partial \Phi_1}{\partial x_{11}} \times \frac{\partial \Phi_1}{\partial x_{12}} \right\| dx_{11} dx_{12} = \iint_{\mathcal{F}} \omega_i(\Psi(\Phi_2(\mathbf{x}_2))) \left\| \frac{\partial(\Psi \circ \Phi_2)}{\partial x_{21}} \times \frac{\partial(\Psi \circ \Phi_2)}{\partial x_{22}} \right\| dx_{21} dx_{22} \quad (10)$$

Egy sor nemlineáris $\{\omega_i\}_{i=1}^{\ell}$ függvényt alkalmazva mindenik ω_i egy új független egyenletet hoz létre, így kapunk egy tetszőleges ℓ egyenletből álló rendszert. Bár tetszőleges ω_i függvényeket használhatunk ebben a mechanizmusban, a hatványfüggvények előnyösebbek [18]:

$$\omega_i(\mathbf{x}_{\mathcal{S}}) = x_1^{l_i} x_2^{m_i} x_3^{n_i}, \quad \text{ú.h. } 0 \leq l_i, m_i, n_i \leq 2 \text{ és } l_i + m_i + n_i \leq 3 \quad (11)$$

Ezáltal képesek vagyunk egy 15 egyenletből álló túlhatározott rendszert konstruálni, amit a *legkisebb négyzetes hiba* értelemben standard *Levenberg-Marquardt* algoritmussal oldunk meg. Az egyenletrendszer megoldása direkt módon eredményezi a keresett \mathbf{H} homográfia paramétereit.

A számítási komplexitáshoz nagyban hozzájárul a (10) szereplő integrálok kiszámítása, de mivel mindkét kamera kalibrálva van a Φ_1 és Φ_2 , vagyis a belső leképezésük ismert, ezért az integrálok bal oldala konstans, tehát elegendő csak egyszer kiszámolnunk azokat. Viszont az ismeretlen \mathbf{H} homográfia részt vesz a jobb oldal felírásában a Ψ függvényen keresztül, tehát ezeket az integrálokat már a Levenberg Marquardt megoldó minden iterációjában újra kell számolni. Természetesen a gömbfelszíni $\mathbf{x}_{\mathcal{S}_2} = \Phi_2(\mathbf{x}_2)$ pontokat is előre kiszámolhatjuk, viszont a felszíni elem kiszámolása sokkal komplexebb művelet. Írjuk fel a $\Psi \circ \Phi_2$

összetett függvény parciális deriváltjait a Ψ Jakobi mátrixa \mathbf{J}_Ψ és a Φ_2 gradiensei szerint kifejezve:

$$\left\| \frac{\partial(\Psi \circ \Phi_2)}{\partial x_{21}} \times \frac{\partial(\Psi \circ \Phi_2)}{\partial x_{22}} \right\| = \left\| \mathbf{J}_\Psi \frac{\partial \Phi_2}{\partial x_{21}} \times \mathbf{J}_\Psi \frac{\partial \Phi_2}{\partial x_{22}} \right\| \quad (12)$$

Megfigyelhető, hogy mivel a Φ_2 gradiensei függetlenek a \mathbf{H} -től, ezeket is előre kiszámolhatjuk. Így tehát csak a $\Psi(\Phi_2(\mathbf{x}_2))$ és $\mathbf{J}_\Psi(\Phi_2(\mathbf{x}_2))$ amit a megoldó minden iterációjában szükséges újra kiszámolnunk, ezért az így kapott algoritmus számítási szempontból hatékonynak bizonyult.

3.2. Normalizálás és inicializálás

Mivel a megoldás az algebrai hiba minimalizálásával történik, a helyes normalizálás kritikus fontosságú a numerikus stabilitás érdekében [18]. A gömbfelszíni koordináták már eleve a $[-1, +1]$ tartományban vannak, ezért [18]-al ellentétben itt már nem szükséges további normalizálás. Viszont az ω_i függvényeket is normalizálnunk kell a $[-1, +1]$ tartományba, hogy biztosítsuk minden egyenlet kiegyensúlyozott hozzájárulásukat az algebrai hibához. Esetünkben ez úgy érhető el, hogy az integrálokat leosztjuk a felületi integrál maximális magnitúdójával a fél egységgömbön. Könnyedén kiszámolhatjuk ezeket az integrálokat, ha paraméterezzük a felszínt az egységkör pontjai által az $x - y$ síkban, vagyis $f(x, y) = (x, y, \sqrt{1 - x^2 - y^2})^T, \forall \|(x, y)\| < 1$. Tehát az ω_i függvény által generált egyenlet normalizáló konstansa N_i a következő formájú:

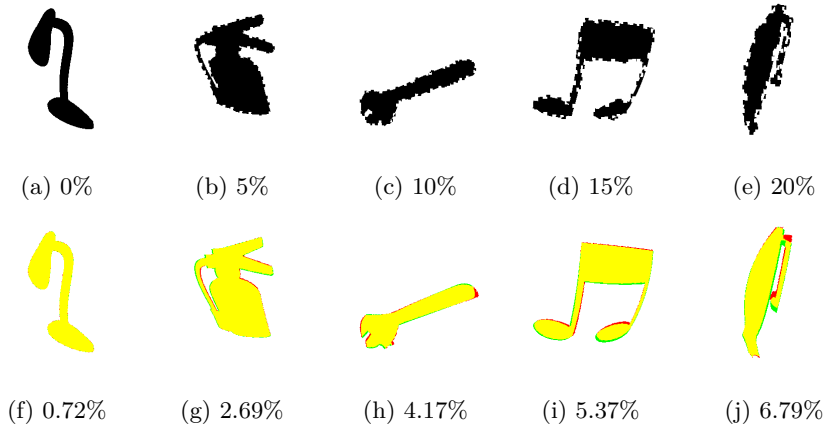
$$N_i = \iint_{\|(x,y)\| < 1} |\omega_i(f(x, y))| \sqrt{\frac{1}{1 - x^2 - y^2}} dx dy \quad (13)$$

Az optimális megoldás érdekében az inicializálás is nagyon fontos. Esetünkben a jó inicializálás a \mathcal{D}_S és \mathcal{F}_S gömbfelszíni foltok minél nagyobb átfedését eredményezi. Ezt úgy érjük el, hogy kiszámoljuk a $\mathcal{D}_S, \mathcal{F}_S$ foltok centroidjait majd a kettő közötti elforgatással inicializáljuk a \mathbf{H} -t.

4. Eredmények

A bemutatott módszer számszerű értékeléséhez összesen 9 adathalmazt generáltunk mindenikben 100 kép párral. 24 különböző alakzatot ábrázoló képet használtunk a helyszín sík régiójaként, és egy pár virtuális omnidirekcionális kamerát, amelyek véletlenszerű pózba beállítva rögzítették az 1 Megapixeles képeket. Feltételezve, hogy ezek a 800×800 pixel méretű alakzatok 5×5 m-es foltoknak felelnek meg a valóságban, a síkot véletlenszerűen helyezzük az első kamera elé 1.5m távolságra, ± 1 m vízszintes eltolással és $\pm [5 - 10]$ fokos elforgatással mindhárom tengely mentén. Majd a második kamera orientációját ismét véletlenszerűen választjuk, ± 5 fokos elforgatással az x és y tengely, valamint ± 10 fok elforgatással a függőleges z tengely mentén, a kamera középpontjának

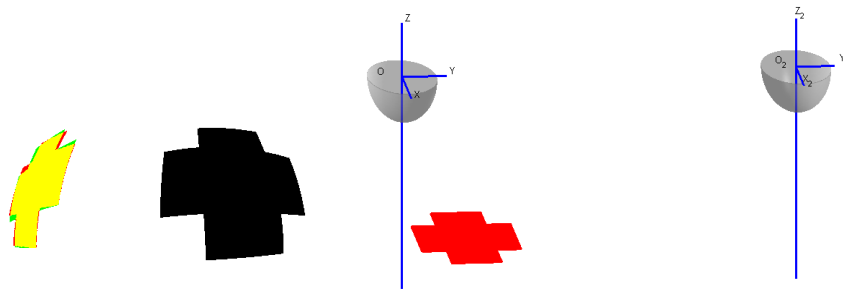
eltolását pedig véletlenszerűen a [45-55]cm, [100-200]cm, és [200-500]cm intervallumok egyikéből választva. Így létrejön az első 3 adathalmazunk 3 különböző tartományban vett bázistávolságokkal.



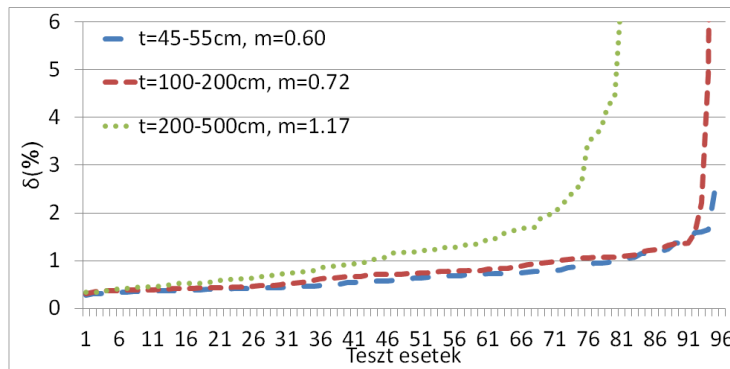
2. ábra. Tipikus regisztrációs eredmények különböző mértékű szegmentálási hibára. Első sor: 0% - 20% szegmentálási hiba hozzáadva az első megfigyelés képezhez. Második sor: a második megfigyelés kép áttanszformálva és az elsőre vetítve, (piros/zöld színben az egymást nem fedő pontok), alattuk a tipikus δ hiba olvasható.

Az illesztési hibát δ -val jelöltük, és úgy definiáltuk, mint az egyik omni régió és a homográfiaival áttanszformált másik omni régió közötti különbség (az egymást nem fedő pontok mennyisége).

Tapasztalati eredményeink alapján kijelentettük, hogy az 5% alatti regisztrációs hiba egy helyes illesztésnek felel meg, ahol a régiók pontosan illeszkednek. A szintetikus adathalmazon mért δ hibák a 4. ábrán láthatóak. A három különböző bázistávolság esetében a δ hiba medián értéke 0.60%, 0.72% és 1.17% (mindhárom eset függetlenül lett sorba rendezve legjobb-legrosszabb értelemben). Az első két esetben, ahol a bázistávolság 200cm alatti, látható, hogy az 5%-os határt csak az esetek 1%-ban lépi át a δ hiba, míg a nagy bázistávolságú adathalmazon (200cm-től 500cm-ig) 84%-a az eredményeknek jónak tekinthető. A kiugróan rossz eredmények tipikusan olyan extrém helyzetekből adódnak ahol a két kamera közötti relatív eltolás olyan irányú, hogy míg az egyik kamera a síkot egy bizonyos szög alatt látja, a másik ellentétes irányból, így teljesen más torzulást tapasztalva látja azt. Az ilyen esetek gondot okozhatnak az algoritmusnak a pontos paraméterek megbecslésében. Egy példa az ilyen esetekre a 3. ábrán látható. Előbb a két kamera képe látható, az első a regisztrációs eredmény kiértékelésére használt visszavetítést is tartalmazza, majd a kamerák közötti relatív pozíciót és a sík helyzetét szemléltető ábra.

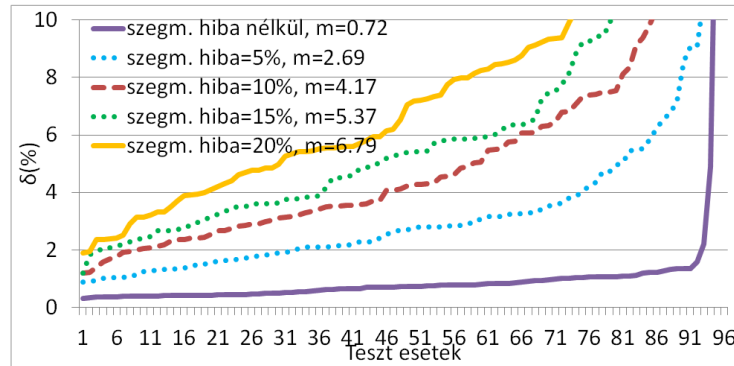


3. ábra. Példa egy extrém szintetikus esetre, előbb a két kamera képe, első tartalmazza a transformált második képet is az elsőre vetítve, majd a kamerák és a sík közötti helyzet.



4. ábra. Illesztési (δ) hiba különböző bázistávolságokra (m a medián érték).

A gyakorlatban ezek a régiók valós képekről kiszegmentálva állnak elő, így az eredmények pontosságát a szegmentálási hiba nagyban befolyásolja. Ezért a módszer szegmentálási hibákkal szembeni robusztusságát is szintetikus adatokon teszteltük. A közepes bázistávolságú (1m-től 2m-ig) adathalmazból kiindulva új halmazokat generáltunk, különböző mértékű körvonal menti zajt adva a kép párok egyikén az alakzathoz. Összesen 4 adathalmazt generáltunk, 5% és 20% közötti szegmentálási zajjal, ezekből egy-egy példa a 2. ábrán látható a regisztrációs eredmények kiértékelésével és a tipikus δ hibával együtt. A 5. ábrán a δ hiba változása követhető a zaj nélküli és a zajos adathalmazokon, esetenként függetlenül sorba rendezve. Megfigyelhető, hogy a δ hiba medián értéke növekszik a hozzáadott szegmentálási hiba mértékével, ami természetes jelenség, viszont a módszer robusztusnak bizonyul akár 15% hiba esetében is.



5. ábra. Illesztési (δ) hiba különböző mértékű szegmentálási hibákra (m a medián érték).

Az algoritmust Matlab-ban implementáltuk, a teszteket egy standard 4 magos i7 processzoros asztali PC-n futtattuk, 5 - 8 másodperces tipikus futási időt mérve optimalizálatlan kódon.

A validációra használt valós képeket egy Canon 50D DSLR kamerával és Canon EF 8-15mm f/4L halszem optikával készítettük 3Mpx méretben. Tesztjeink során a szegmentálást egyszerű, pár kattintással inicializált, régiónövelő módszerrel [22] végeztük, viszont fejlettebb, akár automatikus módszerek használata is lehetséges. A kinyert bináris régió maszkokat majd a módszerünk összeregisztrálta, és a kapott homográfiát felhasználva egyik képet a másikra vetítettük. Három példa is megtekinthető a 6. és 7. ábrákon, ahol mindenik csoportban az első két kép a bemeneti omni kép pár, melyeken a kiszegmentált régió ki van emelve, és a harmadik képen a homográfiával áttranszformált kép éleit vetítettük az első képre. Megfigyelhető, hogy a szegmentálási hibák ellenére, sőt akár bizonyos fokú takarás ellenére is (az első esetben a fa ágai kitakarják a szegmentált régió egy részét) a visszavetített régió élvonalai pontosan illeszkednek az első képre, alig 2-3 pixeles hiba tapasztalható.

5. Relatív helyzet a homográfiából

A *Manhattan világ* feltételezés elég gyakran használt városi vagy beltéri helyszíneket ábrázoló képek esetében [20,21]. Bár elég szigorú megkötések tartalmaz, ezeket nagyjából teljesítik az ember által épített városi környezetek. Egy enyhébb feltételezés az ún. *gyenge Manhattan világ* [5] ami függőleges síkokból áll, tetszőleges orientációval, de párhuzamosan a gravitációs vektorral, és merőlegesen a föld síkjára. Továbbá feltételezzük, hogy a kamera vertikális orientációja is ismert [5], amit akár egy inerciális mérőegység (IMU) segítségével is kiszámolhatunk, ami elég gyakori tartozék a modern kamerákban, okostelefonokban. Míg [5] csak perspektív kamerákkal foglalkozik, mi bemutatjuk, hogy az omnidirekcionális kamerák között becsült homográfia is alkalmazható a kamerák



6. ábra. Homográfia becslési eredmények valós omni képeken. A szegmentált régiók fehérrel megjelölve a képpáron, míg az eredmény a második kép éleit transzformálva és az első képre vetítve látható.



7. ábra. Homográfia becslési eredmények valós omni képeken. A szegmentált régiók fehérrel megjelölve a képpáron, míg az eredmény a második kép élét transzformálva, az első képre vetítve látható.

relatív helyzetének meghatározására, majd szintetikus tesztek által kiértékeljük a módszer teljesítményét.

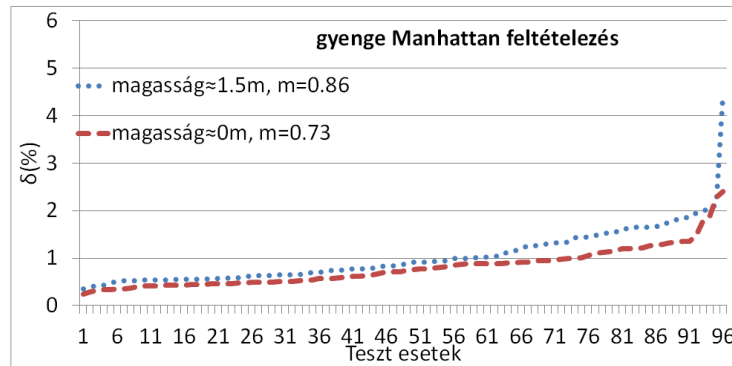
Vegyük a függőleges π síkot és $\mathbf{n} = (n_x, n_y, 0)^T$ normálvektorát (z a függőleges tengely, lásd az 1. ábrán). A sík d távolsága 1-nek választható, mivel a \mathbf{H} homográfia egy szabad skála faktor erejéig van meghatározva. Ismerve a vertikális orientációt, a rotációs mátrix \mathbf{R} az (5) egyenletből csak egy z tengely menti \mathbf{R}_z rotációra redukálódik, így a homográfia felírható:

$$\begin{aligned} \mathbf{H} &= \mathbf{R}_z + (t_x, t_y, t_z)(n_x, n_y, 0)^T \\ &= \begin{pmatrix} \cos(\alpha) + n_x t_x & -\sin(\alpha) + n_y t_x & 0 \\ \sin(\alpha) + n_x t_y & \cos(\alpha) + n_y t_y & 0 \\ n_x t_z & n_y t_z & 1 \end{pmatrix} \\ &= \begin{pmatrix} h_{11} & h_{12} & 0 \\ h_{21} & h_{22} & 0 \\ h_{31} & h_{32} & 1 \end{pmatrix} \end{aligned} \quad (14)$$

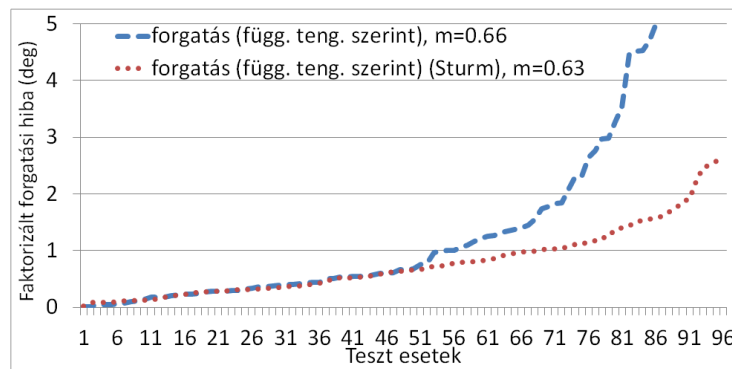
Egy ilyen *gyenge Manhattan* homográfia mátrix hasonló módon megbecsülhető, mint azt az előzőekben bemutatottuk, csak ez esetben a \mathbf{H} utolsó oszlopa $(0, 0, 1)^T$ értékeket vesz fel, így csak 6 szabad paraméterünk van. A módszerünk kiértékelése érdekében erre az esetre is generáltunk 2 szintetikus adathalmazt a *gyenge Manhattan* világ feltételezéseit figyelembe véve. Először a 3D térben elhelyezzük a függőleges síkot, véletlenszerű $[-10, +10]$ fokos elforgatással a függőleges tengely körül, és $\pm[400-800]$ pixeles vízszintes eltolással, ami ekvivalens $[2\text{m}-4\text{m}]$ távolsággal, így biztosítva, hogy a sík látható legyen a kamerából. A második kamera elhelyezésekor $[-10, +10]$ fokos véletlenszerű elforgatást alkalmazunk a függőleges tengely körül, majd $\pm[50-100]\text{cm}$ vízszintes eltolást. A második adathalmaz csak a 3D sík kezdeti z tengely menti magasságában különbözik az elsőtől: így ebben az esetben 150cm-el alacsonyabban lett a sík elhelyezve a z tengely mentén. Az ezeken az adathalmazokon lefuttatott regisztráció pontossága a 8. ábrán megfigyelhető. Amint az elvárt volt, kevesebb szabad paraméterrel növekedett a becslés pontossága (δ hiba kitartóan 2,5% alatti), és ezzel egy időben a számítási idő is csökkent (tipikusan 2-3 másodperc).

A fenti paraméterezésből kiindulva, a homográfiból könnyedén kifaktorizálható az α forgatás és $\mathbf{t} = (t_x, t_y, t_z)^T$ eltolás, vagyis a két kamera közötti relatív helyzet paramétereit, például olyan összefüggéseket felhasználva mint $n_x^2 + n_y^2 = 1$ és $t_z = \pm\sqrt{h_{31}^2 + h_{32}^2}$ (lásd [5] további részletekért).

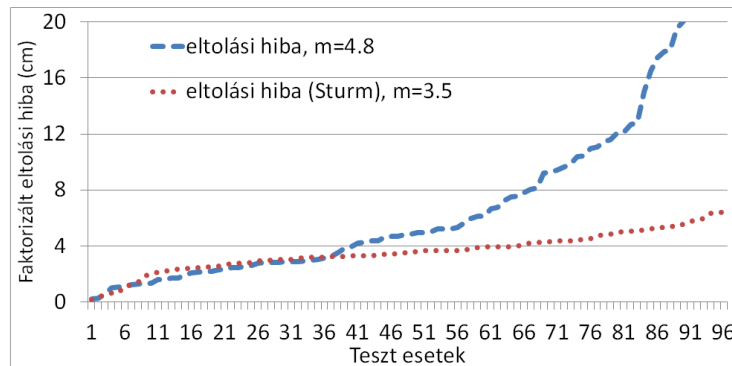
Az említett homográfia faktorizáló módszert használva, a kamera horizontális elforgatását 0.6 fokos pontossággal meghatároztuk, ami a valós elforgatási érték alig 5%-ának felel meg (lásd 9. ábrán). Ami a kamera helyzetének \mathbf{t} eltolás komponensét illeti, kevesebb mint 5cm hibával kinyerhető. Megjegyezzük, hogy az eltolás skála értéke nem meghatározható homográfia faktorizálás során, viszont a szintetikus adathalmaz előállításakor letárolt pontos adatokat felhasználva felskálázhatjuk az egységnyi irányvektort, így az eredeti és a becslült kamera középpontok közötti távolságot értékeljük ki. Ennek eredményei a 10. ábrán láthatóak.



8. ábra. Illesztési (δ) hiba a szintetikus adathalmazon *gyenge Manhattan* megkötéssel.



9. ábra. Vízszintes forgatási hiba a relatív helyzetben (m a medián).



10. ábra. Eltolási hiba a relatív helyzetben (m a medián).

Természetesen klasszikus homográfia faktorizáló módszerek is használhatóak. Például bemutatjuk a Sturm [1] féle SVD felbontás alapú módszer eredményeit ugyanazon az adathalmazon. A 9. és 10. ábra összehasonlítja a két módszer által meghatározott rotáció és eltolás pontosságát. Láthatóan a különbségek elég kicsik, viszont kijelenthető, hogy a Sturm [1] féle módszer stabilabb.

6. Konklúzió

Ebben a cikkben egy új módszert mutattunk be, mellyel a centrális omnidirekcionális kamerák képe közötti homográfia megbecsülhető. A hagyományos módszerekkel ellentétben mi a 3D-s sík régióknak megfelelő szegmentált kép régiókkal dolgozunk, tehát a módszerünk nem használ kulcspont vagy egyéb jellemző kinyerőket. Mivel teljesen az alakzatokra alapul a megközelítésünk, ezért akár multimodális szenzorokkal is használható, amíg a különböző értelmezési tartományokban a megfelelő régiók kiszegmentálhatóak. A homográfia paramétereit direkt módon kapjuk egy nemlineáris egyenletrendszer megoldásaként, aminek a mérete független a bemeneti képektől. Az algoritmus számítási szempontból hatékony, egy optimalizált implementálása akár közel valós időben is futtatható lehet. Különböző szintetikus adathalmazokon végzett számszerű kiértékelése bizonyítja az algoritmus teljesítményét és robusztusságát. Azt is bebizonyítottuk, hogy a becsült homográfia van annyira pontos, hogy megbízhatóan kinyerhessük a kamera külső paramétereit standard faktorizációs módszerek használatával, bizonyos *gyene Manhattan világ* feltételezés mellett.

Köszönetnyilvánítás

A kutatás részben a Domus MTA Magyarország, az Európai Unió és Magyarország támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg a FuturICT.hu (TAMOP-4.2.2.C-11/1/KONV-2012-0013) és TAMOP 4.2.4.A/2-11-1-2012-0001 Nemzeti Kiválóság Program című kiemelt projekt keretében.

Hivatkozások

1. P. Sturm, „Algorithms for plane-based pose estimation,” in *Proc. of International Conference on Computer Vision and Pattern Recognition*, vol. 1, June 2000, pp. 706–711.
2. C. Mei, S. Benhimane, E. Malis, and P. Rives, „Efficient homography-based tracking and 3-D reconstruction for single-viewpoint sensors,” *IEEE Transactions on Robotics*, vol. 24, no. 6, pp. 1352–1364, Dec. 2008.
3. G. Caron, R. Marchand, and E. M. Mouaddib, „Tracking planes in omnidirectional stereovision,” in *Proc. of International Conference on Robotics and Automation*. IEEE, 2011, pp. 6306–6311.
4. A. Makadia, C. Geyer, and K. Daniilidis, „Correspondence-free structure from motion,” *International Journal of Computer Vision*, vol. 75, no. 3, pp. 311–327, Dec. 2007.

5. O. Saurer, F. Fraundorfer, and M. Pollefeys, „Homography based visual odometry with known vertical direction and weak Manhattan world assumption,” in *IEEE/ROCS Workshop on Visual Control of Mobile Robots (ViCoMoR)*, 2012.
6. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.
7. S. K. Nayar, „Catadioptric omnidirectional camera,” in *Proc. of International Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE Computer Society, 1997, pp. 482–488.
8. S. Baker and S. K. Nayar, „A theory of single-viewpoint catadioptric image formation,” *International Journal of Computer Vision*, vol. 35, no. 2, pp. 175–196, 1999.
9. C. Geyer and K. Daniilidis, „A unifying theory for central panoramic systems,” in *Proc. of European Conference on Computer Vision*, 2000, pp. 445–462.
10. B. Mičušík and T. Pajdla, „Para-catadioptric camera auto-calibration from epipolar geometry,” in *Proc. of Asian Conference on Computer Vision*, K.-S. Hong and Z. Zhang, Eds., vol. 2. Seoul, Korea South: Asian Federation of Computer Vision Societies, January 2004, pp. 748–753.
11. D. Scaramuzza, A. Martinelli, and R. Siegwart, „A toolbox for easily calibrating omnidirectional cameras.” in *Proc. of International Conference on Intelligent Robots*. Beijing, October 9–15 2006, pp. 5695–5701.
12. L. Puig and J. J. Guerrero, *Omnidirectional Vision Systems: Calibration, Feature Extraction and 3D Information*. Springer, 2013.
13. J. Kannala and S. S. Brandt, „A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1335–1340, 2006.
14. C. Mei and P. Rives, „Single view point omnidirectional camera calibration from planar grids,” in *Proc. of International Conference on Robotics and Automation*, Roma, Italy, April 2007.
15. D. G. Lowe, „Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
16. L. Puig and J. J. Guerrero, „Scale space for central catadioptric systems: Towards a generic camera feature extractor,” in *Proc. of International Conference on Computer Vision*. IEEE, 2011, pp. 1599–1606.
17. R. Basri and D. W. Jacobs, „Recognition using region correspondences,” *International Journal of Computer Vision*, vol. 25, pp. 141–162, 1996.
18. C. Domokos, J. Nemeth, and Z. Kato, „Nonlinear shape registration without correspondences,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 943–958, 2012.
19. D. Scaramuzza, A. Martinelli, and R. Siegwart, „A flexible technique for accurate omnidirectional camera calibration and structure from motion,” in *Proc. of International Conference on Computer Vision Systems*, ser. ICVS-06. Washington, USA: IEEE Computer Society, 2006, pp. 45–51.
20. J. Coughlan and A. L. Yuille, „Manhattan world: compass direction from a single image by Bayesian inference,” in *Proc. of International Conference on Computer Vision*, vol. 2, 1999, pp. 941–947.
21. Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski, „Manhattan-world stereo,” in *Proc. of International Conference on Computer Vision and Pattern Recognition*. Miami, FL, USA: IEEE Computer Society, 2009, pp. 1422–1429.
22. S.R. Vantaram, and E. Saber, „Survey of contemporary trends in color image segmentation.” *Journal of Electronic Imaging*. 21(4):1–28, 2012.