

# Minimális eljárás fókusz távolság becslésére két affin megfeleltetés segítségével

Baráth Dániel<sup>1</sup>, Tóth Tekla<sup>2</sup>, Hajder Levente<sup>2</sup>

<sup>1</sup> MTA SZTAKI Gépi Érzékelés Kutatócsoport  
barath.daniel@sztaki.mta.hu

<sup>2</sup> ELTE Algoritmusok és Alkalmazásai Tanszék  
hajder@inf.elte.hu

**Absztrakt.** Jelen cikk<sup>1</sup> egy olyan új eljárást mutat be, amelyik sztereó képpárok segítségével félig kalibrált esetben megbecsüli a kamerák fókusz távolságát és a geometriát leíró alapvető mátrixot. A feladat a megoldáshoz szükséges két affin transzformációból képes a becslést elvégezni, ennél kevesebből elvileg sem lehet a feladatot megoldani. Ismereteink szerint a módszer teljesen új, eddig csak pontmegfeleltetéseken alapuló módszerek léteznek. A megoldás többváltozós polinomiális egyenletrendszeren keresztül kapható meg, melyet a rejtett változók technikájával oldunk meg. A kapott gyökök közül a helyes megoldást egy új eljárással választjuk ki, amelyik a kamerák valós paramétereit is figyelembe veszi. A két megfeleltetéssel dolgozó algoritmusunk helyes működését szintetikus és 104 képpárt tartalmazó valós teszten egyaránt igazoljuk.

%

## 1. Introduction

A háromdimenziós világunk rekonstrukciója, kiegészülve a kameraparaméterek becslésével, régóta kutatott terület a háromdimenziós látás területén [14]. A módszerek a kamera paramétereinek becslését tekintve kétfelé oszthatók: vannak a kamera kalibrációk és az autokalibrációk. Ez utóbbi esetben a színterek háromdimenziós rekonstrukciójával párhuzamosan a kamerák belső és külső paramétereit is megbecslik. A mi munkánk is ebbe a kategóriába sorolható, annyi megkötéssel, hogy a kamera belső paramétereit közül csak a fókusz távolság becslésével foglalkozunk. Ezt az esetet hívják félig kalibrált kamerának [20], hiszen a kamera többi paramétereit ismertnek tekintjük: a nyírás nulla, a dőléspont pedig tipikusan a képernyő közepe. Fontos megkötés, hogy a két kamerának azonosak a paramétereit, azaz ugyanolyan kamerával készült a két kép.

A munkák megjelenéséig a hasonló kalibrációs feladatokat pontmegfeleltetéssel oldották meg. Hatpontos algoritmus[20, 31, 33] több is létezik. Ebben a munkában azt szeretnénk megmutatni, hogy a kitűzött feladat két affin megfeleltetésből is kiszámolható.

Háromdimenziós becslési problémák esetében fontos kérdés a robusztusság, azaz a hibás adatok kiszűrése. Háromdimenziós látás esetén komplett algoritmusfolyamokat

<sup>1</sup> A cikk angol nyelvű verziója konferenciakiadványban [7] és folyóiratban [3] egyaránt megjelent.

szokás készíteni [1, 9, 12, 25], melynek egy eleme a rossz adatok kiszűrése, általában a véletlen mintavételezésen alapuló RANSAC algoritmussal [11] vagy valamelyik variánsával. A RANSAC-en belül autokalibrációra az öt- [26] vagy hatpontos módszert [20] szokták javasolni. Minél kevesebb pontot igényel egy módszer, annál kevesebb mintavételre van szükség, ami a futási időt csökkenti, a megbízhatóságot növeli. Ezért lehetőleg minimális pontszámú módszereket szoktak alkalmazni, ahogyan azt mi is tesszük ebben a munkában.

Munkánkban affin megfeleltetéseket alkalmazunk, eltérve a szakma főáramlatától [14], hiszen a legtöbbször kizárólag pontmegfeleltetések felhasználásával készítenek háromdimenziós látó algoritmust. utatási munkánkat a Kató Zoltán-féle kutatócsoport inspirálta, melynek a szerzők is tagjai lehettek [32]. Magukat az affin transzformációkat legjobb tudásunk szerint az elmúlt tíz évben kezdték geometria problémák megoldására használni. Tudomásunk szerint az első munka 2009-ben jelent meg [16] a témában. Azóta felületi normálvektorok becslésére [16, 6], homográfia becslésre [5, 29], epipólus meghatározására [8], kamera póz becslésére [17] vagy akár teljes rekonstrukciós eljárásfolyamra [32, 29] egyaránt használták. Magukat az affin transzformációk előállítását [4, 23] általában ún. affin-kovariáns detektorok segítségével [23] végzik, mint például az Affin-SIFT [24] vagy a Hessian-Affine módszer [22].

Jelen cikkünk az alábbi újdonságokat tartalmazza: Bemutatjuk az affin transzformációk és az epipoláris geometria közötti kapcsolatot. Fókusz távolság becslésére kidolgoztunk egy új módszert, amelyet rejtett változók technikájával [10] oldunk meg. A módszer hatékony, és mindössze két affin megfeleltetés szükséges a végrehajtásához. A kapott gyökök kiválasztásához egy új eljárással dolgoztunk ki, amelyik a kamerák valós fizikai tulajdonságait is kihasználják.

## 2. Előzmények és jelölések

**Epipoláris geometria** A módszerünk bemenetként két kalibrált perspektív képet kap. A két kép belső paramétereit leíró  $\mathbf{K}$  mátrix azonos, azaz ugyanazzal a kamerával készítettük, és a beállítások a két kép között nem változtak meg. A két képet leíró alapvető és lényegi mátrixok <sup>2</sup> elemeit így jelöljük [14]:

$$\mathbf{F} = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix}.$$

Amennyiben a  $\mathbf{K}$  kalibrációs mátrixokat ismerjük, a lényegi mátrixból az alapvető mátrix így kapható:

$$\mathbf{E} = \mathbf{K}^T \mathbf{F} \mathbf{K}. \quad (1)$$

A két képen egymásnak megfeleltetett  $\mathbf{p}_1$  és  $\mathbf{p}_2$  pontok között a jól ismert összefüggést felírhatjuk:

$$\mathbf{p}_2^T \mathbf{F} \mathbf{p}_1 = 0. \quad (2)$$

Alapvető mátrix mindig szinguláris, ezért  $\det(\mathbf{F}) = 0$ . A mátrix elemeit tekintve ez egy harmadfokú megkötést jelent. Miután az  $\mathbf{F}$  mátrix skálázásra érzéketlen a 9 elemre összesen kettővel kevesebb, hét szabadságfok jut.

<sup>2</sup> Az angol fundamental és essential kifejezéseket alapvetőnek és lényeginek magyarosítottuk.

A lényegi mátrix a két kamera közötti kölső elmozdulás alapján írható fel. Ez három eltolás és három forgatás paramétert jelent. Miután a skála itt is bizonytalan, azaz az eltolásnak csak az iránya ismert, a nagysága nem, ezért összesen öt szabadságfok van a mátrixban. Ezt az úgynevezett nyom-megkötés<sup>3</sup> segítségével [20] adhatjuk meg:

$$2\mathbf{E}\mathbf{E}^T\mathbf{E} - \text{tr}(\mathbf{E}\mathbf{E}^T)\mathbf{E} = 0. \quad (3)$$

Bár mind a kilenc elemre egy többváltozós polinomot lehet felírni, valójában az összefüggés csak két független egyenletet jelent.

Félig kalibrált esetben a  $\mathbf{K}$  kalibrációs mátrixból csak az  $f$  fókusz-távolság ismeretlen. A képeken fel lehet úgy venni a koordinátarendszert, hogy az origó éppen a dőfpontban helyezkedjen el. A nyírás nullának tekinthető, ezért félig kalibrált esetben a belső paramétereket tartalmazó mátrix az alábbi alakra egyszerűsödik:  $\mathbf{K} = \mathbf{K}^T = \text{diag}(f, f, 1)$ , ahol  $f$  az ismeretlen fókusz-távolság. A későbbi jelölések egyszerűsítése érdekében vezessük be az alábbi mátrixot:

$$\mathbf{Q} = \text{diag}(1, 1, \tau), \quad \tau = f^{-2}. \quad (4)$$

Ekkor a nyom-megkötést átírhatjuk [18, 28]:

$$2\mathbf{F}\mathbf{Q}\mathbf{F}^T\mathbf{Q}\mathbf{F} - \text{tr}(\mathbf{F}\mathbf{Q}\mathbf{F}^T\mathbf{Q})\mathbf{F} = 0. \quad (5)$$

Ez az összefüggés segíteni fog abban, hogy a fókusz-távot és az  $\mathbf{F}$  alapvető mátrixot két affín megfeleltetés segítségével kiszámítsuk.

**Affín megfeleltetések** A  $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{A})$  vektor-vektor-mátrix hármas két, egymásnak megfelelő pontkoordinátát jelöl, és a körülöttük lévő pixeleket eltranszformáló affín transzformációt. Az  $\mathbf{A}$  mátrix elemeit az alábbi indexekkel jelöljük:

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$$

**Rejtett változók technikája.** A rejtett változók technikája egy algebrai eljárás többváltozós polinomiális rendszerek megoldására. Tegyük fel, hogy  $m$  egyenletünk van, melyek  $n$  darab változó  $(y_1, y_2, \dots, y_n)$  polinomjaiból tevődnek össze. A rejtett változók technikáját akkor alkalmazhatjuk, ha mátrix-vektor szorzássá tudjuk alakítani a rendszert:  $\mathbf{C}(y_1)\mathbf{x} = 0$ , ahol  $\mathbf{C}$  a négyzetes együttható mátrix, mely csak  $y_1$  paramétertől függ,  $\mathbf{x}$  pedig az egyéb változók polinomjait tartalmazza. Ennek a rendszernek  $y_1$  szerinti megoldása a  $\det(\mathbf{C}(y_1)) = 0$  kifejezéssel számítható.

### 3. Fókusz-táv becslés két affín megfeleltetésből

Ez a fejezet azt mutatja meg, hogyan lehet a fókusz-távolságot affín transzformációk segítségével megbecsülni. Először is az affín transzformációk és az alapvető mátrix kapcsolatát mutatjuk be.

#### 3.1. Lokális affín transzformációk felhasználása

<sup>3</sup> Angolul: trace constraint

Tegyük fel, hogy egy  $\mathbf{A}$  affin transzformáció és a megfeleltetéshez tartozó  $\mathbf{p}_1$  és  $\mathbf{p}_2$  pozíciók adottak. Az affin transzformációk legfontosabb tulajdonsága, hogy a pontokon átmenő összes egyenes irányát helyesen megőrzik.

Epipoláris geometria esetén az epipoláris egyenesek irányát ismerjük, ezeket affin transzformáció segítségével fel lehet írni. Ha a két irány  $\mathbf{v}_1$  és  $\mathbf{v}_2$ , írhatjuk [4], hogy:

$$\mathbf{A}\mathbf{v}_1 \parallel \mathbf{v}_2. \quad (6)$$

Ugyanez az összefüggés a normálvektorokra is igaz. A számítógépes grafikából [35] és geometriából jól ismert tényt felhasználva írhatjuk, hogy

$$\mathbf{A}^{-T}\mathbf{n}_1 = \beta\mathbf{n}_2. \quad (7)$$

ahol a  $\beta$  a párhuzamosság skálája,  $\mathbf{n}_1$  és  $\mathbf{n}_2$  az epipoláris egyenesek normálvektorai. Ahogy a függelékben megmutatjuk, a  $\beta$  paraméter pontosan  $-1$  abban az esetben, ha az alapvető mátrixon keresztül számítjuk a jól ismert [14] összefüggés segítségével:  $\mathbf{n}_1 = (\mathbf{1}_1)_{(1:2)} = (\mathbf{F}^T\mathbf{p}_2)_{(1:2)}$  és  $\mathbf{n}_2 = (\mathbf{1}_2)_{(1:2)} = (\mathbf{F}\mathbf{p}_1)_{(1:2)}$ , ahol az alsó index (MATLAB formátumban) azt jelenti, hogy a kapott homogén koordináták koordináták közül az első kettőt kell megtartani, a homogén koordinátát egyszerűen el kell hagyni. Fontos, hogy a vektorokat nem szabad normalizálni!

Ezért az 7. összefüggést így írhatjuk fel:

$$\mathbf{A}^{-T}(\mathbf{F}^T\mathbf{p}_2)_{(1:2)} = -(\mathbf{F}\mathbf{p}_1)_{(1:2)}. \quad (8)$$

Mindez az alapvető mátrix elemeire az alábbi lineáris egyenleteket adja:

$$(u_2 + a_1u_1)f_1 + a_1v_1f_2 + a_1f_3 + (v_2 + a_3u_1)f_4 + a_3v_1f_5 + a_3f_6 + f_7 = 0 \quad (9)$$

$$a_2u_1f_1 + (u_2 + a_2v_1)f_2 + a_2f_3 + a_4u_1f_4 + (v_2 + a_4v_1)f_5 + a_4f_6 + f_8 = 0. \quad (10)$$

Ezért kijelenthetjük, hogy minden egyes affin transzformáció kettő szabadságfokot köt meg az alapvető mátrixból.

### 3.2. Kétpontos algoritmus.

A módszer feltételezi, hogy két affin megfeleltetésünk van:  $(\mathbf{p}_1^1, \mathbf{p}_2^1, \mathbf{A}^1)$  és  $(\mathbf{p}_1^2, \mathbf{p}_2^2, \mathbf{A}^2)$ . Az együtthatómátrix ebben az esetben így néz ki:

$$\mathbf{C}^i = \begin{bmatrix} u_2 + a_1u_1 & a_1v_1 & a_1v_2 + a_3u_1 & a_3v_1 & a_3 & 1 & 0 & 0 \\ a_2u_1 & u_2 + a_2v_1 & a_2 & a_4u_1 & v_2 + a_4v_1 & a_4 & 0 & 1 & 0 \\ u_1u_2 & v_1u_2 & u_2 & u_1v_2 & v_1v_2 & v_2 & u_1 & v_1 & 1 \end{bmatrix}$$

ahol a felső  $i$  index ( $i \in \{1, 2\}$ ) jelöli az affin megfeleltetések számát, és  $\mathbf{x} = [f_1 f_2 f_3 f_4 f_5 f_6 f_7 f_8 f_9]^T$ . A 2, 9, 10 egyenletekből és a  $\mathbf{C}^i\mathbf{x} = 0$  megkötésből jönnek az együtthatók. Két megfeleltetés esetén összevonhatjuk őket egy mátrixba:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^1 \\ \mathbf{C}^2 \end{bmatrix}. \quad (11)$$

A  $\mathbf{C}$  együtthatómátrix mérete  $6 \times 9$ , általános esetben ezért háromdimenziós a nulltere, ebben kell a megoldásunkat keresni

$$\mathbf{x} = \alpha \mathbf{a} + \beta \mathbf{b} + \gamma \mathbf{c} \quad (12)$$

alakban, ahol  $\mathbf{a}$ ,  $\mathbf{b}$  és  $\mathbf{c}$  a nullvektorok, és  $\alpha$ ,  $\beta$ ,  $\gamma$  az ismeretlen súlyok (skalárok).

Ezt behelyettesítve a 5. egyenletbe (nyom-megkötés) tíz harmadfokú polinomiális egyenletet kapunk a  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\tau$  változókra, ahol  $\tau = f^{-2}$  tartalmazza az ismeretlen  $f$  fókusz távolságot.

$\tau$  kiszámítására alkalmazzuk a rejtett változók technikáját: a  $\mathbf{C}(\tau)$  mátrixot  $\tau$ -ban oldjuk meg, Bár  $\alpha$ ,  $\beta$  and  $\gamma$  skálázás erejéig többértelmű, nem rögzítjük a skálát a homogén tulajdonság megőrzése érdekében. A monomial-ok a következők:  $\mathbf{y} = [\alpha^3 \ \alpha^2\beta \ \alpha^2\gamma \ \alpha\beta^2 \ \alpha\beta\gamma \ \alpha\gamma^2 \ \beta^3 \ \beta^2\gamma \ \beta\gamma^2 \ \gamma^3]$

Az 1. táblázat összefoglalja az együtthatómátrixokat. A rendszernek a nemtriviális megoldása akkor adódik, ha

$$\det(\mathbf{C}(\tau)) = 0. \quad (13)$$

Megjegyezzük, hogy megoldásunk lényegi elemeiben hasonlít Li és mtsai. algoritmusához [20]. A determináns  $\tau$ -ra 15-ödfokú polinomot ad, ennek gyökeit kell megvizsgálni. Minden valós gyök esélyes a végső kiválasztásra.  $\alpha$ ,  $\beta$  és  $\gamma$  értékeit  $\mathbf{C}(\tau)$  nullvektorából lehet kiszámolni. A súlyokat behelyettesítve a 12. egyenletbe kapjuk meg az  $\mathbf{F}$  alapvető mátrixot.

$\mathbf{C}(\tau)$	1	2	3	4	5	6	7	8	9	10
	$\alpha^3$	$\alpha^2\beta$	$\alpha^2\gamma$	$\alpha\beta^2$	$\alpha\beta\gamma$	$\alpha\gamma^2$	$\beta^3$	$\beta^2\gamma$	$\beta\gamma^2$	$\gamma^3$
1	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	$c_8$	$c_9$	$c_{10}$
.	.	.	.	.	.	.	.	.	.	.
10	$c_{91}$	$c_{92}$	$c_{93}$	$c_{94}$	$c_{95}$	$c_{96}$	$c_{97}$	$c_{98}$	$c_{99}$	$c_{100}$

1. táblázat:  $\mathbf{C}(\tau)$  együtthatómátrix összetevői, melyeket a nyom-megkötés alkalmazásával kaphatunk meg.

#### 4. A helyes gyök kiválasztása

Ebben a szakaszban azt mutatjuk meg, hogy a kapott 15-ödfokú polinomból kapott gyök közül melyik a helyes. Ehhez elméleti és gyakorlati szempontokat egyaránt figyelembe kell vennünk.

##### 4.1. Érvénytelen fókusz távok kiszűrése

Triviális, hogy komplex számokkal nincsen értelme számolni, ezért a komplex értékeket azonnal ki lehet szűri.

Félig kalibrált kamerák esetén az  $f$  fókusz távolság és az  $\mathbf{F}$  alapvető mátrix segítségével ki lehet a lényegi  $\mathbf{E}$  mátrixot számítani. Ebből négy esetet egyesével megvizsgálva a

két projekciós mátrixokat ( $\mathbf{P}_1$ -et és  $\mathbf{P}_2$ -t) fel lehet írni. Az  $\mathbf{A}$  mátrix és a két pont segítségével  $\mathbf{p}_1$  és  $\mathbf{p}_2$  vetületekből a térbeli  $\mathbf{q} = [x \ y \ z]^T$  pontot ki lehet háromszögelni [14], illetve a felületi  $\mathbf{n} = [n_x \ n_y \ n_z]^T$  normálvektort is meg lehet határozni [6]. Tudjuk, hogy egy felületi pontnak csak az elülső oldalát látjuk, a hátoldalát nem. Ezért az  $\mathbf{n}$  és a  $\mathbf{q} - \mathbf{c}_i$  vektorok által bezárt szög nem lehet derékszögnél nagyobb egyik kamera esetén sem.  $\mathbf{c}_i$  az  $i$ -edik kamera fókuszpontja.

#### 4.2. Lencsék lehetséges fókusz távolságai

A háromdimenziós látásban fókusz távsnak szokás hívni a belső paramétereket tartalmazó  $\mathbf{K}$  mátrix főátlóbeli első két elemét. Ez az elnevezés fizikailag helytelen, hiszen ez valójában a fókusz táv és a pixelméret szorzata. Valós kamera optikája esetén a fizikai megvalósítás korlátja miatt a fókusz táv  $[1 \dots 500]$  mm között van, a pixelméret pedig néhány mikrométer. Ezekből adódik, hogy 100-nál kisebb és 500.000-nél nagyobb értékek nem fordulnak a valóságban elő a belső paraméterek között, a határokon kívül eső értékeket nyugodtan érvénytelennek lehet nyilvánítani.

#### 4.3. Gyök kiválasztása

A fenti gyököket kiszűrve még mindig számtalan hamis gyök lehetséges. Ezeknek a kiszűrésére legvégül a robusztus becslő eljárásokat alkalmazzuk. Miután elég két megfeleltetés a helyes számításhoz, nem elhanyagolható valószínűséggel tudunk két jó megfeleltetést találni. Sokszor ismételve meg tudjuk állapítani, hogy mely megoldások hasonlítanak egymásra. Számtalan robusztus módszer adódik, melyet alkalmazni lehet, a legnépszerűbb RANSAC [11] algoritmustól kezdve a kernel alapú szavazásig [21, 20, 19].

Implementációinkban mi a módkeresés<sup>4</sup> módszert alkalmazzuk [15], ezen a kategórián belül pedig a medián eltolást [30] használjuk, azon belül is Tukey-mediánt [34] alkalmazva.

## 5. Teszteredmények

A szintetikus tesztekhez Matlab implementációt alkalmaztunk, amelyik egyébként az eredeti angol nyelvű cikkben [7] kódszinten megtalálható. A C++ implementáció weboldalunkról letölthető<sup>5</sup>.

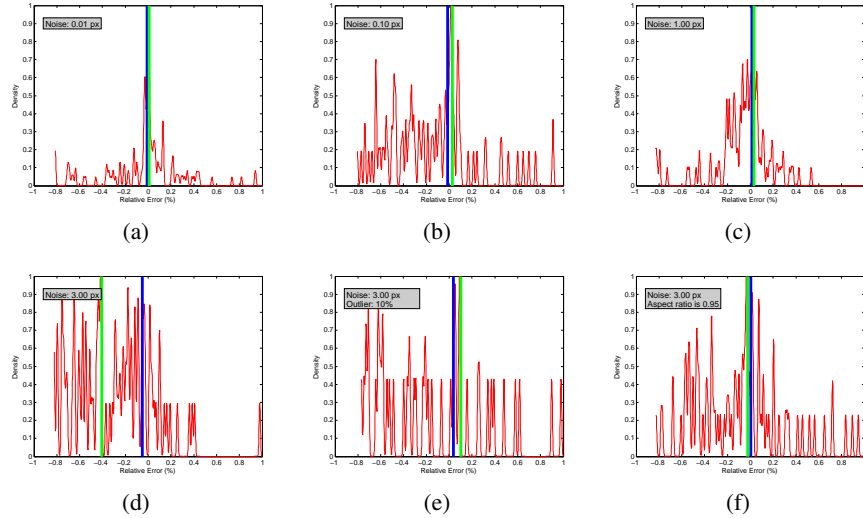
#### 5.1. Szintetikus tesztek

Szintetikus tesztekhez először is két projekciós mátrixot készítettünk. Az első kamera fókuszpontja a  $[0 \ 0 \ 1]^T$ -ben van és az az origóba néz. A második kamera 0.25 távolságra van az elsőtől, véletlen irányban, nézeti iránya szintén véletlenszerű. Öt véletlen síkot generáltunk, melyek átmennek az origón, normálvektoruk véletlen. Mind-egyik sík 50 helyen lett mintavételezve. Ezeket a pontokat a képekre vetítettük, és nulla várható értékű, Gauss-eloszlás szerinti zajt adtunk hozzá. Az affin transzformációk a normálvektor, a pozíciók és a kameraparaméterek segítségével számíthatóak [2].

Az 1. ábrán a kernel eloszlásfüggvénye látszik a relatív zaj függvényében. Az esélyes fókusz távokat az alábbi módszerrel választjuk ki: (1) Két affin megfeleltetést választunk ki, (2) a javasolt módszert alkalmazzuk, (3) az 1-2 lépéseket sokszor ismételjük.

<sup>4</sup> Angulol: mode seeking

<sup>5</sup> <http://web.eec.sztaki.hu/dbarath/>



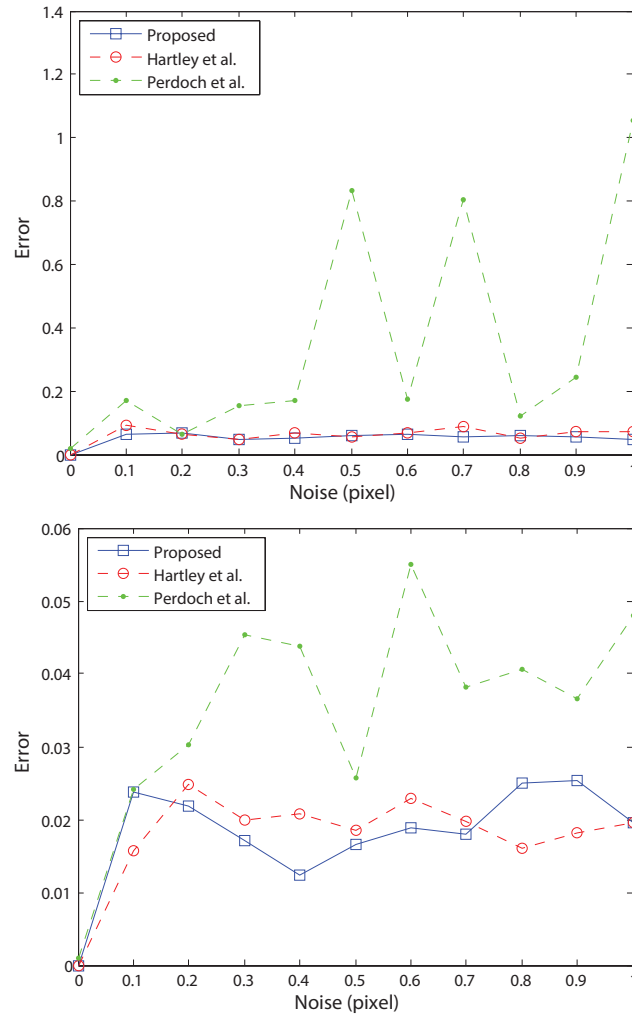
**1. ábra:** A kernel sűrűségfüggvénye (függőleges tengely) a relatív hiba (%) függvényében, 10 szélességű Gauss-kernel esetén. Öt véletlen síkot generáltunk, mindegyiket 20 pontban mintavételeztük. A kék vonal a medián-eltolás, a zöld a kernel szavazás eredményét mutatja.  $\sigma$  jelöli a nulla várható értékű 2D-s zaj szórását, amit a pontmegfeleltetések koordinátáihoz az affin transzformáció elemeihez adtunk hozzá. A zaj mértéke: (a) 0.01 pixel, (b) 0.1 pixel, (c) 1.0 pixel, (d) 3.0 pixel, (e) szintén 3.0 pixel, de 10% outliert is kevertünk az adatok közé, (f) 1.0 pixel és a képerányra is zajt adtunk, azaz a vízszintes és függőleges fókusz nem azonos, az arányuk 19 : 20 A fókusz táv értéke 600 minden esetben. (Az ábra színes nyomtatásban értelmezhető.)

Az ismétlési számot ezernek választottuk. A kék vonal az ábrán a Medián-eltolás eredményét adja, a zöld a kernel szavazását.  $\sigma$  jelöli a zaj szórását az alábbi esetekben: (a) 0.01 pixel, (b) 0.1 pixel, (c) 1.0 pixel, (d) 3.0 pixel, (e) szintén 3.0 pixel, de 10% outliert is kevertünk az adatok közé, (f) 1.0 pixel és a képerányra is zajt adtunk, azaz a vízszintes és függőleges fókusz nem azonos, az arányuk 19 : 20. A valós fókusz a tesztekben mindig 600 volt.

A tesztek alapján megállapítható, hogy a módszerek esetében a csúcs a helyes érték közvetlen környezetében van, továbbá, hogy kernel szavazásos módszernél a javasolt robusztus algoritmus érzékelhetően jobb, különösen akkor, amikor a zajszint magas.

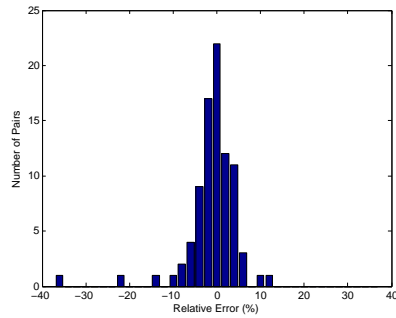
A 2. ábra az alapvető mátrix hibáinak átlagát (felül) és mediánját (alul) mutatja a zaj szórásának ( $\sigma$ ) függvényében. A javasolt módszert Hartley és mtsai.[13], illetve Perdoch és mtsai.[27] algoritmusával hasonlítottuk össze. A hiba mértéke a becslt és az eredeti mátrixok különbségének Frobenius normája. Az eredményeket (átlag és medián) 100 futtatásból kaptuk meg. Jól látható, hogy a javasolt módszer pontossága versenyképer Hartley-ék módszerével, noha lényegesen kevesebb megfeleltetést kell a becsléshez felhasználni.

## 5.2. Tesztelési eredmények valós képpárokon

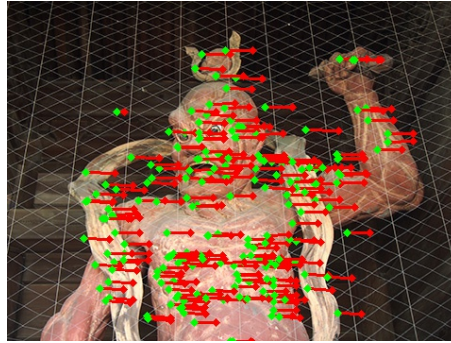


**2. ábra:** 100 futtatás eredményeit összegző grafikon. Az alapvető mátrixok becslésének hibáját a hibamátrix Frobeniusz normájával jellemeztük. A felső ábra az átlaghibát, az alsó a hibák mediánjál jelzi a zaj szórásának függvényében.





(a)



(b)

**3. ábra:** Bal: A fókusz távbecslés hibájából készült hisztogram 104 tesztkép alapján. Jobb: Az első kép az egyik teszpárból. A zöld pontok a megfeleltetések helyei a képpáron, vörös vonal az elmozdulás.

A valós teszteléshez 104 képpárt használtunk fel<sup>6</sup>. A helyes fókusz távokat az EXIF adatokból nyertük ki, lásd például a 4. ábrát. Az affin megfeleltetéseket ASIF algoritmussal [24] kaptuk.

A 3(a). ábrán látható a tesztelesek alapján nyert fókusz távolságok relatív hibája. A kapott eredmények mutatják, hogy a becslés pontossága megfelelő, a hiba egyik esetben sem jelentős, közel vannak az értékek nullához. A 3(b). ábrán a bemeneteként használt jellegetes pontok is láthatóak, a két kép közötti elmozdulással együtt.

A 2. táblázatban a javasolt módszert összehasonlítottuk a Hartley-féle 6 pontos algoritmussal [13], illetve Perdoch és munkatársai módszerével [27]. Ez utóbbi felhasználja az affin transzformációkat pontmegfeleltetések generálására. Az eredmények alapján jól látható, hogy a javasolt kétpontos algoritmus jobban teljesít a konkurens módszereknél, akár az átlag, akár a medián hibát tekintjük.

**2. táblázat:** 100 futtatás eredményeit összegző grafikon. Az alapvető mátrixok becslésének hibáját a hibamátrix Frobenius normájával jellemeztük. Az első eredményoszlop az átlaghibát, a második a hibák mediánját jelzi a zaj szórásának függvényében. Corr # jelöli a szükséges megfeleltetések számát.

Method	Corr #	Avg	Med	$\sigma$
<b>Proposed</b>	2	<b>9.62</b>	<b>3.88</b>	<b>14.08</b>
Perdoch et al. [27]	2	44.66	45.89	26.43
Hartley et al. [13]	6	21.79	8.61	27.48

### 5.3. Az algoritmusok időigénye.

A RANSAC [11] módszert más becslő eljárásokkal is össze lehet hasonlítani, de a javasolt módszer futási ideje várhatóan lényegesen alacsonyabb lesz. A 3. táblázat

<sup>6</sup> <http://www2c.aimet.ne.jp/kawa/photo/ste-idxe.htm>

mutatja meg az elméleti iterációs számaot, ami annál előnyösebb a módszerünkre nézve, minél magasabb az outlier arány.

**3. táblázat:** A RANSAC eljáráshoz szükséges iterációs szám 95%-os valószínűég mellett. Az oszlopok a modellkészítéshez szükséges minimális megfeleltetés számát mutatják. Az outlier-ek számát a sorok adják. Minél több az outlier, annál előnyösebb a kicsi mintavételezési szám.

Outl.	Pontok száma				
	2	5	6	7	8
50%	<b>11</b>	95	191	383	766
80%	<b>74</b>	$\sim 10^3$	$\sim 10^4$	$\sim 10^5$	$\sim 10^6$

## 6. Összefoglalás

Ebben a cikkben bemutattuk, hogy két affin megfeleltetés esetén a félig kalibrált kamerák fókusz távolságát meg lehet becsülni. A kétpontos algoritmust szintetikus és valós adatokon egyaránt teszteltük, és megmutattuk, hogy a jelenlegi módszereknél hatékonyabban képes működni.

Javasoltunk egy robusztus módszert is, amelyik kihasználta, hogy kevés pontból lehet modellt becsülni. A futási idő néhány milliszekundum, ami lényegesen kevesebb, mint a bemeneti adatok, azaz az affin megfeleltetések számítása képpárokon.

## Köszönetnyilvánítás

EFOP-3.6.3-VEKOP-16-2017-00001: Tehetség gondozás és kutatói utánpótlás fejlesztése autonóm járműirányítási technológiák területén – A projekt a Magyar Állam és az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

## Irodalom

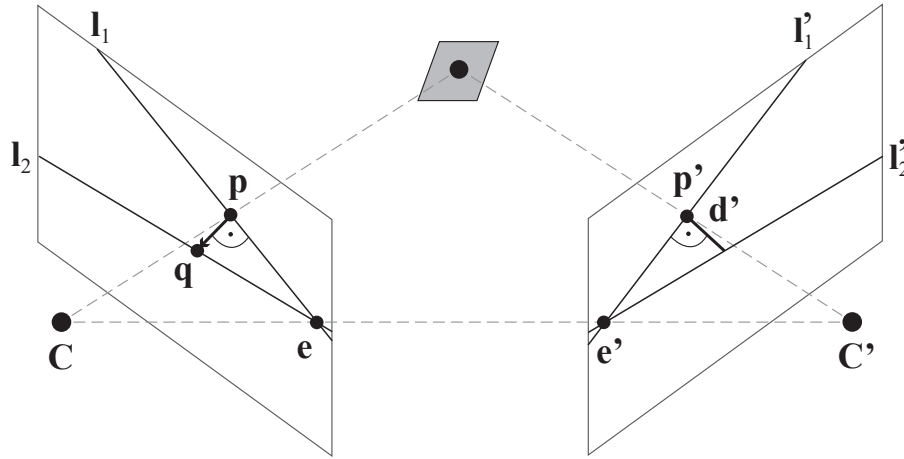
1. S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building rome in a day. *Commun. ACM*, 54(10):105–112, 2011.
2. D. Barath and L. Hajder. Novel ways to estimate homography from local affine transformations. In *Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 434–445, 2016.
3. D. Barath and L. Hajder. Efficient recovery of essential matrix from two affine correspondences. *IEEE Trans. Image Processing*, 27(11):5328–5337, 2018.
4. D. Barath, J. Matas, and L. Hajder. Accurate closed-form estimation of local affine transformations consistent with the epipolar geometry. In *British Machine Vision Conference*, 2016.
5. D. Barath, J. Molnar, and L. Hajder. Novel methods for estimating surface normals from affine transformations. In *Computer Vision, Imaging and Computer Graphics Theory and Applications (Selected and Revised Papers)*, pages 316–337. 2015.



**4. ábra:** A tesztelő képpárok első képei. Zöld és piros pontok: pozíciók az első és a második képen. Vörös vonal: elmozdulás. A helyes fókusztávolságok, a 6-pontos algoritmus [13] eredménye és a javasolt módszer becslése a szürke háromszögön belül olvasható.

6. D. Barath, J. Molnar, and L. Hajder. Optimal Surface Normal from Affine Transformation. In *Proceedings of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 305–316, 2015.
7. D. Barath, T. Toth, and L. Hajder. A minimal solution for two-view focal-length estimation using two affine correspondences. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 2557–2565, 2017.
8. J. Bentolila and J. M. Francos. Conic epipolar constraints from affine correspondences. *Computer Vision and Image Understanding*, 122:105–114, 2014.
9. A. Bódis-Szomorú, H. Riemenschneider, and L. V. Gool. Fast, approximate piecewise-planar modeling based on sparse structure-from-motion and superpixels. In *CVPR*, 2014.
10. D. A. Cox, J. Little, and D. O’Shea. *Using algebraic geometry*. 2006.
11. M. Fischler and R. Bolles. RANdom SAMpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 1981.
12. J. Frahm, P. F. Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y. Jen, E. Dunn, B. Clipp, and S. Lazebnik. Building rome on a cloudless day. In *11th European Conference on Computer Vision*, pages 368–381, 2010.
13. R. I. Hartley and H. Li. An efficient hidden variable approach to minimal-case camera motion estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(12):2303–2314, 2012.
14. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
15. A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM Comput. Surv.*, 31(3):264–323, 1999.
16. K. Köser. *Geometric Estimation with Local Affine Frames and Free-form Surfaces*. Shaker, 2009.
17. K. Köser and R. Koch. Differential spatial resection - pose estimation using a single local image feature. In *IEEE Proceedings of the European Conference on Computer Vision*, 2008.
18. Z. Kukelova, M. Bujnak, and T. Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *Proceedings of the British Machine Vision Conference*, 2008.
19. Z. Kukelova, T. Pajdla, and M. Bujnak. *Algebraic methods in computer vision*. PhD thesis, Center for Machine Perception, Czech Technical University, Prague, Czech republic, 2012.
20. H. Li. A simple solution to the six-point two-view focal-length problem. In *IEEE Proceedings of the European Conference on Computer Vision*, 2006.
21. H. Li and R. Hartley. A non-iterative method for correcting lens distortion from nine-point correspondences. In *In Proc. OmniVision’05, ICCV-workshop*, 2005.
22. K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *IEEE Proceedings of the European Conference on Computer Vision*, 2002.
23. K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *IEEE Proceedings of the International Journal of Computer Vision*, 2005.
24. J. Morel and G. Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM J. Imaging Sciences*, 2(2):438–469, 2009.
25. P. Moulon, P. Monasse, and R. Marlet. Global fusion of relative motions for robust, accurate and scalable structure from motion. In *International Conference on Computer Vision, ICCV 2013*, pages 3248–3255, 2013.
26. D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–777, 2004.
27. M. Perdoch, J. Matas, and O. Chum. Epipolar geometry from two correspondences. In *ICPR*, 2006.
28. Á. Pernek and L. Hajder. Automatic focal length estimation as an eigenvalue problem. *Pattern Recognition Letters*, 34(9):1108–1117, 2013.
29. C. Raposo and J. P. Barreto. Theory and practice of structure-from-motion using affine correspondences. In *IEEE Proceedings on Computer Vision and Pattern Recognition*, 2016.

30. L. Shapira, S. Avidan, and A. Shamir. Mode-detection via median-shift. In *IEEE Proceedings of the International Conference on Computer Vision*, 2009.
31. H. Stewénius, D. Nistér, F. Kahl, and F. Schaffalitzky. A minimal solution for relative pose with unknown focal length. *Image and Vision Computing*, 2008.
32. A. Tanács, A. Majdik, L. Hajder, J. Molnár, Z. Sánta, and Z. Kato. Collaborative mobile 3d reconstruction of urban scenes. In *Computer Vision - ACCV 2014 Workshops - Singapore, Singapore, November 1-2, 2014, Revised Selected Papers, Part III*, pages 486–501, 2014.
33. A. Torii, Z. Kukelova, M. Bujnak, and T. Pajdla. The six point algorithm revisited. In *IEEE Proceedings of the Asian Conference on Computer Vision*, 2010.
34. J. W. Tukey. Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians*, 2:523–531, 1975.
35. K. Turkowski. Transformations of surface normal vectors. In *Technical Report 22, Apple Computer*, 1990.



**5. ábra:** Egy mintadarabka két vetülete. A skálára lehet megkötést adni,  $|p - q|$  és  $d'$  aránya adja meg a skálázást az  $\mathbf{A}^{-T} \mathbf{n}$  és az  $\mathbf{n}'$  vektorok között.

## 7. A lineáris affin megkötés bizonyítása

**Lemma 1** (Megkötés az epipoláris egyenesek normálvektoraira). *Adott egy  $\mathbf{A}$  affin transzformáció, amelyik a két képen a megfelelő pontok körüli "végtelenül közel" levő mintákat transzformálja egymásba. Az epipoláris egyeneseknek megfelelő normákat  $\mathbf{n}_1$ -el és  $\mathbf{n}_2$ -vel jelöljük. Az  $\mathbf{A}$  affin transzformáció akkor és csak akkor érvényes, ha  $\mathbf{A}^{-T} \mathbf{n}_1 = -\mathbf{n}_2$ .*

*Bizonyítás.* Triviális, hogy  $\mathbf{A}$  mátrix az epipoláris egyenesek irányvektorait egymásba transzformálja:  $\mathbf{A} \mathbf{v} \parallel \mathbf{v}'$ , ahol  $\mathbf{v}$  és  $\mathbf{v}'$  a két egymásnak megfelelő egyenes irányvektora.

Számítógépes grafikából [35] jól ismert tény, hogy a normálokra igaz, hogy  $\mathbf{A}^{-T} \mathbf{n} = \beta \mathbf{n}'$ , ahol  $\mathbf{n} = (\mathbf{F}^T \mathbf{p}')_{1:2}$  és  $\mathbf{n}' = (\mathbf{Fp})_{1:2}$  a két képen az egyeneseknek megfelelő normálvektorok. Az  $(1 : 2)$  index az első két sort jelenti, a MATLAB programnyelvéhez nagyon hasonló jelölést alkalmazva.

Az egymásnak megfelelő két pontmegfeleltetést jelölje  $\mathbf{p} = [x, y, 1]^T$  és  $\mathbf{p}' = [x', y', 1]^T$ . Legyen  $\mathbf{n}_1 = [\mathbf{n}_{1,x} \ \mathbf{n}_{1,y}]^T$  és  $\mathbf{n}'_1 = [\mathbf{n}'_{1,x} \ \mathbf{n}'_{1,y}]^T$  az epipoláris egyeneseknek a normálvektora  $\mathbf{l}_1 = \mathbf{F}^T \mathbf{p}' = [l_{1,a} \ l_{1,b} \ l_{1,c}]^T$  és  $\mathbf{l}'_1 = \mathbf{Fp} = [l'_{1,a} \ l'_{1,b} \ l'_{1,c}]^T$ . Felírhatjuk, hogy  $\mathbf{A}^{-T} \mathbf{n}_1 = \beta \mathbf{n}'_1$ , mivel  $\mathbf{A} \mathbf{v} \parallel \mathbf{v}'$ , ahol  $\beta$  egy (egyelőre) ismeretlen valós skála.

Először megmutatjuk, hogyan transzformálja  $\mathbf{A}$  mátrix az  $\mathbf{n}_1$  vektor hosszát, Vezessük be az eredeti  $\mathbf{q}$  ponthoz igen közeli  $\mathbf{q} = \mathbf{p} + \delta \mathbf{n}_1$  pontot, ahol  $\delta$  egy kicsi valós szám. Ez az új  $\mathbf{q}$  pont meghatároz egy epipoláris egyenest a másik síkon: az egyenes implicit paramétereit az  $\mathbf{l}'_2 = \mathbf{Fq} = \mathbf{F}(\mathbf{p} + \delta \mathbf{n}_1) = [l'_{2,a} \ l'_{2,b} \ l'_{2,c}]^T$  vektor adja. A  $\beta$  skálát a  $d'$  távolság adja meg, amely a  $\mathbf{p}'$  és az  $\mathbf{l}'_2$  egyenes közötti távolság, lásd a 6. ábrát.  $d'$  távolságot az alábbi összefüggés adja:

$$d' = \frac{|s_{1,a}x' + s_{2,b}y' + s_{3,c}|}{\sqrt{s_{1,a}^2 + s_{2,b}^2}}, \quad (14)$$

$$s_{i,k} = l'_{1,k} + \delta f_{i1} \mathbf{n}_{1,x} + \delta f_{i2} \mathbf{n}_{1,y},$$

$$i \in \{1, 2, 3\}, k \in \{a, b, c\}$$

A  $\mathbf{p}'$  pont az  $\mathbf{l}'_1$  egyenesen fekszik, azaz  $l'_{1,a}x' + l'_{1,b}y' + l'_{1,c} = 0$ . Ez a tény egyszerűsíti a 14. összefüggést:

$$d' = \frac{|\hat{s}_1 u' + \hat{s}_2 v^2 + \hat{s}_3|}{\sqrt{\hat{s}_1^2 + \hat{s}_2^2}}, \quad (15)$$

ahol  $\hat{s}_i = \delta f_{i1} \mathbf{n}_{1,x} + \delta f_{i2} \mathbf{n}_{1,y}$ ,  $i \in \{1, 2, 3\}$ .  $\beta$  meghatározásához a  $\mathbf{q}$  pontot végtelenül közel kell vinni  $\mathbf{p}$ -hez, azaz  $\delta \rightarrow 0$ .  $\beta$  négyzetére írhatjuk, hogy  $\beta^2 = \lim_{\delta \rightarrow 0} \frac{\delta^2}{d'^2} = \lim_{\delta \rightarrow 0} \frac{s_1^2 + s_2^2}{|\hat{s}_1 u' + \hat{s}_2 v' + \hat{s}_3|^2}$ . Elemi algebrai műveletek segítségével  $\beta$  átalakítható:  $\beta = \sqrt{l'_{1,a} l'_{1,a} + l'_{1,b} l'_{1,b} / (|\tilde{s}_1 x' + \tilde{s}_2 y' + \tilde{s}_3|)}$ , ahol  $\tilde{s}_i = f_{i1} \mathbf{n}_{1,x} + f_{i2} \mathbf{n}_{1,y}$ ,  $i \in \{1, 2, 3\}$ .

A normálvektorok hosszát tegyük egységnyivé. Ez az alábbi összefüggést eredményezi:

$$\mathbf{A}^{-T} \frac{\mathbf{n}}{|\mathbf{n}|} = \beta \mathbf{n}'. \quad (16)$$

A  $\mathbf{p}$  pontnak megfelelő epipoláris egyenes legyen  $[l'_{1,a}, l'_{1,b}, l'_{1,c}] = \mathbf{F}[x, y, 1]^T$ . Ekkor a normálvektor ebben az alakban írható:  $\mathbf{n}' = [l'_{1,a} \ l'_{1,b}]^T = (\mathbf{F} [x' \ y' \ 1]^T)_{(1:2)}$ . Hasonlóan:  $\mathbf{n} = (\mathbf{F}^T [x' \ y' \ 1]^T)_{(1:2)}$ . A 16. egyenletben a normával való osztás miatt  $|\mathbf{n}| = \sqrt{l_{1,a}^2 + l_{1,b}^2}$ . A számláló pedig:

$$\begin{aligned} \tilde{s}_1 u' + \tilde{s}_2 v' + \tilde{s}_3 &= \\ \mathbf{n}_{1,u}(f_{11}u' + f_{21}v' + f_{31}) + \mathbf{n}_{1,v}(f_{12}u' + f_{22}v' + f_{32}) &= \\ \mathbf{n}_{1,u}^2 + \mathbf{n}_{1,v}^2 &= |\mathbf{n}_1|^2. \end{aligned}$$

Tehát  $\beta = \pm |\mathbf{n}_1| / |\mathbf{n}_1|^2 = \pm 1 / |\mathbf{n}_1|$ . Ezért a 16. egyenlet a következőképpen módosul:  $\mathbf{A}^{-T} \mathbf{n} = \pm \mathbf{n}'$ . Az előjelek közül csak a negatív érvényes, hiszen az epipoláris egyenesek irányvektorai egymással ellentétesek a két képen. Ezért a végső összefüggés:  $\mathbf{A}^{-T} \mathbf{n} = -\mathbf{n}'$ .  $\square$