

Epipoláris Geometriával Konzisztens, Legközelebbi Affin Transzformáció Optimális Becslése

Baráth Dániel¹, Jiri Matas², Hajder Levente¹

¹ Gépi Érzékelés Laboratórium, MTA SZTAKI
{barath.daniel,hajder.levente}@sztaki.mta.hu

² Center for Machine Perception, Department of Cybernetics, CTU Prague
matas@cmp.felk.cvut.cz

Kivonat Ebben a tanulmányban egy olyan módszert mutatunk be, amely kiszámítja azt a lokális affin transzformációt, amelyik a legközelebb van egy kezdeti transzformációhoz és konzisztens az epipoláris geometriával. A számítás a négyzetes normát minimalizálja. Azt is megmutatjuk, hogy ennek a normának az alkalmazása geometriai jelentéssel is bír. A bemutatott megoldás legkisebb négyzetes értelemben optimális, az eredmény kiszámítása lineáris egyenletrendszer segítségével történik, ezért rendkívül gyors.

A kapott transzformációkat különböző becslési feladatokban lehet alkalmazni. Szintetikus és valós tesztek egyaránt bemutatunk. A legfontosabb vizsgálatunk a jellegzetes mintákat megfelelő algoritmusok összehasonlítása, ahol a javított affin transzformációk kivétel nélkül jobb eredményt adnak, mint amikor az eredeti transzformációkkal számolnak a módszerek.

1.. Bevezetés

A számítógépes látás egyik alapvető problémája különböző képeken az összefüggő részek megtalálása. Amennyiben nem csak pontmegfeleltetéseket, hanem a pontok körüli foltokat is figyelembe vesszük, a háromdimenzióban látó algoritmusaink pontosabb, részletgazdagabb eredményeket képesek produkálni. A foltok közötti transzformációt többféleképpen lehet leírni, az egyik legkézenfekvőbb megadás az affin transzformáció.

Jelen cikk¹ fő célkitűzése, hogy megmutassa, hogyan lehet statikus háromdimenziós színterekben a lokális affin transzformációt nagyon pontosan megbecsülni. Matas és mtsai. [1] már 2002-ben megmutatták, hogyan lehet az affin transzformációval javítani a sztereo látást. Köser és Koch [2] bebizonyította, hogy a

¹ A cikk angol nyelvű változata a BMVC2016 konferencián jelent meg: Daniel Barath, Jiri Matas and Levente Hajder. Accurate Closed-form Estimation of Local Affine Transformations Consistent with the Epipolar Geometry. In Proceedings of the 27th British Machine Vision Conference. 2016.

kamera helyét és nézeti irányát egyaránt meg lehet becsülni, ha az affin transzformáció mellett a pontmegfeleltetések is adottak. Köser [3] azt is megmutatta, hogy a trianguláció (pont mélységbecslése) is javítható a lokális transzformáció megléte esetén. Bentolila és mtsai. [4] bebizonyították, hogy az affin transzformáció kalibrálatlan sztereo kamera esetén egy másodfokú görbére szűkíti az epipólus helyét. A legújabb háromdimenziós rekonstrukciós eljárások [5,6,7] szintén ki tudják használni az affinitásokból származó többlet-információt, és ezzel még valószínűbb eredményeket képesek előállítani.

A szerzők saját, friss eredménye [8] azt is megmutatja, hogy már egyetlen lokális affinitás segítségével is meg lehet becsülni a perspektív kamerán meglévő sík-sík homográfiát. Korábban azt is megmutattuk [9], hogy a felületi normális és az affin transzformáció kölcsönösen egyértelműen megfelelnek egymásnak.

A cikk fő üzenete, hogy a megbecsült affin transzformációhoz meg lehet találni azt a legközelebbi transzformációt, amelyik megfelel az epipoláris geometriának. Az epipoláris geometriát a fundamentális mátrix írja le, amelyik az affin transzformációkból [4,7] és a pontmegfeleltetésekből [10] egyaránt kiszámítható.

Hartley és Sturm már a 90-es évek végén megmutatta, hogy adott pontmegfeleltetésekhez legközelebbi, epipoláris geometriának megfelelő pontokat optimalisan meg lehet találni. A megoldás alapja, hogy a javított pontok egymásnak megfelelő epipoláris egyenesen helyezkedjenek el. A legkisebb négyzetes értelemben vett optimális megoldást egy hatodfokú polinom egyik gyökeként lehet meghatározni. A jelen cikkben ismertetett módszert ezen eljárás kiterjesztésén is fel lehet fogni: mi az affin transzformáció pontosítására adunk optimális megoldást.

A lokális affin transzformáció előállítását, detektálását végző módszereket tipikusan három kategóriába szokás sorolni: az első kategória közvetlenül állítja elő a transzformációt. Ebből a kategóriából a legismertebb módszer az MSER [11]. A második kategóriába olyan algoritmusok tartoznak, melyek a kezdeti becslést optimalizálják - ide tartoznak a Harris-Affine [12] és a Hessian-Affine [13] módszerek, melyek a Baumberg iterációt [14] alkalmazzák a jó minőségű transzformációk elérése érdekében. Az utolsó kategória módszerei szintetikus torzítják a képet, és a torzított képekre alkalmazzák az egyes detektorokat. A legismertebb ilyen módszer az Affine SIFT (ASIFT) [15], amelyik affin transzformációval torzítja a képet, és a SIFT detektort alkalmazza az eltranszformált képekre. Más detektort is lehet társítani a képtorzításhoz, mint például a SURF [16], ORB [17] és BRISK [18] detektorok. Szintén ilyen eljárás a MODS [19] (Matching On Demand with view Synthesis) algoritmus, amely által visszaadott pontalmaz MSER, ORB és Hessian-Affine pontok keveréke. Ezen felül a módszer paramétere a kimenetként adott pontok elvárt száma, így a lehető legkevesebb nézet-szintézissel² képes annyi megfeleltetést detektálni, amennyi szükséges.

A cikkünk fő hozzájárulásait az alábbiakban foglalhatjuk össze: két kényszert vezetünk be, melyek az epipoláris geometriával konzisztens viselkedést garantálják. Ezen kívül javasunk egy algoritmust, amelyik az EG- L_2 -Optimális (EG-

² A nézetszintézis nagyon költséges művelet, ezért fontos, hogy minél kevesebb nézetet állítsanak elő a módszerek.

L_2 -Opt) affin transzformációt legkisebb négyzetes értelemben optimálisan kiszámolja a két kényszer segítségével. Azt is bebizonyítjuk, hogy az optimalizálandó hibafüggvényünk geometriai és algebrai interpretációval egyaránt rendelkezik.

Tesztelési eredményeink azt is megmutatják, hogy az EG- L_2 -Opt eljárás javítja az affin detektorok minőségét. Érdekes mellékeredménye a kutatásunknak, hogy az affin detektorokat tökéletes (ground truth) adatokon kvantitatívan összehasonlítottuk.

2.. EG- L_2 -Optimális Lokális Affin Transzformáció

Ebben a fejezetben először arról írunk, hogyan lehet az affin transzformációkat a megadott pontpárokból megbecsülni. Ezután a fundamentális mátrix által meghatározott kompatibilitási kényszereket határozzuk meg. Végezetül az EG- L_2 -Optimális transzformációt kiszámítását mutatjuk be.

Lokális affin transzformáció. Még ma is nyitott kérdés, hogy valós környezetben hogyan lehet jó minőségű affin transzformációkat képpárokból kinyerni. Mi azt javasoljuk, hogy affin-kovariáns detektorokat [13] alkalmazzanak a feladatra. Ezek a detektorok a pontmegfeleltetéseket és az affinitásokat egyaránt előállítják. Az általunk javasolt, jelenleg korszerűnek mondható lehetőségek az alábbiak: ASIFT [15], MODS [19], Harris-Affine [20], Hessian-Affine [20]. Ezek a detektorok egy affin transzformációt előállítanak minden i -dik $\mathbf{p}_k^i = [x_k^i \ y_k^i]^T$ ($i \in [1, n]$) pontra a k -dik képen. Az első és második képen levő \mathbf{A}_1^i és \mathbf{A}_2^i affin transzformációkból az eredő \mathbf{A}^i transzformáció így számítható ki:

$$\mathbf{A}^i = \mathbf{A}_2^i (\mathbf{A}_1^i)^{-1}. \quad (1)$$

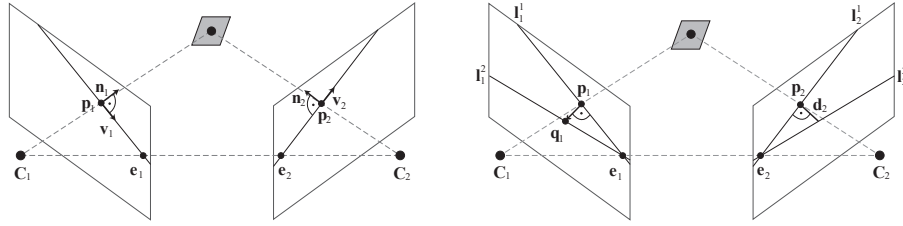
Affin Kompatibilitás – Ertolás. Az első feladat a pontmegfeleltetések korrigálása az epipoláris geometria szerint. Hartley és Sturm [21] megmutatta, hogy legkisebb négyzetes értelemben ez a feladat optimálisan megoldható. A módszerünk a képtérben minimalizálja az euklidészi távolságát az eredeti és a javított pozícióknak. Az eredmény az epipoláros geometria törvényszerűségeinek köszönhetően meg fog felelni.

Affin Kompatibilitás – Orientáció. Ebben a szakaszban az \mathbf{A} affin transzformáció alatt az eredeti transzformáció 2×2 -es bal részmátrixát értjük, mellyel a foltok deformációját lehet leírni.

Tegyük fel, hogy a sztereo képpár között az \mathbf{F} fundamentális mátrix adott, továbbá ismerjük a két képen a \mathbf{p}_1 és \mathbf{p}_2 egymásnak megfelelő pozíciókat. Az \mathbf{A} affin transzformáció az egyenesek irányát az adott pontban helyesen viszi át, ezért igaz, hogy az epipoláris egyenes irányait is jól transzformálja. Jelöljük a két epipoláris egyenest \mathbf{l}_1 -gyel és \mathbf{l}_2 -vel, irányvektoraikat \mathbf{v}_1 -vel és \mathbf{v}_2 -vel. Az affin transzformáció átvizsgálja az egyenes irányait, ezért $\mathbf{A}\mathbf{v}_1 \parallel \mathbf{v}_2$. A számítógépes grafikában [22] jól ismert és gyakran alkalmazott törvényszerűség, hogy

$$\mathbf{A}^{-T} \mathbf{n}_1 = \beta \mathbf{n}_2, \quad (2)$$

ahol \mathbf{n}_k az epipoláris egyenes normálvektora, β a skála az $\mathbf{A}^{-T} \mathbf{n}_1$ és \mathbf{n}_2 vektorok között. A konfiguráció a 1(a). ábrán látható.



(a) Az irányra vonatkozó kompatibilitási kényszer azt mondja ki, hogy $\mathbf{A}\mathbf{v}_1 \parallel \mathbf{v}_2$. Ez ekvivalens $\mathbf{A}^{-T}\mathbf{n}_1 \parallel \mathbf{n}_2$ -vel.

(b) A skálára vonatkozó kompatibilitási kényszer szerint a $\|\mathbf{p}_1 - \mathbf{q}_1\|_2$ és d_2 vektorok hossza közötti arány határozza meg a skálát az epipoláris egyenesekre merőleges irányok között.

1. ábra. Irányra és skálára vonatkozó kompatibilitási kényszerek. Az \mathbf{A} mátrix az affin transzformációt, a \mathbf{v}_k és \mathbf{n}_k vektorok az epipoláris egyenes irány- és normálvektorait jelöli azon az epipoláris egyenesen, melyen a \mathbf{p}_k pont fekszik a k -dik képen ($k \in \{1, 2\}$).

Affin Kompatibilitás – Skála. Azt mutatjuk meg ebben a szakaszban, hogy a β skála az $\mathbf{A}^{-T}\mathbf{n}_1$ és \mathbf{n}_2 vektorok hossza között hogyan határozható meg.

Tegyük fel, hogy az egymásnak megfelelő $\mathbf{p}_1 = [x_1 \ y_1 \ 1]^T$ és $\mathbf{p}_2 = [x_2 \ y_2 \ 1]^T$ homogén koordinátás alakú pontok adóttak. Az $\mathbf{n}_1 = [n_1^x \ n_1^y]^T$ és $\mathbf{n}_2 = [n_2^x \ n_2^y]^T$ vektorok jelölik az epipoláris egyenesek normálirányát. A két epipoláris egyenest az $\mathbf{l}_1^1 = \mathbf{F}^T \mathbf{p}_2 = [l_1^{1,a} \ l_1^{1,b} \ l_1^{1,c}]^T$ és $\mathbf{l}_2^1 = \mathbf{F} \mathbf{p}_1 = [l_2^{1,a} \ l_2^{1,b} \ l_2^{1,c}]^T$ összefüggések adják meg. A feladat azt meghatározni, hogy az \mathbf{A} mátrix hogyan változtatja meg az \mathbf{n}_1 vektor hosszát. Ezt nevezzük skálának. Ennek a kiszámításához vezessünk be egy új pontot: $\mathbf{q}_1 = \mathbf{p}_1 + \gamma \mathbf{n}_1$, ahol γ egy tetszőleges skalár. Ez az új pont egy másik $\mathbf{l}_2^2 = [l_2^{2,a} \ l_2^{2,b} \ l_2^{2,c}]^T$ epipoláris egyenest határoz meg a második képen: $\mathbf{l}_2^2 = \mathbf{F} \mathbf{q}_1 = \mathbf{F}(\mathbf{p}_1 + \gamma \mathbf{n}_1)$.

A β skála a $d_1 = \|\mathbf{p}_1 - \mathbf{q}_1\|_2$ és d_2 távolságok hányadosa, ahol d_2 az \mathbf{l}_2^2 egyenes és \mathbf{p}_2 pont közötti távolság. A feladathoz a 1(b). ábrán láthatunk magyarázó rajtot. A d_2 számításának kifejtése a 3. ábrán található.

$$d_2 = \frac{|(l_2^{1,a} + \gamma f_{11} n_1^x + \gamma f_{12} n_1^y)x_2 + (l_2^{1,b} + \gamma f_{21} n_1^x + \gamma f_{22} n_1^y)y_2 + l_2^{1,c} + f_{31} n_1^x + f_{32} n_1^y|}{\sqrt{(l_2^{1,a} + \gamma f_{11} n_1^x + \gamma f_{12} n_1^y)^2 + (l_2^{1,b} + \gamma f_{21} n_1^x + \gamma f_{22} n_1^y)^2}} \quad (3)$$

Tudjuk, hogy a \mathbf{p}_2 pont az \mathbf{l}_2^1 egyenesen fekszik. Koordinátákkal kifejtve: $l_2^{1,a} x_2 + l_2^{1,b} y_2 + l_2^{1,c} = 0$. Ennek a ténynek a felhasználásával a 3. egyenlet egyszerűsödik:

$$d_2 = \frac{|(\gamma f_{11} n_1^x + \gamma f_{12} n_1^y)x_2 + (\gamma f_{21} n_1^x + \gamma f_{22} n_1^y)y_2 + f_{31} n_1^x + f_{32} n_1^y|}{\sqrt{(l_2^{1,a} + \gamma f_{11} n_1^x + \gamma f_{12} n_1^y)^2 + (l_2^{1,b} + \gamma f_{21} n_1^x + \gamma f_{22} n_1^y)^2}} \quad (4)$$

Ahhoz, hogy β -t meghatározzuk, a vizsgált \mathbf{q}_1 pontot végtelenül közel kell vinni \mathbf{p}_1 -hez, azaz $\gamma \rightarrow 0$. Ez így módosítja a fenti egyenletet:

$$\beta^2 = \lim_{\gamma \rightarrow 0} \frac{\gamma^2}{d_2^2} = \lim_{\gamma \rightarrow 0} \frac{((l_2^{1,a} + \gamma f_{11} n_1^x + \gamma f_{12} n_1^y)^2 + (l_2^{1,b} + \gamma f_{21} n_1^x + \gamma f_{22} n_1^y)^2)}{|(f_{11} n_1^x + f_{12} n_1^y)x_2 + (f_{21} n_1^x + f_{22} n_1^y)y_2 + f_{31} n_1^x + f_{32} n_1^y|^2} \quad (5)$$

Elemi átalakítások utána fenti formula a β skálára így módosul:

$$\beta = \frac{\sqrt{l_2^{1,a} l_2^{1,a} + l_2^{1,b} l_2^{1,b}}}{|s_1 x_2 + s_2 y_2 + s_3|}, \quad s_i = f_{i1} n_1^x + f_{i2} n_1^y, \quad i \in \{1, 2, 3\}. \quad (6)$$

Ez a végső összefüggést, ezt fogjuk a későbbiekben használni.

Az EG- L_2 -Optimális Affin Transzformáció Meghatározása. Tegyük fel, hogy egy megfigyelt \mathbf{A}' lokális affin transzformáció adott. Jelöljük az elemeit a következő módon:

$$\mathbf{A}' = \begin{bmatrix} a'_1 & a'_2 \\ a'_3 & a'_4 \end{bmatrix}. \quad (7)$$

A módszer levezetésének az alapja, hogy találjunk egy \mathbf{A} mátrixot, amelyre a

$$\|\mathbf{A} - \mathbf{A}'\|_2^2 \quad (8)$$

kifejezés minimális, és $\mathbf{A}^{-T} \mathbf{n}_1 = \beta \mathbf{n}_2$ (Lásd a 2. egyenletet). Az invertálás elkerülése érdekében át lehet alakítani ezt a formulát: $\mathbf{n}_1 = \beta \mathbf{A}^T \mathbf{n}_2$. (Az L_2 norma alkalmazását a 3. szakaszban igazoljuk.)

A β skálára a 6. kifejezést alkalmazzuk. Ezért az

$$\mathbf{n}_1 - \beta \mathbf{A}^T \mathbf{n}_2 = 0 \quad (9)$$

feltétel lineáris az \mathbf{A} affin transzformáció paramétereinek a függvényében. A 9. egyenlet az x és y koordinátákra kifejtve így néz ki:

$$n_1^x - \beta n_2^x a_1 - \beta n_2^y a_3 = 0, \quad n_1^y - \beta n_2^x a_2 - \beta n_2^y a_4 = 0. \quad (10)$$

Vezessünk be egy J költségfüggvényt, amely az affin paramétereiből következő hibát bünteti. Ehhez a költségfüggvényhez a feltételeket (8. és 10. egyenletek) Lagrange multiplikátor segítségével hozzá lehet adni. Így kapjuk az alábbi módosított költségfüggvényt:

$$J(\mathbf{A}, \lambda_1, \lambda_2) = \frac{1}{2} \sum_{i=1}^4 (a_i - a'_i)^2 + \lambda_1 (n_1^x - \beta n_2^x a_1 - \beta n_2^y a_3) + \lambda_2 (n_1^y - \beta n_2^x a_2 - \beta n_2^y a_4), \quad (11)$$

ahol λ_1 and λ_2 jelöli a két Lagrange multiplikátort. A 8. egyenlet nemnegatív értékeket tartalmaz, ezért az optimális megoldás J parciális deriváljainak segítségével kapható meg:

$$\begin{aligned} \frac{\partial J}{\partial a_1} &= a_1 - a'_1 - \beta n_2^x \lambda_1 = 0, & \frac{\partial J}{\partial a_2} &= a_2 - a'_2 - \beta n_2^x \lambda_2 = 0, \\ \frac{\partial J}{\partial a_3} &= a_3 - a'_3 - \beta n_2^y \lambda_1 = 0, & \frac{\partial J}{\partial a_4} &= a_4 - a'_4 - \beta n_2^y \lambda_2 = 0, \\ \frac{\partial J}{\partial \lambda_1} &= n_1^x - \beta n_2^x a_1 - \beta n_2^y a_3 = 0, & \frac{\partial J}{\partial \lambda_2} &= n_1^y - \beta n_2^x a_2 - \beta n_2^y a_4 = 0. \end{aligned}$$

Ezeket összefoglalhatjuk egy inhomogén, lineáris egyenletrendszerbe, amely $\mathbf{C}\mathbf{x} = \mathbf{b}$ alakban írható fel, ahol $\mathbf{x} = [a_1 \ a_2 \ a_3 \ a_4 \ \lambda_1 \ \lambda_2]^T$ és $\mathbf{b} = [a'_1 \ a'_2 \ a'_3 \ a'_4 \ -n_1^x \ -n_1^y]^T$. A 6×6 -os méretű \mathbf{C} mátrix tartalmazza az együtthatókat:

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & -\beta n_2^x & 0 \\ 0 & 1 & 0 & 0 & 0 & -\beta n_2^x \\ 0 & 0 & 1 & 0 & -\beta n_2^y & 0 \\ 0 & 0 & 0 & 1 & 0 & -\beta n_2^y \\ -\beta n_2^x & 0 & -\beta n_2^y & 0 & 0 & 0 \\ 0 & -\beta n_2^x & 0 & -\beta n_2^y & 0 & 0 \end{bmatrix}.$$

A megoldás triviális: $\mathbf{x} = \mathbf{C}^{-1}\mathbf{b}$. Az 1. algoritmus részletesen leírja a javasolt módszer pszeudo-kódját.

Algorithm 1 EG- L_2 -Optimális Affin Transzformáció

```

1: procedure CORRECTAFFINETRANSFORMATION
2:   Bemenet:
3:    $\mathbf{F}$  – fundamentális mátrix.
4:    $\mathbf{p}_1, \mathbf{p}_2$  – pontmegfeleltetés.
5:    $\mathbf{A}'$  – bemeneti affin transzformáció.
6:   Kimenet:
7:    $\mathbf{A}$  – optimálisan finomított affin transzformáció.
8:   Algoritmus:
9:    $\mathbf{l}_1 := \mathbf{F}^T \mathbf{p}_2; \mathbf{l}_2 := \mathbf{F} \mathbf{p}_1; \mathbf{n}_1 := [l_1^a; l_1^b] / \|[l_1^a; l_1^b]\|_2; \mathbf{n}_2 := [l_2^a; l_2^b] / \|[l_2^a; l_2^b]\|_2;$ 
10:   $s_1 := f_{11}n_1^x + f_{12}n_1^y; s_2 := f_{21}n_1^x + f_{22}n_1^y; s_3 := f_{31}n_1^x + f_{32}n_1^y;$ 
11:   $\beta := (1 / |s_1x_2 + s_2y_2 + s_3|) \sqrt{l_2^a l_2^a + l_2^b l_2^b};$ 
12:   $\mathbf{C} := \text{eye}(6, 6); \mathbf{C}_{55} := 0; \mathbf{C}_{66} := 0;$ 
13:   $\mathbf{C}_{15} := -\beta n_2^x; \mathbf{C}_{26} := -\beta n_2^x; \mathbf{C}_{35} := -\beta n_2^y; \mathbf{C}_{46} := -\beta n_2^y;$ 
14:   $\mathbf{C}_{51} := -\beta n_2^x; \mathbf{C}_{62} := -\beta n_2^x; \mathbf{C}_{53} := -\beta n_2^y; \mathbf{C}_{64} := -\beta n_2^y;$ 
15:   $\mathbf{b} := [a'_1; a'_2; a'_3; a'_4; -n_1^x; -n_1^y];$ 
16:   $\mathbf{x} := \mathbf{C}^{-1}\mathbf{b};$ 
17:   $\mathbf{A} := [x_1, x_2; x_3, x_4];$ 

```

3.. A Legkisebb Négyzetes Módszer Igazolása

Ebben a szakaszban megmutatjuk, hogy a minimalizáláskor alkalmazott (Frobenius) norma geometriai és algebrai jelentéssel egyaránt bír.

A 2×2 -es \mathbf{A} affin mátrixot két pontmegfeleltetés alapján meg lehet határozni, hiszen a transzformációnak négy paramétere van, egy pontmegfeleltetés pedig két koordinátára vonatkozik, azaz két pontmegfeleltetés négy egyenletet produkál. Válasszuk ki az $[1 \ 0]^T$ és $[0 \ 1]^T$ pontokat az első képen. A minimalizálandó

kifejezés az első válaszott pontra így alakul:

$$\begin{aligned} \left\| \mathbf{A} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \mathbf{A}' \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\|_2^2 &= \left\| (\mathbf{A} - \mathbf{A}') \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\|_2^2 = \\ \left\| \begin{bmatrix} a_1 - a'_1 & a_2 - a'_2 \\ a_3 - a'_3 & a_4 - a'_4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\|_2^2 &= \left\| \begin{bmatrix} a_1 - a'_1 \\ a_3 - a'_3 \end{bmatrix} \right\|_2^2 = \\ &= (a_1 - a'_1)^2 + (a_3 - a'_3)^2 = 0. \end{aligned} \quad (12)$$

A második pontra a minimalizálandó kifejezés hasonlóan számítható:

$$\left\| \mathbf{A} \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \mathbf{A}' \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\|_2^2 = \left\| \begin{bmatrix} a_2 - a'_2 \\ a_4 - a'_4 \end{bmatrix} \right\|_2^2 = (a_2 - a'_2)^2 + (a_4 - a'_4)^2 = 0. \quad (13)$$

Jól látszik, hogy a két pontra vett költség éppen az affin transzformáció elemeinek négyzetösszege, azaz maga a teljes költségfüggvény. Azaz kijelenthetjük, hogy az algebrai hiba megegyezik a kiválasztott két pontra vonatkozó geometriai hibával.

4.. Tesztek

A tesztelési leírás első felében megmutatjuk, hogyan állítunk elő tökéletes (ground truth) adatokat, majd részletesen ismertetjük a kapott futtatási eredményeket.

4.1.. Affin Transzformációk Becslése Homográfiából

Ahogy azt Molnár és mtsai. [23] megmutatták, az \mathbf{A} lokális transzformáció a vonatkozó homográfia kétdimenziós irány szerinti parciális deriváltjaiból számítható – ezt az összefüggést alkalmazzuk itt is. Tegyük fel, hogy a \mathbf{H} homográfia adott. Az első és a második képen ekkor két pont ($\mathbf{p}_1 = [x_1 \ y_1 \ 1]^T$ és $\mathbf{p}_2 = [x_2 \ y_2 \ 1]^T$) homogén koordinátás alakja között a homográfia teremt kapcsolatot: $\mathbf{H}\mathbf{p}_1 \sim \mathbf{p}_2$. A 2×2 -es affin mátrix a parciális deriváltak segítségével így írható le:

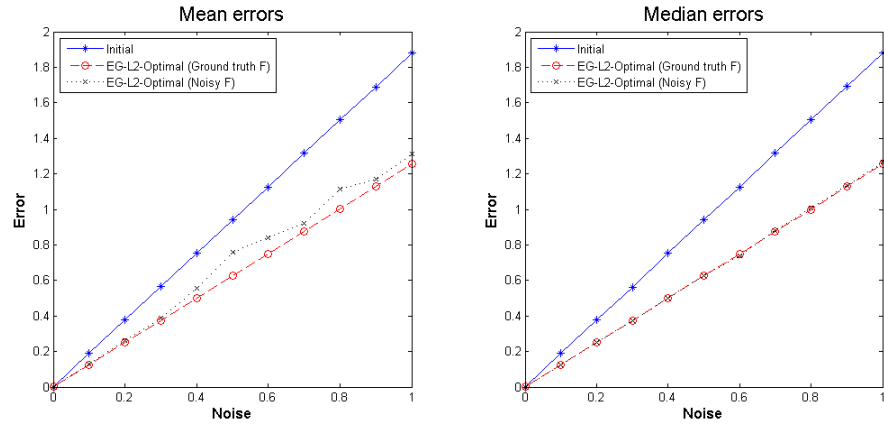
$$a_{1j} = \frac{h_{1j} - h_{3j}x_2}{s} \quad a_{2j} = \frac{h_{2j} - h_{3j}y_2}{s} \quad j \in \{1, 2\}, \quad (14)$$

ahol $s = \mathbf{h}_3^T \mathbf{p}_1$ a projektív mélység, \mathbf{h}_i^T a homográfia i -dik sora, h_{ij} pedig a \mathbf{H} homográfia megfelelő eleme. A levezetés részletesen megtalálható korábbi cikkünkben [8]. Az affin transzformációhoz tartozó eltolás a pontmegfeletetésekből számítható. A tesztek során a ground truth adatokhoz pontmegfeletetéseket, homográfiákat generáltunk, és a fenti módszerrel készítettük el az affin transzformációkat.

4.2.. Szintetikus Tesztek

A szintetikus tesztekhez két perspektív kamerát vettünk, a projekciós mátrixait jelöljük \mathbf{P}_1 -gyel és \mathbf{P}_2 -vel. A kamerák helyét a $Z = 60$ síkon vettük fel

véletlenszerűen. A sík értelemszerűen párhuzamos az XY síkkal. Mindkét kamera az origó felé néz. A kamerák fókusz távolsága 600, a dőléspont helye $[300 \ 300]^T$. A kamerák generálása után 50 pontot véletlenszerűen előállítottunk egy véletlen síkon, mely sík tartalmazza az origót. A pontokat levetítettük, a homográfiákat kiszámítottuk, és az affin transzformációkat a 14. összefüggés segítségével megkaptuk. A kapott adatokhoz véletlen zajt kevertünk. A tesztelést 500-szor megismételtük különböző zajszintekre. A zajosítás során nulla várható értékű, normál-eloszlású zajt adtunk hozzá különböző szórással a pontmegfeleltetésekhez és az affin transzformációk elemeihez.



2. ábra. Az eredeti és az optimális affin transzformációk hibája változó zajszinten. A ground truth transzformációtól vett átlagos L_2 távolságot mutatjuk be a Gauss-zaj σ szórásának (pixelben értendő) függvényében. A zajt mind az affin transzformációkhoz, mind a pontmegfeleltetésekhez hozzáadtuk. Piros egyenes: A ground truth fundamentális mátrixot használtuk a becsléshez. Fekete egyenes: A fundamentális mátrixot a normalizált 8-pontos algoritmussal becsültük a zajos pontmegfeleltetésekből, majd a szimmetrikus epipoláris hibát a Levenberg-Marquardt optimalizációval minimalizáltuk. A medián ábrán a fekete és piros egyenesek egybeesnek.

A 2. ábra mutatja az affin transzformációk becslésére kapott átlagos (mean) és medián (median) hibát. A két görbe mutatja, hogy az EG- L_2 javítással és a javítás nélkül a hiba hogyan alakul a ground truth értékekhez képest. A függőleges tengely mutatja a becsült affin transzformációkra az L_2 -es norma szerinti hibát, a vízszintes tengely a Gauss-eloszlású zaj σ szórását jelöli.

A kék görbe az eredeti affin transzformációk hibáját mutatja. A javított affinitások eredményeit a vörös és a fekete görbe adja. A vörös görbe azt a hibát jelöli, amit úgy kaptunk, hogy a fundamentális mátrix tökéletes volt, azaz a ground truth értékeket alkalmaztuk a kiszámításához. A fekete görbe esetén a fundamentális mátrixot a zajos pontokból becsültük meg. Ez a becslés a nor-

malizált 8-pontos algoritmussal történt, utána Levenberg-Marquardt iterációval a szimmetrikus epipoláris hibát minimalizáltuk. A tesztek alapján egyértelmű, hogy a javított affin transzformációk lényegesen jobbak. A medián- és átlaghibák karakterisztikája között nem látunk lényegi eltérést.

A **futási idő** a javasolt módszerre elhanyagolható, hiszen az algoritmus csak néhány műveletet tartalmaz. A C++ implementáció a másodperc tört része alatt lefut egy 2.3 GHz-es PC-n.

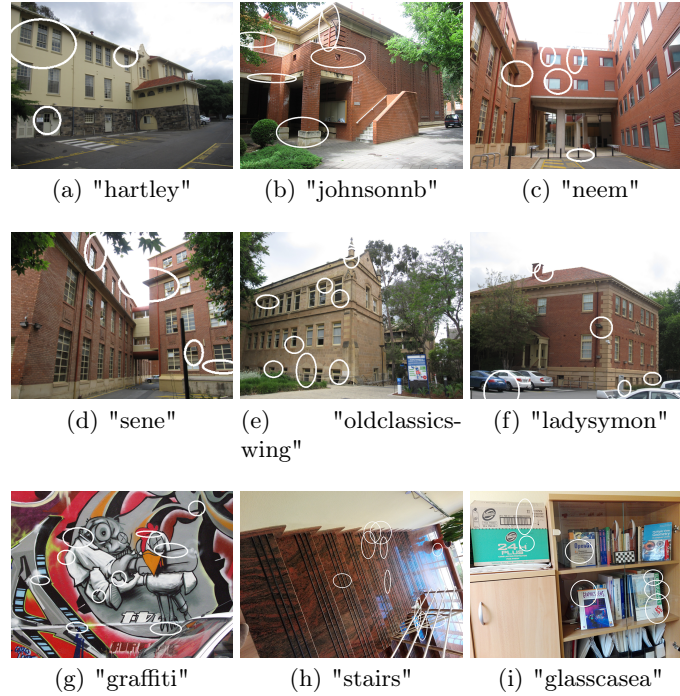
4.3.. Valós Tesztek

Detektor		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	átlag	medián
AAKAZE	Megfigyelt	0.26	0.30	0.17	0.30	0.26	0.18	0.25	0.62	0.38	0.30	0.26
	EG-L_2-Opt	0.21	0.22	0.12	0.19	0.19	0.14	0.16	0.54	0.26	0.23	0.19
ABRISK	Megfigyelt	0.28	0.33	0.27	0.38	0.28	0.30	0.28	1.31	0.31	0.42	0.30
	EG-L_2-Opt	0.21	0.25	0.19	0.24	0.22	0.18	0.18	0.50	0.20	0.24	0.21
AHES-AFF	Megfigyelt	0.19	0.23	0.18	0.20	0.14	0.17	0.21	0.24	0.22	0.20	0.20
	EG-L_2-Opt	0.14	0.17	0.11	0.13	0.09	0.11	0.13	0.14	0.15	0.13	0.13
AORB	Megfigyelt	0.34	0.34	0.15	0.45	0.23	0.24	0.27	-	0.28	0.29	0.28
	EG-L_2-Opt	0.27	0.28	0.10	0.29	0.17	0.18	0.18	-	0.20	0.20	0.19
ASIFT	Megfigyelt	0.27	0.28	0.27	0.26	0.21	0.22	0.27	0.23	0.29	0.26	0.27
	EG-L_2-Opt	0.20	0.21	0.15	0.17	0.14	0.17	0.16	0.17	0.18	0.17	0.17
ASURF	Megfigyelt	0.23	0.27	0.17	0.30	0.22	0.17	0.25	0.26	0.27	0.24	0.25
	EG-L_2-Opt	0.18	0.20	0.11	0.21	0.16	0.12	0.17	0.18	0.19	0.18	0.18
HAR-AFF	Megfigyelt	0.24	0.25	0.15	0.24	0.16	0.27	0.20	0.38	0.28	0.24	0.24
	EG-L_2-Opt	0.18	0.18	0.09	0.19	0.12	0.19	0.13	0.35	0.17	0.16	0.18
HES-AFF	Megfigyelt	0.24	0.22	0.20	0.22	0.13	0.20	0.19	-	0.24	0.21	0.21
	EG-L_2-Opt	0.17	0.16	0.10	0.17	0.09	0.09	0.12	-	0.15	0.13	0.14
MODS	Megfigyelt	0.29	0.40	0.23	0.31	0.26	0.25	0.61	0.24	0.47	0.34	0.29
	EG-L_2-Opt	0.20	0.25	0.13	0.22	0.19	0.17	0.42	0.19	0.32	0.23	0.20
MSER	Megfigyelt	0.42	0.69	0.46	0.34	0.29	0.31	0.42	0.51	0.34	0.42	0.42
	EG-L_2-Opt	0.24	0.32	0.23	0.25	0.20	0.22	0.25	0.31	0.21	0.25	0.24

1. táblázat. Az affin-kovariáns detektorok által "megfigyelt" és az EG- L_2 -Optimális affin transzformációk hibája. A hiba az átlaga a becült és a ground truth mátrixok különbségmátrixainak L_2 normát használva. Tesztpárok: (a) hartley, (b) johnsonnb, (c) neem, (d) sene, (e) oldclassicswing, (f) ladysymon (g) graffiti (h) stairs (i) glasscasea

Az algoritmust az AdelaideRMF adatbázison teszteltük a következő képpárokra: "graffiti", "stairs" és "glasscasea". (A 3. ábra mutatja meg a képeket.) A pontmegfeleltetéseket kézzel végeztük el, a síkok homográfiáit ezekből a pontokból becültük saját, publikus módszerünkkel.³ Az affin transzformációkat a parciális deriváltakból kaptuk meg.

³ Elérhető a <http://web.eee.sztaki.hu/home4/node/56> címen.



3. ábra. A kiválasztott képpárok első képei ellipszisekkel, melyek az affin transzformációt ábrázolják.

Különböző affin-kovariáns detektorokat futattunk a képpárokon. Ezek a detektorok a következők: AAKAZE, ABRISK, AORB, ASIFT, ASURF, AHessian-Affine⁴, MODS⁵, MSER, Harris-Affine és Hessian-Affine⁶

A teszt során az egymásnak megfelelő jellegzetes pontokat detektáltuk[24], és a legközelebbi annotált síkhoz rendeltük. A távolságot a pont és a homográfia között a visszavetítési hiba adja: $(H\mathbf{p}_1 \sim \mathbf{p}_2)$. Ha egy távolság nagyobb, mint 1.0 pixel, akkor kiteszük az adathalmazból. Az affin transzformációt az 1. összefüggés segítségével számítjuk. A fundamentális mátrixot a 8-pontos algoritmus segítségével számítjuk először, majd a Levenberg-Marquard iteráció [25] segítségével finomítjuk.

A kapott hibaértékeket az 1. táblázat tartalmazza. A leírt hiba az L_2 -es norma szerinti hibaértékek átlaga a ground truth értékekhez képest. Minden

⁴ Az ASIFT algoritmus letölthető a <http://www.ipol.im/pub/art/2011/my-asift> címről. Az "A*" módszerek esetén a képszintézissel kapott képeket az AKAZE, BRISK, ORB, SIFT, SURF, Hessian-Affine módszerekkel is házasítottuk.

⁵ A MODS algoritmust a <http://cmp.felk.cvut.cz/wbs> címről töltöttük le.

⁶ Az MSER, Har-Aff és Hes-Aff módszereket a <http://www.robots.ox.ac.uk/~vgg/research/> címről töltöttük le.

egyes oszlop egy tesztpárt jelképez az utolsó kettő oszlopot leszámítva. Ezek a teljes sorozatra az átlag és a medián hibát tartalmazzák. A páros és a páratlan sorok összefüggenek abban az értelemben, hogy a páratlan sorok az eredeti, a párosak a javított értékeket tartalmazzák. A hibaértékek definíciója ugyanaz, mint a szintetikus tesztek esetében.

A detektáló módszerekhez a letöltéshez tartozó eredeti (default) paraméterezést használtuk. Az átlag- és mediánhibák ugyanazt a trendet mutatják: az EG- L_2 -Optimális affin transzformációk kivétel nélkül mindig jobbak az eredeti értékeknél.

	AAKAZE	ABRISK	AHES-AFF	AORB	ASIFT	ASURF	HAR-AFF	HES-AFF	MODS	MSER
Inlierek	239	110	1420	145	2082	837	64	73	941	78
Idő	81.91	81.38	89.30	86.39	81.34	84.00	4.10	3.22	52.92	0.41

2. táblázat. Az átlagos inlier szám a különböző detektorokhoz. Inlier-ek azok a pontok, amelyek egy annotált síkon fekszenek. A feldolgozási idő másodpercben értendő. A teszthez egy Intel Core4Quad 2.33 GHz PC-t használtunk egy maggal, párhuzamosítás nélkül.⁸

Érdekessége a tesztnak, hogy a Hessian-Affine módszer az ASIFT eljárás képszintetizáló módszerével kiegészítve (AHES-AFF-el rövidítettük a táblázatban) adja a legjobb eredményt, és viszonylag sok pontmegfeleltetést is ad.

Ha a kapott pontok száma lehet kevesebb, jó alternatíva képszintézis nélkül a Hessian-Affine módszer, amelyiknek közel azonos a minősége, de jóval kisebb a futtatási igénye.

4.4.. Homográfia és Felület Normális Becslésének Javítása

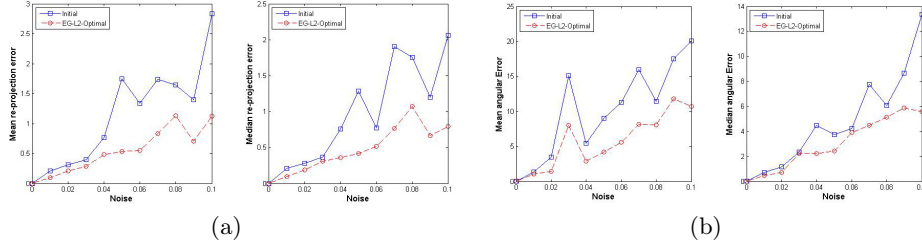
Ez a szakasz megmutatja, hogy az EG- L_2 -Optimális affin transzformációk segítségével olyan módszerek minősége is javítható, melynek a bemenete maga az affin transzformáció. Két konkrét algoritmust vizsgálunk meg: homográfia-becslést és felületi normális becslést.

A homográfia-becslésre a már korábban bemutatott (4.2.. fejezet) szintetikus tesztrendszer alkalmaztuk: véletlen síkon tíz pontot mintavételeztünk, majd ezeket levetítettük a kamerára, a kapott pontokat és az affin transzformációkat zajosítottuk. Koeser [3] módszerével becsültük a homográfiaát. 500-szor ismételtük a tesztet minden egyes zajszintre. A 4(a). ábra mutatja a kapott homográfiaakra vett visszavetítési hibákat, amelyek a pontokra számított visszavetítési hiba átlaga és mediánja.

A felületi normálisokat korábbi módszerünk [9] segítségével számítottuk. Ebben a cikkünkben megmutattuk, hogy kölcsönösen egyértelmű megfeleltetés létezik az affin transzformáció és a felületi normális között. A vizsgálati eredményekhez a cikkben közölt tesztlő környezetet alkalmaztuk, a módszerek közül pedig a leggyorsabb FNE (Fast Normal Estimator) algoritmust választottuk.⁹

⁹ Letölthető a <http://web.eee.sztaki.hu/home4/node/53> címről.

A 4(b). ábrán látható eredmények meggyőzhetik az Olvasót is, hogy a javított affin transzformációk a felületi normálisok becslését is javítják.



4. ábra. (a) A homográfia-becslés[3] átlagos (mean) és medián (median) hibája. A hibát a visszavetítésből (reprojection error) számoljuk. (b) A felületi normális szöghibájának átlagos és medián értékei. A grafikonokon a kezdeti és az EG-L2 optimális értékek szerepelnek. A vízszintes tengely a zéró várható értékű zaj szórásának felel meg.

5. Összefoglalás

Ebben a cikkben megmutattuk, hogy a detektáló algoritmusok segítségével (sztereo) képen meghatározott lokális affin transzformációkat hogyan lehet javítani. A javasolt algoritmus a definiált hibafüggvény minimalizálására legkisebb négyzetes értelemben optimális megoldást ad. A szükséges számítási idő korszerű processzorokon futtatva elhanyagolhatónak tekinthető.

Azt is megmutattuk, hogy a bevezetett költségfüggvény geometriai jelentéssel is bír. A bevezetett EG- L_2 -Optimális eljárást szintetikus és valós adatokon egyaránt megvizsgáltuk. Kivétel nélkül az összes teszten sikerült a minőséget javítani. Az affin detektorok hibáját körülbelül az eredeti érték 65%-ra csökkenteni az eljárás. Azt is megmutattuk, hogy a javított affin transzformációk a homográfia becslését és a felületi normálisokat javítja.

Érdekes mellékszál a kutatásnak, hogy a detektorok közül a Hessian-Affine érte el a legjobb eredményt, főleg akkor, ha a képeket az ASIFT módszerből kiemelt képtranszformációval torzítottuk.

A módszer forráskódja elérhető a <http://web.eee.sztaki.hu/home4/node/56> címen.

Támogatók.

A munkát az Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal VKSZ 14-1-2015-0072. pályázata támogatta. J. Matas was supported by the GACR P103/12/G084 grant.

Hivatkozások

1. J. Matas, S. Obdržálek, and O. Chum, „Local affine frames for wide-baseline stereo,” in *ICPR, Quebec, Canada, August 11-15, 2002.*, 2002, pp. 363–366.
2. K. Köser and R. Koch, „Differential spatial resection - pose estimation using a single local image feature,” in *ECCV*, 2008, pp. 312–325.
3. K. Köser, *Geometric Estimation with Local Affine Frames and Free-form Surfaces*. Shaker, 2009.
4. J. Bentolila and J. M. Francos, „Conic epipolar constraints from affine correspondences,” *Computer Vision and Image Understanding*, vol. 122, pp. 105–114, 2014.
5. Y. Furukawa and J. Ponce, „Accurate, dense, and robust multi-view stereopsis,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362–1376, 2010.
6. 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.
7. C. Raposo and J. P. Barreto, „Theory and practice of structure-from-motion using affine correspondences,” 2016.
8. D. Barath and L. Hajder, „Novel ways to estimate homography from local affine transformations,” in *11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*, 2016, pp. 432–443.
9. 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*. Springer International Publishing, 2016, pp. 316–337.
10. R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
11. J. Matas, O. Chum, M. Urban, and T. Pajdla, „Robust wide baseline stereo from maximally stable extremal regions,” in *Proc. BMVC*, 2002, pp. 36.1–36.10.
12. K. Mikolajczyk and C. Schmid, „An affine invariant interest point detector,” in *ECCV*. Springer, 2002, pp. 128–142.
13. —, „Scale & affine invariant interest point detectors,” *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
14. A. Baumberg, „Reliable feature matching across widely separated views,” in *Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2000, pp. 774–781.
15. J.-M. Morel and G. Yu, „ASIFT: A new framework for fully affine invariant image comparison,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.
16. H. Bay, T. Tuytelaars, and L. Van Gool, „Surf: Speeded up robust features,” in *European conference on computer vision*. Springer, 2006, pp. 404–417.
17. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, „Orb: An efficient alternative to sift or surf,” in *2011 International conference on computer vision*. IEEE, 2011, pp. 2564–2571.
18. S. Leutenegger, M. Chli, and R. Y. Siegwart, „Brisk: Binary robust invariant scalable keypoints,” in *2011 International conference on computer vision*. IEEE, 2011, pp. 2548–2555.
19. D. Mishkin, J. Matas, and M. Perdoch, „MODS: Fast and robust method for two-view matching,” *Computer Vision and Image Understanding*, vol. 141, pp. 81–93, 2015.
20. K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, „A comparison of affine region detectors,” *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

21. R. I. Hartley and P. Sturm, „Triangulation,” *Computer Vision and Image Understanding: CVIU*, vol. 68, no. 2, pp. 146–157, 1997.
22. K. Turkowski, „Transformations of surface normal vectors,” in *Tech. Rep. 22, Apple Computer*, 1990.
23. J. Molnár and D. Chetverikov, „Quadratic transformation for planar mapping of implicit surfaces,” *Journal of Mathematical Imaging and Vision*, vol. 48, pp. 176–184, 2014.
24. D. G. Lowe, „Object recognition from local scale-invariant features,” in *Proceedings of the International Conference on Computer Vision*, ser. ICCV, 1999, pp. 1150–1157.
25. J. Moré, „The levenberg-marquardt algorithm: implementation and theory,” in *Numerical analysis*. Springer.