



Viziune Artificială

TEZĂ DE DOCTORAT

- REZUMAT -

Metode Liniare de Estimare a Odometriei Vizuale

Student-doctorand:
Cătălin Cosmin Golban

Conducător științific:
Prof. Dr. Ing. Sergiu Nedevschi

Comisia de evaluare a tezei de doctorat:

Președinte: Prof. Dr. Ing. **Rodica Potolea** - Universitatea Tehnică din Cluj-Napoca

Conducător științific: Prof. Dr. Ing. **Sergiu Nedevschi** - Universitatea Tehnică din Cluj-Napoca

Referenți:

- Prof. Dr. Ing. **Vladimir-Ioan Crețu** - Universitatea Politehnică Timișoara
- Prof. Dr. Ing. **Vasile Ion Manta** - Universitatea Tehnică "Gheorghe Asachi" Iași
- Prof. Dr. Ing. **Florin Oniga** - Universitatea Tehnică din Cluj-Napoca

- Cluj-Napoca -
2022

Cuprins

1. Introducere și context	3
2. Stadiul actual al cunoașterii	5
3. Contribuții personale	5
3.1. Modele liniare de estimare a odometriei vizuale - contribuții teoretice	5
3.2. Metode liniare de estimare a odometriei vizuale - rezultate experimentale	10
3.3. Senzori LiDAR pentru estimarea odometriei	14
3.3.1. Înlocuirea informației stereo de adâncime cu LiDAR	14
3.3.2. Generalizarea modelului matematic de odometrie stereo pentru LiDAR	15
3.3.3. Determinarea distanțelor trăsăturilor video pe bază de LiDAR	15
3.3.4. Rezultate experimentale și concluzii	16
3.4. Folosirea de senzori LiDAR pentru corecția odometriei vizuale	17
3.4.1. Prezentare generală a sistemului	17
3.4.2. Modelul matematic	19
3.4.3. Rezultate și concluzii	20
4. Concluzii	21
4.1. Concluzii generale	21
4.2. Originalitatea și contribuțiile inovative ale tezei	22
4.3. Posibile dezvoltări ulterioare	22
Referințe	23

1. Introducere și context

Disponibilitatea camerelor digitale este în continuă creștere, acestea găsimu-se în numeroase dispozitive și accesorii care în general cresc calitatea vieții și oferă posibilități nebănuite în urmă cu două decenii. Camera este transformată într-un dispozitiv de măsurare datorită capabilităților de reconstrucție 3D care se pot obține folosind procesarea datelor de la camerele video.

Odometria vizuală este capabilitatea unui sistem pe baza unei camere sau a unui sistem de camere video de a estima mișcarea camerei în timp, în spațiul tridimensional în care aceasta se află, pe baza imaginilor achiziționate succesiv la momente diferite de timp. Termenul de odometrie vizuală a fost pentru prima dată introdus în [18] și terminologia a devenit general acceptată în comunitățile de cercetare.

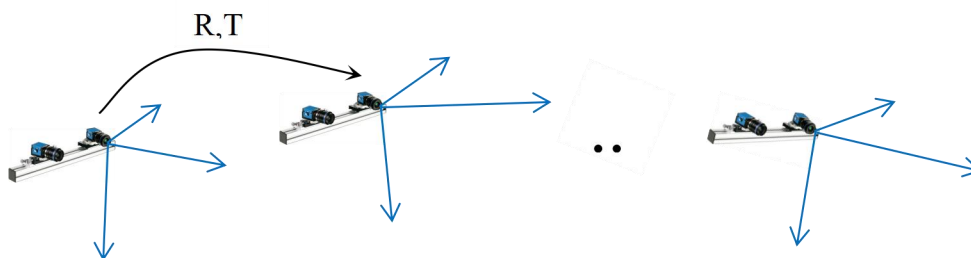


Figura 1 Problema odometriei vizuale

Din punct de vedere formal, odometria vizuală poate fi înțeleasă ca și o transformare între sisteme de coordonate. Imaginea achiziționată de o camera, la un moment dat de timp, este proiecția spațiului tridimensional în planul imagine la acel moment de timp. Acest proces de formare a imaginii poate fi descris ca și o transformare de proiecție dintr-un sistem de coordonate tridimensional atașat camerei într-un sistem de coordonate bidimensional în planul imagine (Figura 2). Mișcarea camerei într-o poziție nouă implică mișcarea sistemului de coordonate tridimensional al camerei într-o poziție nouă, și proiecția elementelor tridimensionale din acest nou sistem de coordonate în planul imagine. Această mișcare poate fi modelată printr-un vector de translație și printr-o matrice de rotație, având în total șase grade de libertate.

Odometria vizuală este estimarea transformării (rotație și translație) dintre sistemele de coordonate tridimensional atașate camerei în momente diferite de timp, pe baza imaginilor preluate de către cameră (stereo) în aceste momente diferite de timp (Figura 1). Se estimează practic mișcarea camerei între momente diferite de timp, pe baza variației în aparența proiecției scenei, în imaginile produse de această cameră la momente diferite de timp. Aceste mișcări între imagini succesive preluate de o cameră video, aflate în mișcare pot fi desigur concatenate, construind o traiectorie a mișcării camerei în timp. În cazul camerelor monoculare, rotația camerei precum și două din cele trei grade de libertate ale translației pot fi estimate. Practic, putem estima direcția vectorului de translație, dar nu și magnitudinea acestuia.

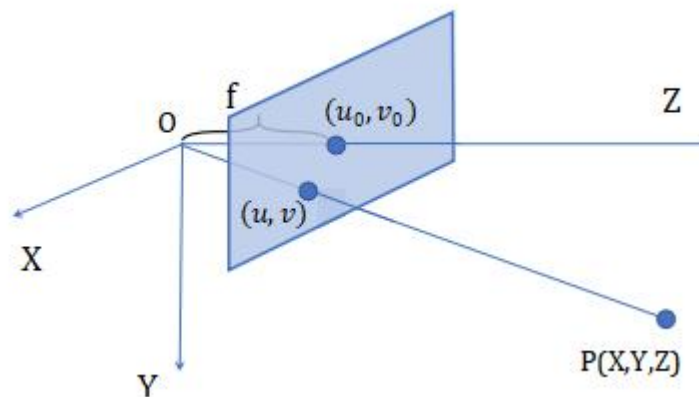


Figura 2 Modelul de proiecție perspectivă. Un punct P de coordonate (X, Y, Z) se proiectează în punctul (u, v). Intersecția axei optice cu planul imagine este punctul principal (u_0, v_0) iar f este distanța focală a camerei

În cazul sistemelor de camere stereo putem estima și magnitudinea translației, deci se pot estima toate cele șase grade de libertate ale mișcării, trei pentru rotație, respectiv trei pentru translație. Geometria unui sistem stereo este ilustrată în Figura 3. Principalele metode explorate în cadrul tezei vor fi metode stereo, dar sunt punctate și considerente teoretice, care ilustrează diferențele dintre metodele monoculare și cele stereo, precum și metode de a combina aceste metode pentru a crește precizia estimării.

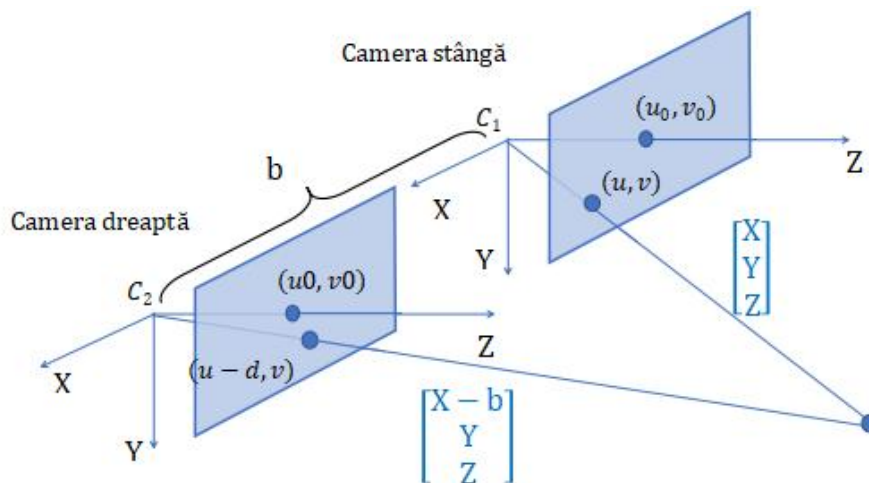


Figura 3 Geometria unei camere stereo. Corespondențele proiecției (u, v) a punctului de coordonate (X, Y, Z) în imaginea stângă se va situa pe aceeași linie în imaginea dreaptă la un deplasament d spre stânga numit disparitate ($u - d, v$)

Odometria vizuală facilitează algoritmi și metode de viziune artificială precum:

- ✓ Asocierea și integrarea în timp a informației extrase din imagini consecutive
- ✓ Compensarea mișcării camerei pentru aplicații de urmărire a obiectelor cu o camera dinamică
- ✓ Reconstrucția structurilor tridimensionale din scenă pe baza de poziții succesive ale aceleiași camere
- ✓ Determinarea traiectoriei unei camere aflate în mișcare

- ✓ Localizarea și construcția de hărți pe bază de sisteme video folosite sub formă de dispozitive de detecție și de măsurare

Odometria vizuale își găsește aplicabilitate în diverse domenii: roboți care se ghidează autonom în medii off-road [19] [20] [21] sau în medii în care nu avem metode de localizare precisă precum sisteme GPS [22], pentru inspecția halelor industriale [34], realitate virtuală [33], mașini autonome [35], obiecte zburătoare la altitudini mari [36], drone [37], telefoane inteligente [33], accesorii pentru realitate argumentată și pentru ajutorul persoanelor cu dizabilități sau a soldaților aflați în misiuni de training sau de salvare [23], industria marina si sub-marina [28], trenuri inteligente [29] [30], roboți agricoli [31] etc. Odometria vizuală și-a găsit aplicabilitatea și în domeniul medical. Metodele de localizare și cartografiere simultană sau odometrie vizuală au fost aplicate cu succes pe fluxul de imagini primit de la un endoscop echipat cu o cameră mono [25] [27] sau stereo [24] [26].

2. Stadiul actual al cunoașterii

Lucrarea este influențată de numeroase publicații în domeniu. Am grupat capitolul de analiză al stadiului actual al cunoașterii în trei mari sub-capitole. Un prim sub-capitol este dedicat analizei evoluției metodelor de odometrie vizuală de la bazele teoretice [1][2][3][4] ale elementelor de geometrie proiectivă până la primele sisteme în timp real [18], incluzând și referințe înspre primele metode de odometrie vizuală pe care le-am publicat [8][9].

Al doilea sub-capitol trece în revistă principalele metode de robustificare a sistemelor de odometrie vizuală pe bază de trăsături.

Al treilea sub-capitol prezintă un set de noi tendințe de cercetare în domeniul odometriei vizuale: sisteme multi-cameră unde nu există neapărat o fixare rigidă a camerelor, metode pe bază de LiDAR, si folosirea rețelelor neuronale convoluționale pentru estimarea odometriei vizuale.

3. Contribuții personale

3.1. Modele liniare de estimare a odometriei vizuale - contribuții teoretice

Algoritmii evaluați vor avea ca și intrare două perechi de imagini achiziționate cu o cameră stereo în două momente diferite de timp t_1 și t_2 , $t_1 < t_2$. Vom folosi denumirile de imagine anterioară stângă și imagine anterioară dreaptă pentru cele două imagini achiziționate la momentul t_1 și denumirile de imagine curentă stângă respectiv imagine curentă dreaptă pentru imaginile achiziționate la momentul t_2 . Modelul matematic se bazează pe un set de trăsături corespondente între cele patru imagini: imaginea anterioară stângă, imaginea anterioară dreaptă, imaginea curentă stângă și imaginea curentă dreaptă. Aceste

trăsături reprezintă proiecțiile perspectivă ale aceleiași primitive tridimensionale (puncte 3D de exemplu) în cele patru imagini.

Vom folosi în continuare următoarele notații (Figura 4) pentru coordonatele trăsăturii numărul i :

$$q_{L1}^i = \begin{bmatrix} u_{L1}^i \\ v_{L1}^i \\ 1 \end{bmatrix} \text{ și } q_{R1}^i = \begin{bmatrix} u_{R1}^i \\ v_{R1}^i \\ 1 \end{bmatrix} \text{ coordonate pixeli în imaginea anterioară stângă respectiv dreaptă}$$

$$q_{L2}^i = \begin{bmatrix} u_{L2}^i \\ v_{L2}^i \\ 1 \end{bmatrix} \text{ și } q_{R2}^i = \begin{bmatrix} u_{R2}^i \\ v_{R2}^i \\ 1 \end{bmatrix} \text{ coordonatele pixeli în imaginea curentă stângă respectiv dreaptă}$$

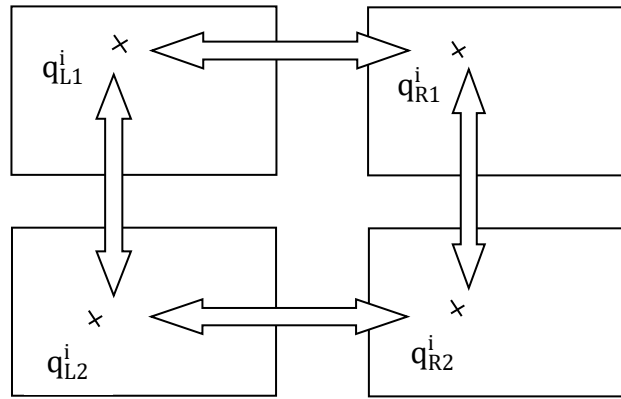


Figura 4 Corespondențe de trăsături între imagini succesive

Prin înmulțirea acestora cu inversa matricii parametrilor intrinseci obținem coordonatele normalizate ale trăsăturilor:

$$p_{L1}^i = K^{-1} \cdot q_{L1}^i = \begin{bmatrix} x_{L1}^i \\ y_{L1}^i \\ 1 \end{bmatrix} \quad p_{R1}^i = K^{-1} \cdot q_{R1}^i = \begin{bmatrix} x_{R1}^i \\ y_{R1}^i \\ 1 \end{bmatrix}$$

$$p_{L2}^i = K^{-1} \cdot q_{L2}^i = \begin{bmatrix} x_{L2}^i \\ y_{L2}^i \\ 1 \end{bmatrix} \quad p_{R2}^i = K^{-1} \cdot q_{R2}^i = \begin{bmatrix} x_{R2}^i \\ y_{R2}^i \\ 1 \end{bmatrix}$$

În general matricea de rotație R se exprimă folosind funcții trigonometrice și cele trei unghiuri Euler: roll, pitch, yaw. Dacă rotațiile care apar sunt mici atunci matricea de rotație se poate aproxima astfel:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} 1 & -\gamma & -\beta \\ \gamma & 1 & -\alpha \\ \beta & \alpha & 1 \end{bmatrix} \quad (1)$$

Figura 5 arată norma Forbenius a diferenței dintre matricea de rotație și aproximarea acesteia din ecuația (1). Unghiurile pitch (α) și yaw (β) sunt variate în intervalul $[-3^\circ, 3^\circ]$ iar unghiul roll (γ) este păstrat constant la 0° respectiv 3° . Observăm diferențe de ordinul 10^{-4} (pentru $\gamma = 0^\circ$) respectiv 10^{-3} (pentru $\gamma = 3^\circ$), și de asemenea faptul că norma Forbenius este mare atunci când două sau mai multe dintre cele trei unghiuri au valori mari simultan, lucru rar întâlnit în practică.

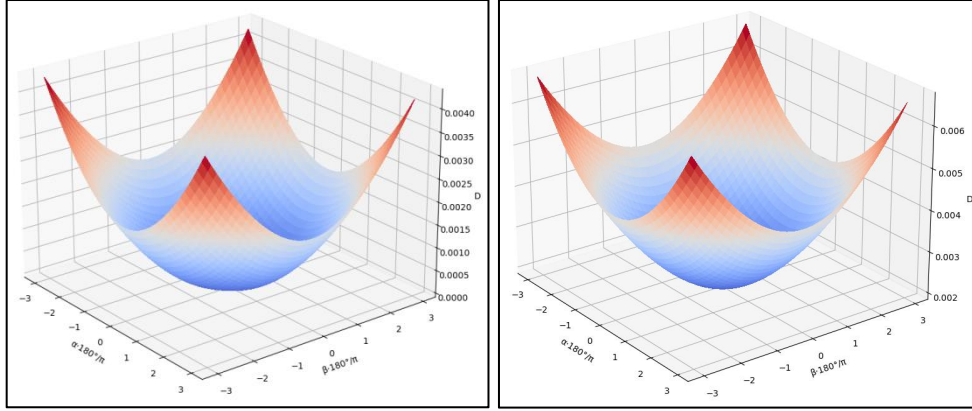


Figura 5 Norma Forbenius a diferenței dintre matricea de rotație și aproximarea acesteia. Parametrii α și β sunt variați în intervalul $[-3^\circ, 3^\circ]$ iar γ este păstrat constant la 0° (stânga) respectiv 3° (dreapta)

Matricea de translație o notăm cu: $T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$. Notăm cu $\delta_1^i = x_{L1}^i - x_{R1}^i$, $\delta_2^i = x_{L2}^i - x_{R2}^i$, $d_1^i =$

$u_{L1}^i - u_{R1}^i$, $d_2^i = u_{L2}^i - u_{R2}^i$ disparitățile în coordonate normalizate și respectiv în pixeli. De asemenea presupunem ca pixelii sunt pătratici: $f_x = f_y$, presupunere validă în cardul sistemului stereo în care se va face implementarea practică a acestui model teoretic și în general valabilă după etapa de eliminare a distorsiunilor și de rectificare a imaginilor inițial obținute de la camere. Folosind reprezentarea aproximativă a matricii de rotație obținem un set de ecuații liniare care vor fi folosite pentru estimarea odometriei vizuale (2), și un set de ecuații pentru predicția mișcării în panul imagine (3) pe baza parametrilor de mișcare [6].

$$\begin{cases} x_{L2}^i(t_x, t_y, t_z) = f \cdot \frac{x_{L1}^i - \gamma y_{L1}^i - \beta f + \frac{t_x \delta_1^i}{b}}{\beta x_{L1}^i + \alpha y_{L1}^i + f + \frac{t_z \delta_1^i}{b}} \\ y_{L2}^i(\alpha, \beta, \gamma, t_x, t_y, t_z) = f \cdot \frac{\gamma x_{L1}^i + y_{L1}^i - \alpha f + \frac{t_y \delta_1^i}{b}}{\beta x_{L1}^i + \alpha y_{L1}^i + f + \frac{t_z \delta_1^i}{b}} \\ \delta_2^i(\alpha, \beta, \gamma, t_x, t_y, t_z) = \frac{f \delta_1^i}{\beta x_{L1}^i + \alpha y_{L1}^i + f + \frac{t_z \delta_1^i}{b}} \end{cases} \quad (2)$$

$$\begin{cases} \alpha \cdot y_{L1}^i \cdot x_{L2}^i + \beta \cdot (x_{L1}^i \cdot x_{L2}^i + f^2) + \gamma \cdot f \cdot y_{L1}^i + t_x \cdot \left(\frac{-f \delta_1^i}{b} \right) + t_z \cdot \frac{x_{L2}^i \delta_1^i}{b} = \\ \quad = f \cdot (x_{L1}^i - x_{L2}^i) \\ \alpha \cdot (y_{L1}^i \cdot y_{L2}^i + f^2) + \beta \cdot x_{L1}^i \cdot y_{L2}^i + \gamma \cdot (-f \cdot x_{L1}^i) + t_y \cdot \left(\frac{-f \delta_1^i}{b} \right) + t_z \cdot \frac{y_{L2}^i \delta_1^i}{b} = \\ \quad = f \cdot (y_{L1}^i - y_{L2}^i) \\ \alpha \cdot y_{L1}^i \cdot \delta_2^i + \beta \cdot x_{L1}^i \cdot \delta_2^i + t_z \cdot \frac{\delta_2^i \delta_1^i}{b} = f \cdot (\delta_1^i - \delta_2^i) \end{cases} \quad (3)$$

Așa cum am propus inițial în [12], pornind de la ecuațiile (2), (3) și considerând doar acele puncte care sunt îndepărtate (i.e $\delta_1^i \rightarrow 0$ și $\delta_2^i \rightarrow 0$) obținem:

$$x_{L2}^i(\alpha, \beta, \gamma) = f \cdot \frac{x_{L1}^i - \gamma y_{L1}^i - \beta f}{\beta x_{L1}^i + \alpha y_{L1}^i + f} \quad (4)$$

$$y_{L2}^i(\alpha, \beta, \gamma) = f \cdot \frac{\gamma \cdot x_{L1}^i + y_{L1}^i - \alpha \cdot f}{\beta \cdot x_{L1}^i + \alpha \cdot y_{L1}^i + f} \quad (5)$$

Sistemul de ecuații care derivă din aceste relații este:

$$\begin{cases} \alpha \cdot y_{L1}^i \cdot x_{L2}^i + \beta \cdot (x_{L1}^i \cdot x_{L2}^i + f^2) + \gamma \cdot f \cdot y_{L1}^i = f \cdot (x_{L1}^i - x_{L2}^i) \\ \alpha \cdot (y_{L1}^i \cdot y_{L2}^i + f^2) + \beta \cdot x_{L1}^i \cdot y_{L2}^i + \gamma \cdot (-f \cdot x_{L1}^i) = f \cdot (y_{L1}^i - y_{L2}^i) \end{cases} \quad (6)$$

Rezolvând acest sistem de ecuații folosind doar puncte îndepărtate obținem rotația dintre două imagini succesive. De remarcat ca indicele 'L' este oarecum redundant în această notație deoarece rotația se determină strict pe baza coordonatelor trăsăturilor din două imagini succesive, nemaifiind nevoie de imaginea dreaptă.

În cazul aplicațiilor automotivă când sistemul de senzori video pe baza căruia este estimată odometria din trăsături video este poziționat pe o mașină și prin urmare aproximarea din (1) este validă și odometria poate fi estimată eficient folosind ecuațiile (2) și (3). Se pune întrebarea în ce măsură putem să reutilizăm modelul matematic în cazul în care avem mișcări de o amplitudine mult mai mare între imagini consecutive (de exemplu în situația în care senzorii video sunt poziționați pe roboți zburători, avioane, drone, sau în cazul în care avem platforme mobile care se deplasează pe teren accidentat). În aceste situații propunerea noastră din [16] este să efectuăm o estimare inițială a rotației. Utilizăm apoi această rotație pentru a transforma punctul $(x_{L1}^i, y_{L1}^i, \delta_{L1}^i)$ într-un sistem de coordonate intermediar apoi folosim (3) pentru a determina transformata dintre sistemul de coordonate intermediar și poziția finală a camerei.

$$x_{L1}^i(R) = f \cdot \frac{r_{11} \cdot x_{L1}^i + r_{12} \cdot y_{L1}^i + r_{13} \cdot f}{r_{31} \cdot x_{L1}^i + r_{32} \cdot y_{L1}^i + r_{33} \cdot f} \quad (7)$$

$$y_{L1}^i(R) = f \cdot \frac{r_{21} \cdot x_{L1}^i + r_{22} \cdot y_{L1}^i + r_{23} \cdot f}{r_{31} \cdot x_{L1}^i + r_{32} \cdot y_{L1}^i + r_{33} \cdot f} \quad (8)$$

$$\delta_1^i(R) = \frac{f \delta_{L1}^i}{r_{31} \cdot x_{L1}^i + r_{32} \cdot y_{L1}^i + r_{33} \cdot f} \quad (9)$$

Valorile rotite $(x_{L1}^i, y_{L1}^i, \delta_1^i)$, sunt folosite în continuare pentru re-estimarea odometriei vizuale.

Pasul 1. Detectăm set de trăsături corespondente între imagini succesive $(x_{L1}^i, y_{L1}^i, \delta_1^i) \leftrightarrow (x_{L2}^i, y_{L2}^i, \delta_2^i)$, $i = \overline{1, N}$

Pasul 2. Estimăm matricea de rotație R având unghiurile (α, β, γ) și opțional translația $T = [t_x \ t_y \ t_z]$ folosind (3) și (2) sau (6) și (4)(5) într-un framework RANSAC

Pasul 3. Opțional re-estimarea corespondențelor de trăsături și rafinarea selecției acestora

Pasul 4.

Pasul 4.1. Transformăm $(x_{L1}^i, y_{L1}^i, \delta_1^i)$ în $(x_{L1}^i, y_{L1}^i, \delta_1^i)$ folosind ecuațiile (7), (8) și (9)

Pasul 4.2. Estimăm rotația parțial R' având unghiurile $(\alpha', \beta', \gamma')$ și translația $T = [t_x \ t_y \ t_z]$ folosind din nou (3) și (2) într-un framework RANSAC, dar pe baza valorilor $(x_{L1}^i, y_{L1}^i, \delta_1^i)$ și $(x_{L2}^i, y_{L2}^i, \delta_2^i)$

Pasul 4.3. Compunem rotația totală astfel: $R = R' \cdot R$

Pentru evaluare am generat un set de 400 de trăsături (u_1^i, v_1^i) distribuite aleatoriu în spațiul unei imagini având rezoluția de 1024 x 768 pixeli. Punctul principal (u_0, v_0) este setat

în centrul imaginii iar distanța focală a acestora este $f_x = f_y = 1000$ px. Pentru fiecare punct (u_1^i, v_1^i) generăm aleatoriu o distanță Z_1^i pe baza căreia calculăm valoarea corespunzătoare a disparității d_1^i . Din (u_1^i, v_1^i, d_1^i) putem reconstrui P_1^i , poziția 3D a punctului. Generăm apoi o valoare de referință pentru matricea de rotație și pentru vectorul de translație și le folosim pentru a determina poziția punctului 3D în sistemul de coordonate al celei de-a doua camere:

$$P_2^i = R_{\text{ref}} \cdot P_1^i + T_{\text{ref}}$$

Proiectând punctul P_2^i în planul celei de-a doua camere, obținem poziția (u_2^i, v_2^i) și disparitatea d_2^i corespunzătoare acestora. Obținem astfel un set de trăsături corespondente $(u_1^i, v_1^i, d_1^i) \leftrightarrow (u_2^i, v_2^i, d_2^i)$, generate sintetic pe baza unei mișcări de referință $R_{\text{ref}}, T_{\text{ref}}$.

Vom genera în continuare mai multe seturi de astfel de puncte pentru a evalua beneficiul reestimării rotației așa cum am propus în pasul 4 al algoritmul de mai sus.

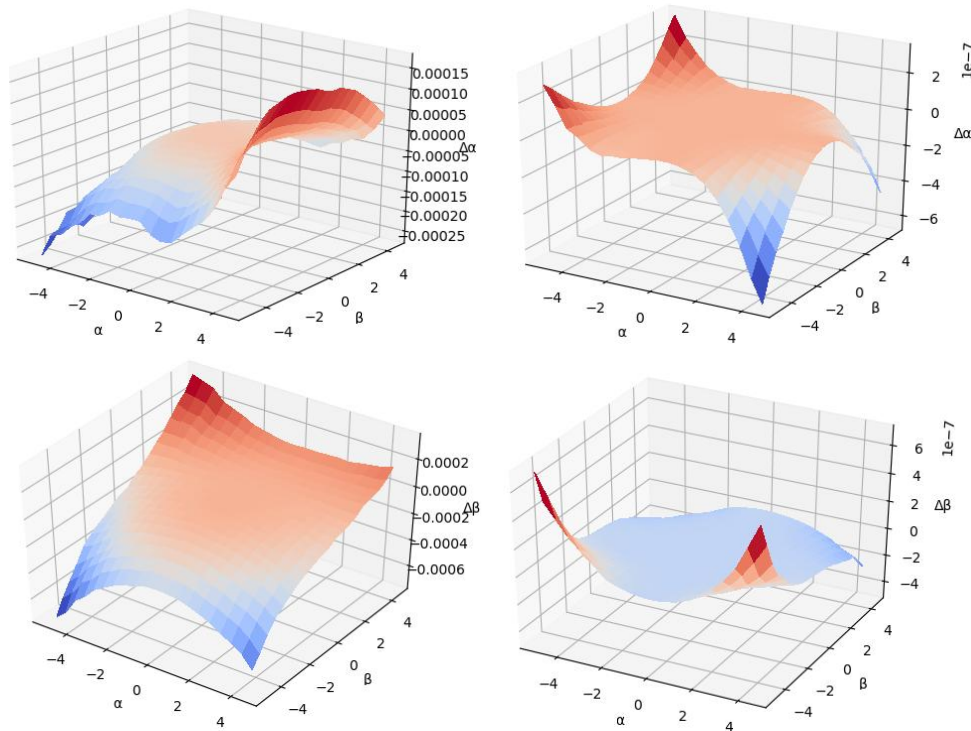


Figura 6 Erorile de estimare ale unghiurilor pitch (α , rândul de sus) și yaw (β , rândul de jos). Coloana stângă prezintă rezultatele după o singură estimare pe baza metodei liniare (pasul 2 al algoritmului). Coloana dreaptă prezintă rezultatele după reestimarea efectuată la pasul 4 al algoritmului

Figura 6 prezintă erorile de estimare pentru o configurație în care T_{ref} este fixat la o mișcare de translație în direcția axei Z, unghiul γ este de asemenea fixat la 0° iar valorile lui α, β sunt variate în intervalul $[-5^\circ, 5^\circ]$. Pentru fiecare combinație de α și β , folosind valorile fixe pentru γ și T_{ref} generăm un set de corespondențe $(u_1^i, v_1^i, d_1^i) \leftrightarrow (u_2^i, v_2^i, d_2^i)$. Pe baza corespondențelor $(u_1^i, v_1^i, d_1^i) \leftrightarrow (u_2^i, v_2^i, d_2^i)$ folosim metodele liniare de estimare a mișcării și comparăm această mișcare cu mișcarea de referință, rezultând erorile de estimare prezentate în Figura 6. Observăm un ordin de magnitudine de 10^{-3} pentru o singură estimare, respectiv un ordin de magnitudine de 10^{-7} pentru o reiterare a estimării, maximele fiind atunci când ambele valori α, β se apropie de extremele intervalului $[-5^\circ, 5^\circ]$.

Experimentele prezentate în Figura 6 se desfășoară în condiții ideale. Pentru a confirma beneficiul adus de reestimarea liniară a odometriei atunci când avem unghiuri mari de mișcare, am repetat experimentul cu adăugare de zgomot Gaussian cu o deviație standard de 0.5px punctelor (u_1^i, v_1^i, d_1^i) și (u_2^i, v_2^i, d_2^i) pe baza cărora efectuăm estimarea. Ordinul de magnitudine al erorilor relativ la valorile de referință $R_{\text{ref}}, T_{\text{ref}}$ sunt prezentate în Tabelul 1.

Observând ultimele patru coloane ale Tabelului 1, remarcăm că erorile de rotație sunt de ordinul 10^{-4} radiani, iar cele de translație de ordinul 10^{-2} metri chiar și la mișcări de $\pm 10^\circ$ între imagini consecutive. Concluzionăm astfel că metodele liniare de odometrie vizuală sunt generalizabile pentru orice tip de mișcare efectuată de o cameră stereo.

Cu zgomote adăugate la trăsăturile generate sintetic	Metoda liniară, o singură iterație				Metoda liniară, o iterație urmată de reestimare			
	$\Delta\alpha$ (rad)	$\Delta\beta$ (rad)	$\Delta\gamma$ (rad)	ΔT (m)	$\Delta\alpha$ (rad)	$\Delta\beta$ (rad)	$\Delta\gamma$ (rad)	ΔT (m)
α, β, γ $\in [-3^\circ, 3^\circ]$	1e-4	1e-4	1e-3	1e-2	1e-4	1e-4	1e-4	1e-3
α, β, γ $\in [-5^\circ, 5^\circ]$	1e-3	1e-3	1e-2	1e-1	1e-4	1e-4	1e-4	1e-2
$\alpha, \beta, \gamma \in$ $[-10^\circ, 10^\circ]$	1e-2	1e-2	1e-2	1e-0	1e-4	1e-4	1e-4	1e-2

Tabel 1 Evaluarea ordinului de magnitudine al erorilor pentru o translație longitudinală $T_{ref} = [0 \ 0 \ -1]$ și pentru unghiurile de rotație variind în intervalul indicat pe fiecare rând al tabelului, și adăugând zgomot Gaussian trăsăturilor generate sintetic

Rezolvarea prin metoda celor mai mici pătrate (least squares) a sistemului supradeterminat de ecuații de mai sus este echivalent cu determinarea minimului următoarei funcții:

$$\begin{aligned}
 & F(x_{L1}^i, y_{L1}^i, x_{L2}^i, y_{L2}^i, \delta_1^i, \alpha, \beta, \gamma, t_x, t_y, t_z) = \\
 & = \sum_{i=0}^{N-1} \left(\left(\alpha \cdot y_{L1}^i \cdot x_{L2}^i + \beta \cdot (x_{L1}^i \cdot x_{L2}^i + f^2) + \gamma \cdot f \cdot y_{L1}^i + t_x \cdot \left(\frac{-f \cdot \delta_1^i}{b} \right) + t_z \cdot \frac{x_{L2}^i \cdot \delta_1^i}{b} - f \cdot (x_{L1}^i - \right. \right. \\
 & \left. \left. x_{L2}^i) \right)^2 + \left(\alpha \cdot (y_{L1}^i \cdot y_{L2}^i + f^2) + \beta \cdot x_{L1}^i \cdot y_{L2}^i + \gamma \cdot (-f \cdot x_{L1}^i) + t_y \cdot \left(\frac{-f \cdot \delta_1^i}{b} \right) + t_z \cdot \frac{y_{L2}^i \cdot \delta_1^i}{b} - f \cdot (y_{L1}^i - \right. \right. \\
 & \left. \left. y_{L2}^i) \right)^2 \right) \quad (10)
 \end{aligned}$$

Notăm cu $\theta = [\theta_i, i = 0, 5] = [\alpha \ \beta \ \gamma \ t_x \ t_y \ t_z]$, vectorul parametrilor de estimat la fiecare pereche de imagini consecutive. De asemenea, notăm cu $p = [p_i, i = 0, N - 1]$, unde $p_i = [p_{i,j}, j = 0, 4] = [x_{L1}^i \ x_{L2}^i \ y_{L1}^i \ y_{L2}^i \ \delta_1^i]$, vectorul parametrilor de intrare pentru fiecare pereche de imagini consecutive, variabili de la o imagine la cealaltă. Cu aceste notații problema odometriei vizuale se transformă în determinarea valorii θ care minimizează $F(\theta, p)$. Pe baza acestor notații, o perturbație \sum_p asupra vectorului de parametrii, reprezentând datele de intrare pentru estimarea odometriei vizuale, va duce la o matrice de covarianță \sum_θ în urma estimării parametrilor de mișcare θ . Detaliile acestei metode sunt expuse în [13]:

$$\sum_\theta = A^{-1} \cdot B \cdot \sum_p \cdot B^T \cdot A^{-1} \quad (11)$$

3.2. Metode liniare de estimare a odometriei vizuale - rezultate experimentale

În [6] am propus o metodă de simplificare a modelului non-linear de estimare din [5]. Pornind de la reprezentarea simplificată a matricii de rotație din ecuația (1), determinarea odometriei vizuale se reduce la rezolvarea unui sistem supradeterminat de ecuații. Pentru fiecare asociere de trăsături între perechi de imagini stereo succesive obținem un set de trei

ecuații precum cel din (3). Dacă avem un număr de K trăsături, atunci vom obține un sistem liniar de $3K$ ecuații.

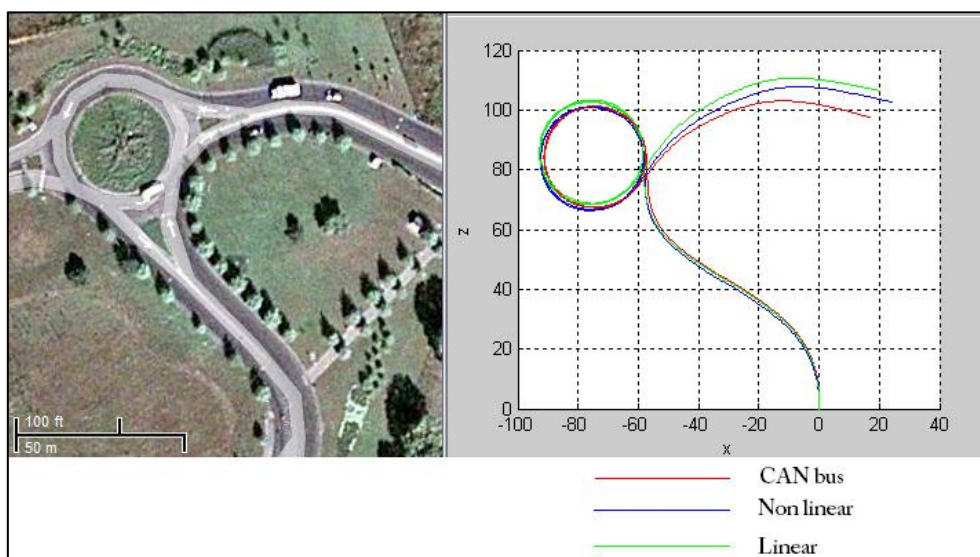


Figura 7 O comparație a traiectoriilor generate cu metoda liniară (verde) și metoda ne-liniară (albastru). Includem ca și referință traiectoria obținută pe baza unor valori de translație și yaw rate citite de pe magistrala CAN a mașinii

Determinarea trăsăturilor care verifică parametrii estimați (specifică metodelor de tip RANSAC), se face prin comparația pozițiilor trăsăturilor din imaginea curentă cu poziția prezisă a acestora pe baza parametrilor de mișcare conform ecuațiilor (2). Figura 7 prezintă o comparație între traiectoria obținută prin compunerea rotațiilor și translațiilor dintre imagini consecutive obținute prin metoda liniară (verde) și metode non-lineare (albastru). Observăm diferențe minore între cele două traiectorii. Ca și referință am inclus și o traiectorie (desenată cu roșu) generată pe baza unghiului yaw și a translației citite de pe magistrala CAN a mașinii.

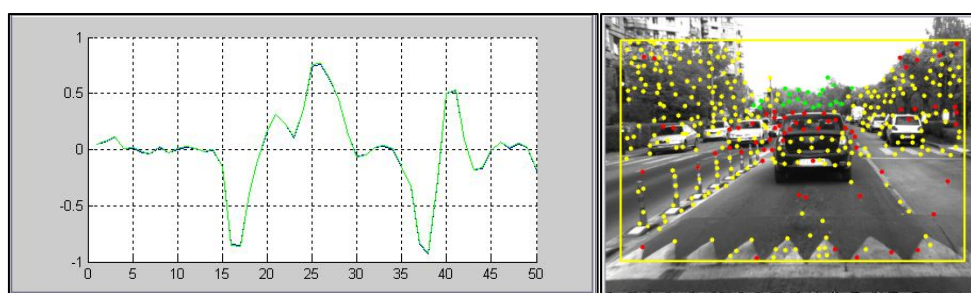


Figura 8 Estimarea unghiului pitch (stânga) la trecerea peste limitatorul de viteză din imaginea dreaptă. Cu albastru este estimarea obținută cu o metodă stereo iar cu verde este estimarea obținută cu metoda pe bază de trăsături aflate la distanță. Trăsăturile de la distanță marcate cu verde sunt folosite pentru estimarea rotației

Un experiment de estimare a rotației folosind puncte de la distanță este ilustrat în Figura 8. Acesta se bazează pe modelul matematic din ecuațiile (4), (5), (6), rezultatele fiind detaliate atât în teză cât și în [12].

Estimarea odometriei vizuale prezintă anumite provocări atunci când imaginile pe baza cărora aceasta este efectuată sunt influențate de factori externi precum condițiile de iluminare care afectează senzorii video (ex: suprasaturare), condiții meteorologice (ceață, ploaie, ninsoare) sau obiecte mișcătoare care ocluzionează părți predominante din scenă. În [10] am propus o metodă de robustificare pe baza transformatei rank inițial propusă în [38]

pentru a îmbunătăți reconstrucția stereo atunci când apar condiții diferite de iluminare între camera stângă și camera dreaptă dintr-un sistem stereo.

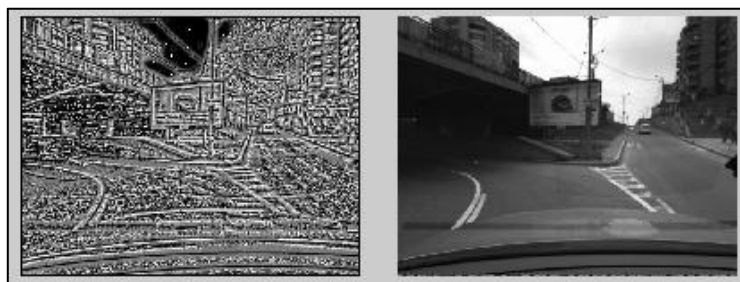


Figura 9 Transformata rank (stânga) pentru o imagine de trafic (dreapta)

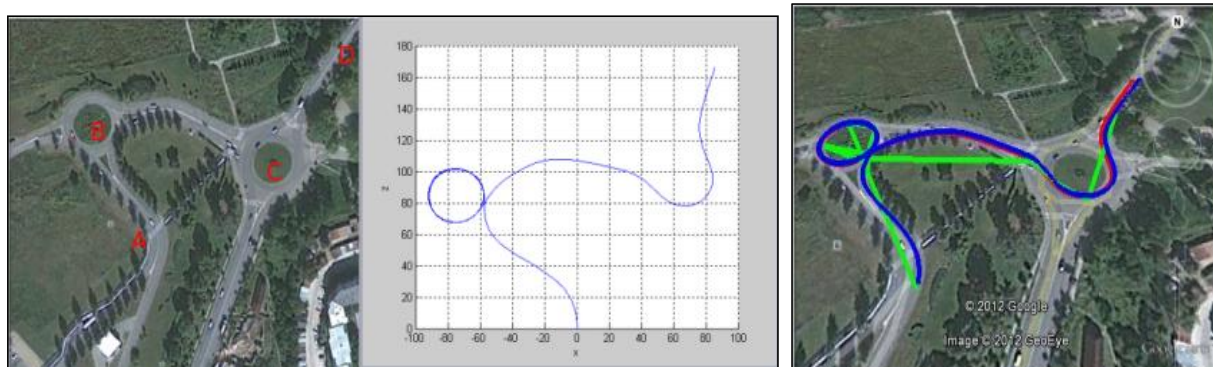


Figura 10 Comparație între traiectoria generată pe baza odometriei vizuale obținute din asocieri de trăsături obținute prin metode diferențiale (albastru în dreapta) și asocieri pe baza de rank transform (roșu în dreapta)

Graficul din mijloc este echivalentul liniei roșii din imaginea dreapta. Cu verde este reprezentată o linie poligonală care unește coordonatele GPS obținute pe ruta A-B-C-D, făcând mai multe rotații în sensul giratoriu B

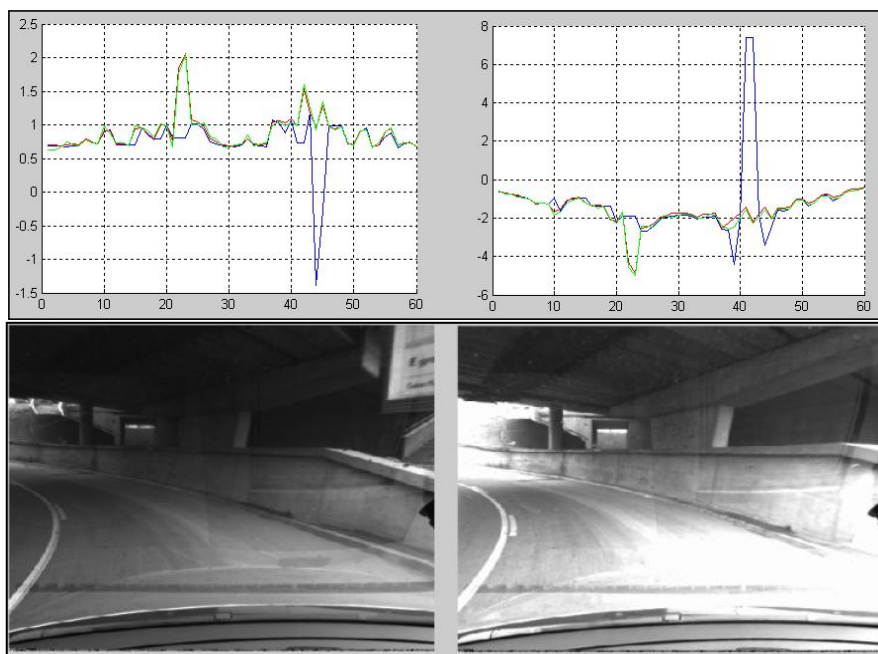


Figura 11 Partea de sus arată evoluția unghiului yaw la intrarea sub un pod. În rândul de jos sunt două imagini succesive având condiții de iluminare diferite datorită trecerii de la soare la umbră. Cu roșu este reprezentat unghiul citit de la magistrala CAN a mașinii, cu verde este reprezentată estimarea din metode de asocieri diferențiale (flux optic), cu verde este estimarea din asocieri de trăsături pe bază de rank transform

Propunerea din [10] calculează trăsături pe baza derivatelor imaginii în perechi de imagini consecutive. În același timp se calculează transformata rank a acestor imagini (Figura 9 ilustrează vizual cum arată această reprezentare, scalată pentru a lua valori între 0 și 255). Imaginea rezultată are proprietatea de a fi invariantă la schimbări de iluminare între imagini succesive. Figurile ilustrează beneficiul adus de trăsăturile invariante la iluminare.

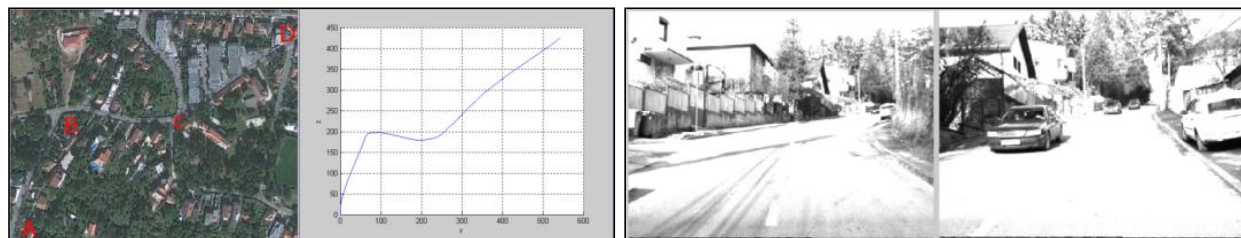


Figura 12 Traiectoria obținută prin integrarea odometriei vizuale obținute pe traseul A-B-C-D din asociere de trăsături pe bază de rank transform. Traseul prezintă zone cu probleme de iluminare precum cele din dreapta

Modelul matematic din (10) și (11) propune determinarea unor formule directe pentru a calcula matricea de covarianță a estimării odometriei vizuale, cunoscând erorile pe care le avem în datele de intrare, și anume, precizia de localizare a trăsăturilor pe baza cărora facem estimarea odometriei. În Figura 13 din stânga am reprezentat translația obținută pe baza modelului liniar de odometrie vizuală, considerând atât o estimare care folosește toate trăsăturile din scenă (verde) cât și două versiuni care consideră doar trăsături aflate la o distanță mai mare decât un anumit prag (15m albastru, 30m violet). Observăm în Figura 13 din dreapta că deviația standard crește odată cu creșterea pragului de filtrare a trăsăturilor, lucru așteptat deoarece erorile de reconstrucție 3D sunt mai mari pentru trăsăturile de la distanță. Cu alte cuvinte, metoda propusă va estima o deviație standard mai mare a translației atunci când nu avem trăsături în apropierea camerei.

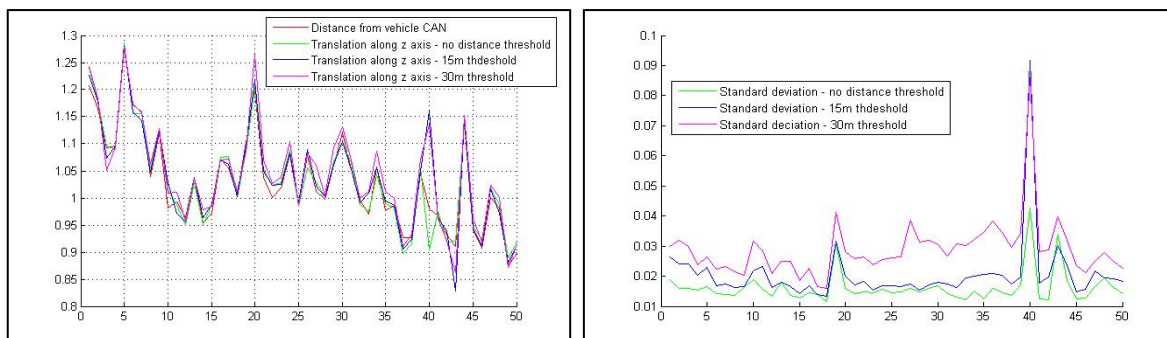


Figura 13 Stânga - translația estimată în lungul axei longitudinale în metri (axa verticală) pentru 50 de imagini consecutive (axa orizontală) cu diverse praguri de distanță pentru filtrarea trăsăturilor. Dreapta - deviația standard pentru translația longitudinală exprimată în metri (axa verticală) pentru cele 50 de imagini consecutive din stânga (axa orizontală) și pentru diverse praguri de distanță

Figura 14 din stânga reprezintă cu verde unghiul yaw pentru aceleași 100 de imagini consecutive ca și cele din Figura 13, în care mașina echipată cu o cameră video se deplasează printr-un sens giratoriu. În graficul din dreapta am reprezentat deviația standard a unghiului yaw, calculată conform ecuației (11). Observăm cele două puncte de pe grafic care sunt încercuite cu roșu. Acestea corespund unor imagini consecutive pentru care variația de iluminare de la una la alta este ridicată. Conform așteptărilor din punct de vedere teoretic, imprecizia de localizare a trăsăturilor cauzată de schimbarea condițiilor de iluminare este reflectată în modelul propus de propagare a erorilor de la nivel de trăsături și asocieri de trăsături la nivel de parametrii de odometrie.

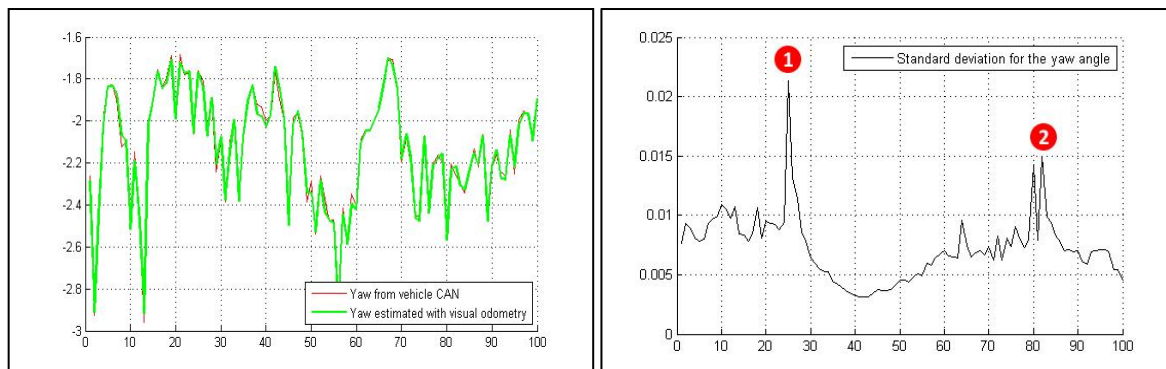


Figura 14 Stânga - unghiul yaw (verde, în grade) calculat pentru aceleași imagini ca și Figura 13 și unghiul yaw citit de la mașină (roșu). Dreapta - deviația standard a unghiului yaw (vertical, exprimată în grade). Cu 1 și 2 sunt marcate imaginile unde avem valori mari ale deviației standard a unghiului yaw datorită schimbărilor bruște de iluminare

Concluzionăm prin a preciza faptul că metoda de propagare a erorilor de detecție și localizare de trăsături vine să complementeze metodele care presupun descriptori invarianți la condiții de iluminare precum și metodele de robustificare a asocierilor de trăsături pentru a obține algoritmi de odometrie vizuală cu un grad ridicat de robustețe.

3.3. Senzori LiDAR pentru estimarea odometriei

Metodele prezentate în capitolul anterior se bazează pe o cameră stereo pentru informația de adâncime, și pentru a determina magnitudinea vectorului de translație. Alternativ sau complementar camerelor stereo, sistemele de percepție din mașinile cu funcții de conducere asistată și automată tind să includă și senzori LiDAR în fuziune cu senzorii video. Datorită preciziei superioare, LiDAR-ul este uneori considerat ca și o alternativă mai bună în ciuda costurilor mai ridicate. Acest lucru a motivat studiul nostru din [11] unde am propus o integrare a informației LiDAR cu o cameră mono, obținând rezultate superioare metodei stereo din [5] și [6].

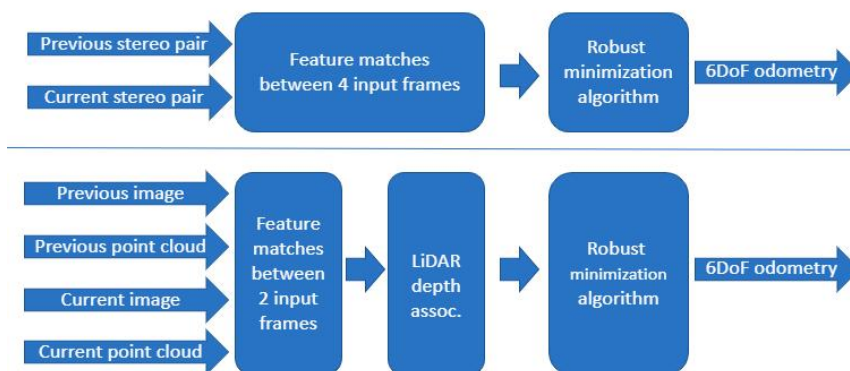


Figura 15 Overview al metodelor pe bază de stereo (sus) comparativ cu metoda pe bază de LiDAR și o cameră monoculară (jos) [11]

3.3.1. Înlocuirea informației stereo de adâncime cu LiDAR

Pornind de la modelul stereo din (2) și (3) am dezvoltat o generalizare care să funcționeze cu orice informații de adâncime, în particular cu un senzor LiDAR. Presupunerea generală este că avem disponibile date de intrare calibrate (imagini rectificate în cazul unui sistem stereo, respectiv calibrare spațială și temporală între sistemul de coordonate al

dispozitivului LiDAR și sistemul de coordonate al camerei. Figura 15 prezintă o comparație din punct de vedere al design-ului de sistem între metoda de odometrie vizuală stereo și cea de odometrie vizuală mono combinat cu LiDAR. Observăm că modulul de asociere de trăsături precum și cel de optimizare robustă (RANSAC) sunt reutilizabile cu mențiunea că aceasta se face între două perechi de imagini stereo, respectiv între două imagini consecutive provenite de la o cameră monoculară. Apare în plus un modul de asociere al măsurătorilor de LiDAR la trăsăturile din planul imagine. Folosirea aceluiași model matematic și prin urmare a aceluiași algoritm de optimizare robustă face posibilă reutilizarea versiunii de implementare paralelă pe care am propus-o în [7].

3.3.2. Generalizarea modelului matematic de odometrie stereo pentru LiDAR

Pornim de la modelul matematic descris anterior. Considerăm ecuația (2) și împărțim ultima ecuație la b:

$$\left\{ \begin{array}{l} \alpha \cdot y_{L1}^i \cdot x_{L2}^i + \beta \cdot (x_{L1}^i \cdot x_{L2}^i + f^2) + \gamma \cdot f \cdot y_{L1}^i + t_x \cdot \left(\frac{-f \cdot \delta_1^i}{b} \right) + t_z \cdot \frac{x_{L2}^i \cdot \delta_1^i}{b} = \\ \quad = f \cdot (x_{L1}^i - x_{L2}^i) \\ \alpha \cdot (y_{L1}^i \cdot y_{L2}^i + f^2) + \beta \cdot x_{L1}^i \cdot y_{L2}^i + \gamma \cdot (-f \cdot x_{L1}^i) + t_y \cdot \left(\frac{-f \cdot \delta_1^i}{b} \right) + t_z \cdot \frac{y_{L2}^i \cdot \delta_1^i}{b} = \\ \quad = f \cdot (y_{L1}^i - y_{L2}^i) \\ \alpha \cdot y_{L1}^i \cdot \frac{\delta_2^i}{b} + \beta \cdot x_{L1}^i \cdot \frac{\delta_2^i}{b} + t_z \cdot \frac{\delta_2^i \cdot \delta_1^i}{b \cdot b} = f \cdot \frac{(\delta_1^i - \delta_2^i)}{b} \end{array} \right. \quad (12)$$

Am subliniat cu albastru cantitățile care sunt specifice unui sistem stereo de camere.

Din modelul de reconstrucție stereo, $\frac{\delta_1^i}{b}$ și $\frac{\delta_2^i}{b}$ pot fi înlocuite cu $\frac{f}{Z_1^i}$ respectiv cu $\frac{f}{Z_2^i}$, valori care depind doar de distanța focală a camerei și de distanțele la care se află trăsăturile în prima, respectiv în a doua imagine. Altfel spus, dacă avem distanța calculată din LiDAR sau alte surse de măsurare a profunzimii obținem un sistem de ecuații identic cu cel utilizat pentru metoda stereo. Datorită densității mai mici a măsurătorilor de LiDAR comparativ cu cele de video, determinarea distanțelor Z_1^i și Z_2^i nu este trivială și o vom explica în subcapitolul următor.

3.3.3. Determinarea distanțelor trăsăturilor video pe bază de LiDAR

În [11] am folosit date de la un senzor LiDAR care este aliniat cu camera. În prima fază proiectăm în imagine punctele 3D transformate anterior din sistemul de coordonate al senzorului LiDAR în sistemul de coordonate al camerei și reținem în continuare poziția 3D a acestora (Figura 16).

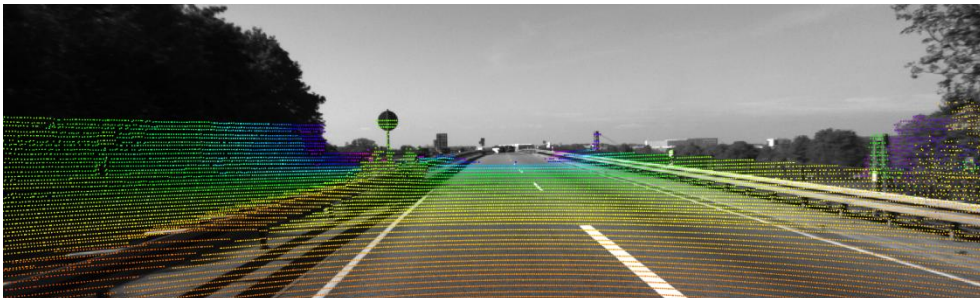


Figura 16 Puncte 3D provenite de la un senzor LiDAR, transformate în sistemul camerei și proiectate în planul imagine

Vom construi apoi o triangulare Delaunay [43] pe baza acestor proiecții. Figura 17 ilustrează o reprezentare grafică a acestei triangulări. Pentru fiecare dintre trăsătura din imagine, candidată pentru a fi folosită pentru estimarea odometriei vizuale, vom determina triunghiul în care aceasta este situată, apoi vom estima distanța acesteia printr-o interpolare liniară a distanțelor corespunzătoare coordonatelor 3D a vârfurilor acestui triunghi. De asemenea, efectuăm o filtrare a trăsăturilor pe care le folosim în continuare în cadrul procesului de estimare a odometriei, pe baza proprietăților geometrice triunghiurilor în care acestea se găsesc. Motivația acestei filtrări este dată de faptul că atunci când facem interpolarea pe baza distanțelor corespunzătoare vârfurilor triunghiului presupunem că acel triunghi corespunde unei suprafețe 3D planare și prin urmare filtrăm acele triunghiuri care nu au astfel de proprietăți.

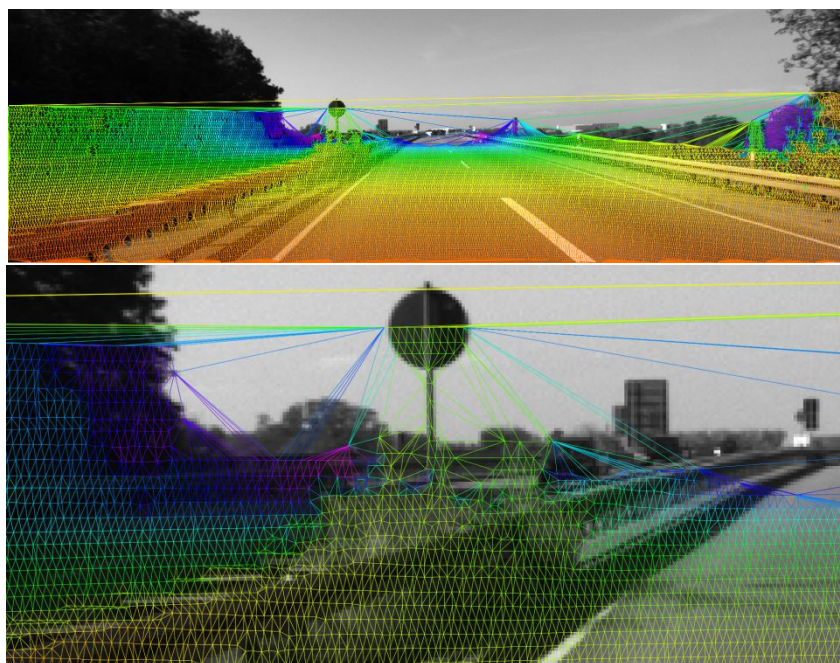


Figura 17 Triangulare Delaunay construită pe baza punctelor LiDAR proiectate în imagine (sus) și o versiune mărită (jos) pentru o mai bună vizualizare a triangulării

3.3.4. Rezultate experimentale și concluzii

Expunem în continuare o parte dintre rezultatele experimentale din [11] pentru a ilustra superioritatea utilizării distanțelor din LiDAR în condiții bune de calibrare, precum și efectele decalibrării dintre senzori asupra metodei. Evaluăm comparativ 3 metode:

- A. metoda liniară al cărei model matematic este detaliat în capitolul 3.1
- B. o metodă hibrid care preia distanțele LiDAR acolo unde aceasta este posibil și păstrează trăsăturile neacoperite de LiDAR cu disparitățile corespunzătoare acestora
- C. metoda care preia distanțele LiDAR acolo unde aceasta este posibil și ignoră trăsăturile neacoperite de măsurători LiDAR

	Eroarea de translație	Eroarea de rotație
A	1.31%	0.00441 [deg/m]
B	1.24%	0.00441 [deg/m]
C	1.1%	0.00412 [deg/m]

Tabel 2 Erorile de translație și rotație pentru configurațiile A (stereo), B (LiDAR combinat cu stereo) și C (LiDAR)

Tabelul 2 ilustrează performanța celor trei configurații pe setul de 10 secvențe KITTI [32]. Observăm performanța superioară obținută pentru estimarea translației, configurația A atingând o eroare de 1.31% din distanța parcursă în comparație cu C care are o eroare de 1.1% din distanța parcursă.

Un alt experiment pe care l-am făcut în [11], și ale cărui rezultate experimentale sunt reluate în teză se referă la evaluarea impactului decalibrărilor care pot apărea între LiDAR și camera video asupra odometriei vizuale. Concluzionăm că erori de decalibrare pe translație de +/- 0.2m în direcția longitudinală au un efect mic asupra estimării, iar cu erori de decalibrare mai mari de 0.2m precizia odometriei pe bază de LiDAR și video devine mai puțin mică decât cea a unei metode stereo. De asemenea, observăm că valori de decalibrare de ordinul +/-2px conduc la erori de estimare ale odometriei vizuale similare cu cele ale metodei stereo, în timp ce erori de ordinul 5-10px degradează semnificativ precizia de măsurare a odometriei. Valorile decalibrărilor de rotație sunt exprimate în pixeli pentru o mai bună intuiție a gradului de aliniere necesar între cei doi senzori. Practic, unghiul de decalibrare se poate calcula ca fiind $\varphi = \text{atan}(\Delta p / f_x)$, unde Δp este decalibrarea exprimată în pixeli și invers $\Delta p = f_x \cdot \tan(\varphi)$.

Concluzionăm prin a puncta faptul că generalizarea metodei stereo de odometrie vizuală înspre alți senzori de măsurare a distanțelor a fost evaluată pe bază de LiDAR, demonstrând rezultate superioare, dar și reutilizând în același model matematic, aceleași module de asociere de trăsături și același algoritm de optimizare robustă. De asemenea, considerăm că metoda poate funcționa și cu alte surse de estimare a distanțelor precum metode din ce în ce mai promițătoare pe bază de deep learning [39][40].

3.4. Folosirea de senzori LiDAR pentru corecția odometriei vizuale

Pentru a rezolva probleme de localizare și cartografiere, estimarea odometriei vizuale între imagini succesive se compun succesiv, obținând traiectoria camerei. Prin această compunere temporală a estimărilor succesive, erorile care apar la fiecare estimare dintre imagini succesive se compun la rândul lor, rezultând într-o deviere de la traiectoria reală care se amplifică în timp. Cu cât compunem mai multe estimări succesive, cu atât devierea este mai mare, deoarece fiecare mișcare relativă dintre imagini consecutive pe care o compunem pentru a obține traiectoria globală contribuie cu erorile sale de estimare.

În [17] am propus o metodă de atenuare a erorilor acumulate la integrarea măsurătorilor succesive de odometrie vizuală pe bază de măsurători LiDAR. Deoarece aceste erori sunt suficient de mici ca și ordin de magnitudine, am propus o metodă de formulare a problemei sub forma unui sistem liniar de ecuații, eliminând în acest mod nevoia de a folosi metode de optimizare neliniară care sunt costisitoare din punct de vedere computațional. Aceasta conduce la o metodă mai eficientă computațional decât alte abordări din literatură. Lucrarea noastră din [17] este influențată mai ales de [41][42]. Vom prelua din [41] ideea de a detecta puncte de muchie și puncte planare într-un moment de timp, de a le asocia muchiilor corespunzătoare respectiv planelor corespunzătoare achiziționate într-un alt moment de timp. Folosind aceste trăsături, dezvoltăm un model matematic care presupune rezolvarea unui sistem de ecuații pentru a determina și corecta erorile de integrare a mișcărilor succesive.

3.4.1. Prezentare generală a sistemului

Figura 18 prezintă o diagramă block a sistemului de corecție a erorilor de integrare temporală a odometriei vizuale și contextul mai larg de estimare a odometriei vizuale între

imagini succesive. Odometria între imagini succesive este reprezentată prin blocuri de culoare gri.

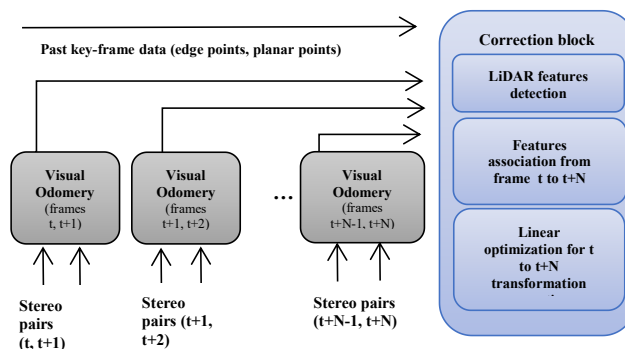


Figura 18 Diagrama bloc a sistemului de corecție a odometriei vizuale [17]

Imaginile cheie între care se execută rutina de corecție sunt selectate pe baza distanței parcurse de la ultima corecție (sau de la inițializarea sistemului). Rutina de corecție se bazează pe măsurătorile LiDAR asociate cu imaginea cheie trecută și respectiv cu imaginea cheie curentă. Într-o primă etapă detectăm puncte de muchie și puncte planare (vezi Figura 19). Punctele de muchie asociate cu imaginea cheie anterioară sunt asociate cu muchii 3D din imaginea cheie curentă. La fel, puncte planare din imaginea cheie anterioară sunt asociate cu plane 3D din imaginea cheie curentă. Transformatele cunoscute dintre imagini consecutive pot fi compuse, obținând o transformată totală dintre imaginea cheie anterioară (t) și imaginea cheie curentă ($t+N$). Această transformată dintre imaginea (t) și imaginea ($t+N$) este folosită pentru a reduce spațiul de căutare la asocierea dintre puncte de muchie și de plan din imaginea cheie anterioară (t) la muchii și plane 3D în imaginea cheie curentă ($t+N$).

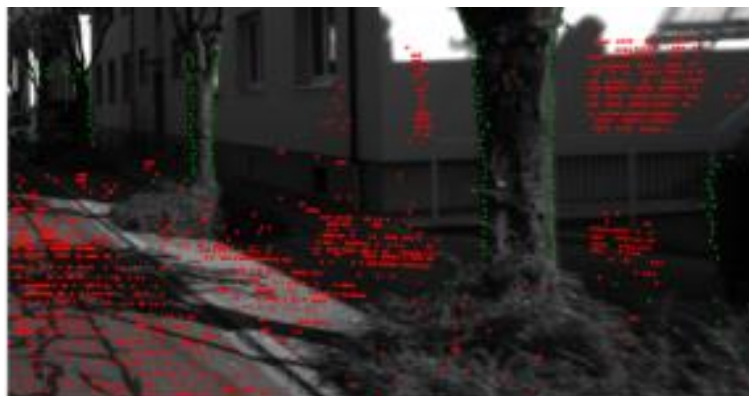


Figura 19 Clasificarea punctelor LiDAR în puncte planare (roșu) și puncte de muchie (verde)

Odată ce corespondențele sunt disponibile formulăm o problemă de optimizare care estimează o transformată de corecție prin rezolvarea unui sistem liniar de ecuații specificând următoarele constrângeri:

- ✓ punctele de muchie propagate de la imaginea cheie anterioară (t) trebuie să aparțină muchiilor cu care acestea sunt asociate în imaginea cheie curentă ($t+N$)
- ✓ punctele planare propagate de la imaginea cheie anterioară (t) trebuie să aparțină planelor cu care acestea sunt asociate în imaginea cheie curentă ($t+N$)

Calcularea acestor transformări de corecție între imagini cheie dintr-un stream video și utilizarea lor atunci când facem integrarea temporală a estimărilor de odometrie vizuală conduc la o acuratețe mai bună a traiectoriei estimate. Practic erorile acumulate la integrarea în timp a odometriei vizuale sunt corectate de acest mecanism care intervine după

parcurgerea unei anumite distanțe prin selecția acelor imagini cheie între care se face rutina de corecție.

3.4.2. Modelul matematic

Fie $\{P_i^t\}$ setul de trăsături LiDAR corespunzătoare imaginii numărul t . Pentru o trăsătură de muchie $P_i^{t_1}$ din sistemul de coordonate asociat imaginii numărul t_1 asociem cele mai apropiate două puncte de muchie aflate pe linii de achiziție diferite ale LiDAR-ului $P_j^{t_2}, P_k^{t_2}$. Punctul $P_i^{t_1}$ este propagat în sistemul de coordonate al imaginii t_2 rezultând $\bar{P}_i^{t_2} = R^{t_2, t_1} \cdot P_i^{t_1} + T^{t_2, t_1}$. După transformarea lui $P_i^{t_1}$ setul de puncte $\bar{P}_i^{t_2}, P_j^{t_2}, P_k^{t_2}$ sunt în același sistem de coordonate și aparțin aceleiași muchii (deci sunt colineare). Cu toate acestea, datorită erorilor de estimare a odometriei vizuale care sunt incluse în valorile parametrilor de transformare $R^{t_2, t_1}, T^{t_2, t_1}$ cele trei puncte nu vor fi perfect colineare, și scopul nostru va fi acela de a determina o transformată de corecție R_c, T_c care să le facă colineare. Făcând presupunerea că unghiurile de rotație sunt mici putem considera:

$$R_c = \begin{bmatrix} 1 & -\gamma & -\beta \\ \gamma & 1 & -\alpha \\ \beta & \alpha & 1 \end{bmatrix} \quad T_c = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (13)$$

Pentru simplitatea notațiilor, în restul acestui capitol vom renunța la indexul t_2 când reprezentăm explicit coordonatele 3D ale punctelor:

$$\bar{P}_i^{t_2} = \begin{bmatrix} X^i \\ Y^i \\ Z^i \end{bmatrix}, P_j^{t_2} = \begin{bmatrix} X^j \\ Y^j \\ Z^j \end{bmatrix}, P_k^{t_2} = \begin{bmatrix} X^k \\ Y^k \\ Z^k \end{bmatrix} \quad (14)$$

Matricea de corecție este aplicată pentru $\bar{P}_i^{t_2}$. Vom scrie condiția de coliniaritate pentru $R_c \cdot \bar{P}_i^{t_2} + T_c, P_j^{t_2}$ și $P_k^{t_2}$ astfel:

$$\frac{X^i - \gamma \cdot Y^i - \beta \cdot Z^i + t_x - X^k}{X^j - X^k} = \frac{\gamma \cdot X^i + Y^i - \alpha \cdot Z^i + t_y - Y^k}{Y^j - Y^k} \quad (15)$$

$$\frac{X^i - \gamma \cdot Y^i - \beta \cdot Z^i + t_x - X^k}{X^j - X^k} = \frac{\beta \cdot X^i + \alpha \cdot Y^i + Z^i + t_z - Z^k}{Z^j - Z^k} \quad (16)$$

Rezultă că dintr-o asociere de punct de muchie obținem două ecuații care constrâng vectorul necunoscut de parametri ai transformatei de corecție: $[\alpha \quad \beta \quad \gamma \quad t_x \quad t_y \quad t_z]$.

Similar cu modul de gestionare a trăsăturilor de muchie vom gestiona și trăsăturile planare. Pentru o trăsătură planară $P_i^{t_1}$ din sistemul de coordonate asociat imaginii numărul t_1 asociem cele mai apropiate trei puncte de planare aflate pe linii de achiziție diferite ale LiDAR-ului $P_j^{t_2}, P_k^{t_2}, P_l^{t_2}$. Punctul $P_i^{t_1}$ este propagat în sistemul de coordonate al imaginii t_2 astfel: $\bar{P}_i^{t_2} = R^{t_2, t_1} \cdot P_i^{t_1} + T^{t_2, t_1}$. Apoi scopul este de a găsi o transformată de corecție R_c, T_c care face punctele $R_c \cdot \bar{P}_i^{t_2} + T_c, P_j^{t_2}, P_k^{t_2}, P_l^{t_2}$ coplanare. Aceasta poate fi exprimată sub forma unui determinant, rezultând o nouă ecuație liniară care constrânge transformata de corecție R_c, T_c reprezentată ca și vector de șase parametri:

$$\begin{vmatrix} X^i - \gamma \cdot Y^i - \beta \cdot Z^i + t_x - X^k & X^j - X^k & X^l - X^k \\ \gamma \cdot X^i + Y^i - \alpha \cdot Z^i + t_y - Y^k & Y^j - Y^k & Y^l - Y^k \\ \beta \cdot X^i + \alpha \cdot Y^i + Z^i + t_z - Z^k & Z^j - Z^k & Z^l - Z^k \end{vmatrix} = 0 \quad (17)$$

3.4.3. Rezultate și concluzii

Experimentele le-am făcut pe setul de date KITTI [32] care dispune atât de date stereo calibrate cât și de măsurători achiziționate cu un LiDAR și care sunt transformate în sistemul de coordonate al camerei stângi, permițând astfel evaluarea acestor metode de fuziune. Corecția se bazează pe selecția de trăsături de muchie și de trăsături planare estimate în ordinea de achiziție a măsurătorilor LiDAR (linii de măsurare orizontale). Aceasta poate să conducă la o capacitate mai redusă de a constrânge anumite grade de libertate, prin urmare evaluăm potențialul metodei de a corecta diferite componente ale parametrilor de mișcare $[\alpha \ \beta \ \gamma \ t_x \ t_y \ t_z]$. Considerăm patru subseturi de parametri pe care încercăm să-i corectăm:

- ✓ corectăm setul întreg de parametri $[\alpha \ \beta \ \gamma \ t_x \ t_y \ t_z]$
- ✓ excludem unghiul roll și corectăm restul $[\alpha \ \beta \ t_x \ t_y \ t_z]$
- ✓ corectăm unghiul yaw și translația $[\beta \ t_x \ t_y \ t_z]$
- ✓ corectăm yaw și translația în planul orizontal $[\beta \ t_x \ t_z]$

Pentru dimensiunile vectorului de parametri pe care le excludem din procesul de corecție setăm coeficienții corespunzători din ecuațiile (15), (16) și (17) la zero. Tabelul 3 subliniază faptul că varianta optimă este să corectăm unghiul yaw și vectorul de translație. Figura 20 arată exemple de traiectorii corectate comparativ cu cele necorectate.

Parameter subset	Translation error (% from traveled distance)	Rotation error (radians/m)
Baseline (no correction)	1.10	0.000063
$[\alpha \ \beta \ \gamma \ t_x \ t_y \ t_z]$	1.21	0.000101
$[\alpha \ \beta \ t_x \ t_y \ t_z]$	1.07	0.000081
$[\beta \ t_x \ t_y \ t_z]$	0.87	0.000064
$[\beta \ t_x \ t_z]$	0.89	0.000064

Tabel 3 Evaluarea performanței de corecție a odometriei vizuale pe bază de LiDAR. Liniile tabelului indică diferite configurații de parametri pe care-i corectăm

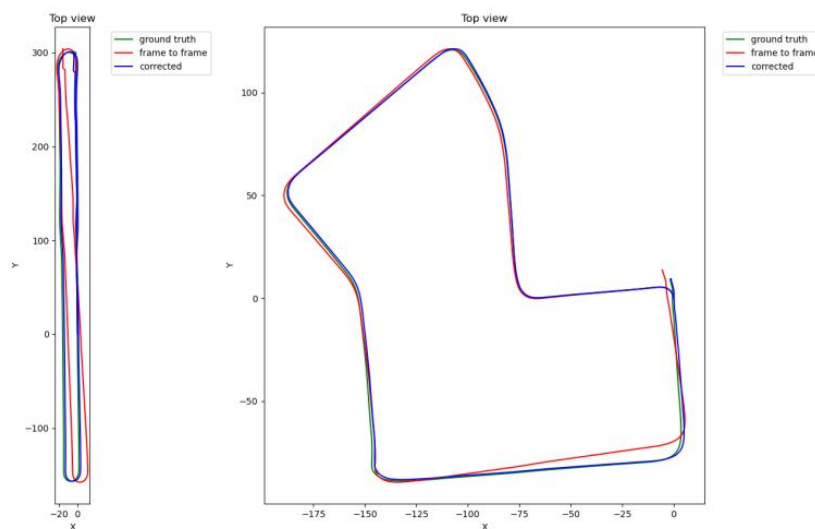


Figura 20 Exemple de traiectorii corectate (secvența KITTI 06 în stânga și secvența 07 în dreapta). Cu verde este reprezentată traiectoria de referință (în foarte mare parte acoperită de traiectorii estimate și corectate), cu albastru este traiectoria corectată prin metodele propuse în [17], cu roșu este traiectoria generată cu o metodă care face estimarea strict din imagini succesive

Concluzionăm prin a puncta că am propus o metodă de corecție a erorilor care se acumulează la compunerea traiectoriei din estimări din imagini succesive ale odometriei vizuale. Aceasta se bazează pe trăsături de tip muchie și pe trăsături planare dintr-o imagine cheie anterioară, asociate la muchiile și respectiv la planele corespunzătoare în imaginile cheie curente. Deoarece erorile de acumulare sunt mici, am propus o metodă de liniarizare a problemei de corecție, formulând-o sub forma unui sistem liniar de ecuații care se poate rezolva eficient din punct de vedere computațional. Rezultatele experimentale arată o îmbunătățire a preciziei de estimare a traiectoriei pe setul de date KITTI [32] comparativ cu abordările care se bazează doar pe compunerea estimărilor dintre imagini succesive. Considerăm că printr-o rafinare a parametrilor algoritmului se pot obține în continuare performanțe ușor îmbunătățite (o reducere de 2-3‰ din distanța parcursă a erorii de translație). De asemenea există oportunitatea de a experimenta în ce măsură este posibil să folosim o categorizare semantică a punctelor din LiDAR pentru o clasificare mai robustă a punctelor de muchie și a celor planare. Nu în ultimul rând includerea unor muchii orizontale care să constrângă mai bine unghiurile de roll și de pitch poate să fie benefică și să aducă un plus de performanță în reconstrucția traiectoriei.

4. Concluzii

4.1. Concluzii generale

Lucrarea de față abordează problema estimării eficiente a odometriei pe baza unei camere video stereo sau a unei camere video mono complementată de un senzor de măsurare a distanțierelor. Odometria vizuală este o problemă consacrată și intens studiată în literatura de specialitate așa cum am arătat în paginile care descriu stadiul actual al cunoașterii, punctând atât primele abordări care au pus bazele teoretice ale odometriei vizuale și în general ale viziunii artificiale cât și primele metode care au funcționat în timp real. De asemenea am făcut un studiu al principalelor metode de robustificare a algoritmilor de odometrie vizuală pe bază de trăsături, metode care stau la baza contribuțiilor din această lucrare. Considerăm că metodele de odometrie vizuală bazate pe o cameră mono, stereo sau omnidirecțională vor evolua în trei direcții: sisteme multi-cameră, fuziune cu alți senzori, utilizarea rețelelor neuronale convoluționale. Din acest motiv am dedicat câte un subcapitol de analiză al stadiului actual al cunoașterii în aceste trei direcții, punctând și posibilele dezvoltări ulterioare ale contribuțiilor din această teză în cele trei direcții menționate.

Lucrarea construiește o formulare matematică unitară pentru modelele stereo de odometrie vizuală și detaliază ulterior aspectele specifice legat de formularea problemei de optimizare pentru estimarea odometriei vizuale mono și stereo ca și un sistem liniar de ecuații.

Din punct de vedere al implementării practice explicăm în detaliu trăsăturile folosite pentru odometria vizuală și prezentăm apoi rezultatele experimentale obținute, folosind metode liniare de rezolvare a problemei comparativ cu metode de optimizare non-liniară, concluzia fiind că metodele liniare propuse ating aceeași performanță ca și metodele non-liniare din literatură și necesită resurse de calcul mai puține.

Punctăm de asemenea că generalizarea metodelor stereo pentru a fi utilizare pentru alți

senzori de adâncime (de exemplu LiDAR) este o contribuție semnificativă a acestei lucrări. De asemenea metodele de postprocesare a traiectoriei obținute pe bază de senzori LiDAR pentru a reduce erorile de acumulare la concatenarea măsurătorilor succesive ale odometriei vin să completeze metodele de estimare dintre imagini succesive, scăzând eroarea de poziționare pe bază de odometrie vizuală.

Estimarea vitezei relative a obiectelor în mișcare folosind metode de odometrie vizuală este o contribuție care merită exploatată în continuare ca și aplicație a metodelor de odometrie vizuală.

4.2. Originalitatea și contribuțiile inovative ale tezei

Din punct de vedere formal, putem grupa contribuțiile și inovațiile acestei lucrări în aspecte teoretice și aspecte practice. Contribuțiile teoretice sunt următoarele:

- ✓ Dezvoltarea unui model matematic de estimare a odometriei vizuale prin rezolvarea unui sistem de ecuații [6]
- ✓ Particularizarea modelului liniar de estimare a odometriei vizuale pentru a estima rotația camerei folosind puncte de la distanță [12]
- ✓ Dezvoltarea unui model analitic de propagare a erorilor pe care le prezintă trăsăturile folosite în odometria vizuală pentru a estima erorile de estimare ale mișcării [13]
- ✓ Dezvoltarea unei generalizări a metodei liniare de odometrie vizuală care permite utilizarea acestora și în scenarii în care rotația relativă dintre imagini succesive este mare, eventual folosind rotația estimată din puncte de la distanță [16]
- ✓ Dezvoltarea unui model matematic pentru a corecta erorile cumulate la concatenarea măsurătorilor succesive de odometrie vizuală pe bază de LiDAR [17]

Din punct de vedere practic, punctăm următoarele contribuții:

- ✓ Evaluarea metodelor liniare de odometrie vizuală comparativ cu metode non-lineare [5] și cu metode bazate pe tensorul trifocal
- ✓ Paralelizarea algoritmilor de odometrie vizuală pe arhitecturi paralele de calcul [7]
- ✓ Studiul trăsăturilor și a descriptorilor din literatura de specialitate și evaluarea unor metode de robustificare a algoritmilor de odometrie vizuală pe bază de descriptori invarianți la schimbarea condițiilor de iluminare [10]
- ✓ Preprocesarea și verificarea trăsăturilor de odometrie vizuală pentru a elimina trăsăturile eronate: consistența unghiurilor dintre perechi de trăsături, verificare circulară, inițializarea metodelor diferențiale și reducerea spațiului de căutare pe baza mișcării estimate anterior [8][9][10]
- ✓ Generalizarea metodei liniare de odometrie vizuală propusă în [6] pentru a folosi alte surse de informație de adâncime precum LiDAR [11]
- ✓ Folosirea metodelor de odometrie vizuală pentru a estima vitezele relative dintre obiecte [14][15]

4.3. Posibile dezvoltări ulterioare

Punctăm de asemenea câteva direcții noi care considerăm că merită explorate în continuare:

- ✓ Folosirea de trăsături dedicate pentru odometrie vizuală, calculate folosind rețele neuronale convoluționale
- ✓ Folosirea de hărți de adâncime pe bază de rețele neuronale convoluționale
- ✓ Robustificarea metodei liniare de odometrie vizuală prin ponderarea diferită a ecuațiilor

din sistem

- ✓ Folosirea odometriei vizuale în contextul localizării și cartografierii simultane
- ✓ Generalizarea modelului matematic liniar de estimare a odometriei vizuale pentru a considera mai multe imagini consecutive (sliding window)
- ✓ Generalizarea modelului matematic liniar de estimare a odometriei vizuale pentru sisteme cu mai multe camere

Încheiem prin a menționa faptul că deși metodele de odometrie vizuală au ajuns la o maturitate suficient de mare pentru a fi incluse în produse disponibile pentru publicul larg, oportunitățile de cercetare în domeniu rămân în continuare numeroase mai ales datorită unor factori precum: creșterea rezoluțiilor camerelor, nevoia de sisteme multi-cameră pentru o percepție surround view, disponibilitatea puterii de calcul crescute și a metodelor de percepție pe bază de rețele neuronale convoluționale, sisteme multi-senzor în care măsurătorile video sunt fuzionate cu informații de la senzori de localizare (ex: GPS), senzori inerțiali sau alți senzori de percepție (LiDAR, camere infraroșii etc).

Referințe

- [1] David A. Forsyth, Jean Ponce, Computer Vision: A Modern Approach, 2nd Edition, November 5, 2011
- [2] Richard Hartley, Andrew Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2nd edition, April 19, 2004
- [3] Richard Szeliski, "Computer Vision: Algorithms and Applications", Springer, November 24, 2010
- [4] Trucco, Emanuele, and Alessandro Verri. Introductory techniques for 3-D computer vision. Vol. 201. Englewood Cliffs: Prentice Hall, 1998.
- [5] Golban, Cătălin, Ionuț Golban, and Sergiu Nedeveschi. "Vision based three-dimensional vehicle motion detection by minimizing nonlinear functions." Proceedings of the 2010 IEEE 6th International Conference on Intelligent Computer Communication and Processing. IEEE, 2010.
- [6] Golban, Cătălin, and Sergiu Nedeveschi. "Linear vs. non linear minimization in stereo visual odometry." 2011 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2011.
- [7] Istvan, Szakats, Cătălin Golban, and Sergiu Nedeveschi. "Fast vision based ego-motion estimation from stereo sequences—A GPU approach." 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, 2011.
- [8] Nedeveschi, Sergiu, Cătălin Golban, and Cosmin Mitran. "Improving accuracy for Ego vehicle motion estimation using epipolar geometry." 2009 12th International IEEE Conference on Intelligent Transportation Systems. IEEE, 2009.
- [9] Golban, Cătălin, Cosmin Mitran, and Sergiu Nedeveschi. "A practical method for ego vehicle motion estimation from video." 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing. IEEE, 2009.
- [10] Golban, Cătălin, Szakats Istvan, and Sergiu Nedeveschi. "Stereo based visual odometry in difficult traffic scenes." 2012 IEEE Intelligent Vehicles Symposium. IEEE, 2012.
- [11] Golban, Catalin, Petrut Cobarzan, and Sergiu Nedeveschi. "A Comparison Study on Replacing Stereo Disparity with LiDAR in Visual Odometry Methods." 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE, 2020.

- [12] Golban, Catalin, and Sergiu Nedevschi. "An experiment on relative rotation estimation from distant points with monocular vision." 2013 IEEE 9th International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE, 2013.
- [13] Golban, Catalin, Petrut Cobarzan, and Sergiu Nedevschi. "Direct formulas for stereo-based visual odometry error modeling." 2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE, 2015.
- [14] Golban, Catalin, and Sergiu Nedevschi. "Speed estimation for scene objects using stereo visual odometry methods." 2013 IEEE 9th International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE, 2013.
- [15] Golban, Catalin, and Sergiu Nedevschi. "Moving rigid objects segmentation in 3D dynamic traffic scenes using a stereovision system." 2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE, 2014.
- [16] Golban, Catalin, Petrut Cobarzan, and Sergiu Nedevschi. "On visual odometry estimation using linear methods - a mathematical framework" 2020 Automation, Computers, Applied Mathematics (ACAM).
- [17] Golban, Catalin, Petrut Cobarzan, and Sergiu Nedevschi. "Visual Odometry Drift Reduction Based on LiDAR Point Clouds Alignment" 2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE, 2021.
- [18] Nistér, David, Oleg Naroditsky, and James Bergen. "Visual odometry." Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.. Vol. 1. IEEE, 2004.
- [19] Agrawal, M., & Konolige, K. (2007, August). Rough terrain visual odometry. In Proceedings of the International Conference on Advanced Robotics (ICAR) (Vol. 25, pp. 28-30).
- [20] Agrawal, M., & Konolige, K. (2006, August). Real-time localization in outdoor environments using stereo vision and inexpensive gps. In 18th International conference on pattern recognition (ICPR'06) (Vol. 3, pp. 1063-1068). IEEE.
- [21] Howard, A. (2008, September). Real-time stereo visual odometry for autonomous ground vehicles. In 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 3946-3952). IEEE.
- [22] Cheng, Y., Maimone, M., & Matthies, L. (2005, October). Visual odometry on the Mars exploration rovers. In 2005 IEEE International Conference on Systems, Man and Cybernetics (Vol. 1, pp. 903-910). IEEE.
- [23] Badino, Hernán, and Takeo Kanade. "A Head-Wearable Short-Baseline Stereo System for the Simultaneous Estimation of Structure and Motion." MVA. 2011.
- [24] Chang, Ping-Lin, Ankur Handa, Andrew J. Davison, Danail Stoyanov and Philip J. Edwards. "Robust Real-Time Visual Odometry for Stereo Endoscopy Using Dense Quadrfocal Tracking." IPCAI (2014).
- [25] Grasa, Óscar G., Ernesto Bernal, Santiago Casado, Ismael Gil and J. M. M. Montiel. "Visual SLAM for Hand-Held Monocular Endoscope." (2013).
- [26] Lourenço, Miguel & Stoyanov, Danail & Barreto, Joao P. (2014). Visual Odometry in Stereo Endoscopy by Using PEaRL to Handle Partial Scene Deformation. 33-40. 10.1007/978-3-319-10437-9_4.
- [27] Stap, Nanda van der, Ferdinand van der Heijden and Ivo Amj Broeders. "Towards automated visual flexible endoscope navigation." Surgical Endoscopy 27 (2013): 3539-3547.
- [28] Wirth, Stephan, Pep Lluís Negre Carrasco, and Gabriel Oliver Codina. "Visual odometry for autonomous underwater vehicles." 2013 MTS/IEEE OCEANS-Bergen. IEEE, 2013.
- [29] Etxeberria-Garcia, M., Labayen, M., Zamalloa, M., & Arana-Arexolaleiba, N. (2020, January). Application of Computer Vision and Deep Learning in the railway domain for autonomous train stop operation. In 2020 IEEE/SICE International Symposium on System Integration (SII) (pp. 943-948). IEEE.
- [30] Tschopp, F., Schneider, T., Palmer, A. W., Nourani-Vatani, N., Cadena, C., Siegwart, R., & Nieto, J. (2019). Experimental comparison of visual-aided odometry methods for rail vehicles. IEEE Robotics and Automation Letters, 4(2), 1815-1822.

- [31] Zaman, Shahzad, et al. "Cost-effective visual odometry system for vehicle motion control in agricultural environments." *Computers and Electronics in Agriculture* 162 (2019): 82-94.
- [32] Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012.
- [33] Schöps, Thomas, Jakob Engel, and Daniel Cremers. "Semi-dense visual odometry for AR on a smartphone." 2014 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, 2014.
- [34] Cvišić, Igor, et al. "SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles." *Journal of field robotics* 35.4 (2018): 578-595.
- [35] Geiger, A., Ziegler, J., & Stiller, C. (2011, June). Stereoscan: Dense 3d reconstruction in real-time. In 2011 IEEE Intelligent Vehicles Symposium (IV) (pp. 963-968). IEEE.
- [36] Warren, Michael, and Ben Upcroft. "High altitude stereo visual odometry." *Proceedings of Robotics: Science and Systems IX* (2013).
- [37] Kanellakis, C., & Nikolakopoulos, G. (2017). Survey on computer vision for UAVs: Current developments and trends. *Journal of Intelligent & Robotic Systems*, 87(1), 141-168.
- [38] Ramin Zabih, John Woodfill, "Non-parametric local transforms for computing visual correspondence", ECCV 1994, *Proceedings of the Third European Conference on Computer Vision - Volume II*, pp. 151-158
- [39] Amiri, Ali Jahani, Shing Yan Loo, and Hong Zhang. "Semi-supervised monocular depth estimation with left-right consistency using deep neural network." 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2019
- [40] Tang, J., Ambrus, R., Guizilini, V., Pillai, S., Kim, H., & Gaidon, A. (2019). Self-Supervised 3D Keypoint Learning for Ego-motion Estimation. arXiv preprint arXiv:1912.03426.
- [41] Zhang, Ji, and Sanjiv Singh. "LOAM: Lidar Odometry and Mapping in Real-time." *Robotics: Science and Systems*. Vol. 2. 2014.
- [42] Zhang, Ji, and Sanjiv Singh. "Visual-lidar odometry and mapping: Low-drift, robust, and fast." 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2015.
- [43] B. Delaunay, 'Sur la sphère vide. A la mémoire de Georges Voronoï, *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, 1934, no. 6, 793-800.