Universitatea Tehnica din Cluj-Napoca
Facultatea de Automatica si Calculatoare
Departamentul Calculatoare

PERCEPTIA MULTISPECTRALA A MEDIULUI PRIN FUZIUNEA DATELOR SENZORIALE 2D SI 3D DIN SPECTRUL VIZIBIL SI INFRA-ROSU

(MULTISPECTRAL ENVIRONMENT PERCEPTION BY FUSION OF 2D AND 3D SENSORIAL DATA FROM THE VISIBLE AND INFRARED SPECTRUM)

Cod proiect: PN-III-P4-ID-PCE-2016-0727
Contract nr: 60 / 12.07.2017

Etapa 2 / 2018

Director proiect:

Prof. dr. ing. Sergiu Nedevschi

Colectiv:

Conf. dr. ing. Tiberiu Marita
Conf. dr.ing. Florin Oniga
Sef lucrari dr. ing. Raluca Brehar
Asistent dr. ing. Robert Varga
Doctorand ing. Cristian Vancea
Doctorand ing. Arthur Costea
Doctorand ing. Mircea Muresan
Masterand ing. Horatiu Florea
Masterand ing. Zelia Blaga
Masterand ing. Selma Goga
Stud. Alexandru Kis

Cluj-Napoca, Decembrie 2018

Raport stiintific

privind implementarea proiectului in perioada: ianuarie – decembrie 2018

Titlul proiectului: "Perceptia multispectrala a mediului prin fuziunea datelor senzoriale 2D si 3D din spectrul vizibil si infra-rosu"

Denumire etapa 2: Dezvoltare de modele si metode originale pentru perceptia multispectrala si multisenzoriala a mediului

Activitatile etapei 2/2018:

A.2.1. Dezvoltarea de modele si metode originale pentru alinierea si reprezentarea 3D spatio-temporala si bazata pe aparente a datelor multi-sensoriale si multi-spectrale primare (SO2.1-2, SO2.2-2).

A.2.2. Dezvoltarea generatoarelor de regiuni de interes pentru obiecte (SO3.1-2)

A.2.3. Dezvoltarea mecanismului de clasificare redunant multi-senzorial si multi-spectral al obiectelor (SO3.3-2)

A.2.4. Dezvoltarea uneltei de adnotare si construirea seturilor de date pentru antrenare şi testare (SO3.2-2)

A.2.5. Proiectarea demonstratorului (SO4.1-1)

A.2.6. Diseminare rezultate (SO4.2-1)

Cuprins

2.1. Dezvoltarea de modele si metode originale pentru alinierea si reprezentarea 3D spatio-temporala si bazata pe aparente a datelor multi-sensoriale si multi-spectrale primare (SO2.1-2, SO2.2-2)

2.1.1. Sincronizarea spatio-temporala a datelor senzoriale

2.1.1. a. Sincronizarea datelor provenite de la senzori de imagine multispectrali

Obiectivul acestei proceduri este de a obţine imagini de la senzorul infraroşu (LWIR – Long Wavelength Iinfra-Red) sicronizate cu imaginile capturate de camerele video. Camera infraroşu funcţionează la o frecvenţă de achiziţie constantă fără a oferi facilităţi de captură la cerere. Spre deosebire, sistemul stereo de camere video oferă imagini la cerere. Acestea sunt sincronizate având in vedere faptul că cererea se realizează întotdeauna simultan. Diferenţele înregistrate între cele două camera video sunt nesemnificative.

În scopul realizării obiectivului propus s-a definit un model temporal de achiziţie pentru camere care nu oferă captură la cerere. Acest model defineşte corespondenţa dintre momentul capturii imaginii, momentul realizării cererii de către utilizator şi momentul în care imaginea este pusă la dispoziţie. Aceste detalii sunt prezentate în Fig. 2.1.1.

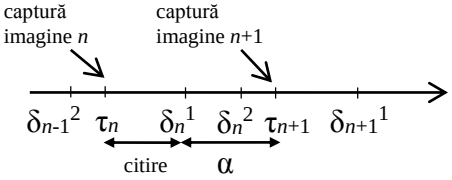


Fig. 2.1.1. Modelul propus pentru camera fără captură la cerere

Modelul propus are următoarea particularitate. Considerând captura la momentul τ_n pentru imaginea a *n*-a, orice cerere in intervalul orar $[\delta_{n-1}^2, \delta_n^1]$ va fi întârziată până la momentul δ_n^1, iar orice cerere în intervalul de timp următor $[\delta_n^1, \delta_n^2]$ nu va suferi întârzieri, răspunsul fiind imediat. Modelul defineşte umătoarea constrângere pentru aceşti parametri:

$$|\tau_{n+1} - \delta_n^2| = |\tau_n - \delta_{n-1}^2| \quad (2.1.1)$$

Deasemenea intervalul de citire din senzor are proprietatea de a fi constant de la captură la captură pe baza următoarei relaţii:

$$\delta_n^1 - \tau_n = \delta_{n+1}^1 - \tau_{n+1}\ , \quad (2.1.2)$$

La modelul propus se asociază şi următoarele proprietăţi de cronologie:

$$\tau_n < \delta_n^1 < \tau_{n+1} \quad (2.1.3)$$

$$\delta_n^1 \leq \delta_n^2 < \delta_{n+1}^1 \quad (2.1.4)$$

Conform modelului perioada capturii unei imagini are următoarea definiţie:

$$\tau = \tau_{n+1} - \tau_n \quad (2.1.5)$$

Considerând această perioadă τ constantă se defineşte *intervalul de relaxare* α ca fiind perioada dintre momentul punerii la dispoziţie a unei imagini şi momentul achiziţiei următoarei imagini:

$$\alpha = \tau_{n+1} - \delta_n^1 \quad (2.1.6)$$

În ceea ce priveşte camerele cu captură la cerere modelul utilizat este prezentat in Fig. 2.1.2:

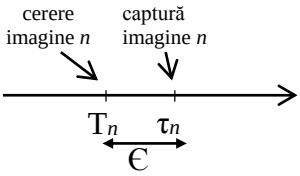


Fig. 2.1.2. Modelul uzual pentru camera cu captură la cerere

Cunoscând intervalul de timp de expunere Є s-a putut determina experimental perioada de relaxare α asociată camerei infraroşu. În acest scop a fost necesară sincronizarea firului de execuţie al aplicaţiei cu momentul δ_n^1. Conform modelului propus pentru camera fără captură la cerere (Fig. 2.1.1) o cerere pentru imagine va fi amânată până la momentul δ_n^1 sau nu va fi amânată defel dacă ea are loc în intervalul dintre δ_n^1 şi δ_n^2. Ca urmare a perioadei de incertitudine dintre momentele δ_n^1 şi δ_n^2 s-au iniţiat două cereri consecutive în loc de una. Cea de-a doua a avut loc imediat după răspunsul primit la prima cerere, fără întârziere. Conform (4) răspunsul la prima cerere va fi intre momentele δ_n^1 şi δ_n^2, iar la cea de-a doua cerere va fi cu siguranţă la momentul δ_{n+1}^1. Astfel după două cereri consecutive sincronizarea s-a realizat incepând cu ce-a de-a treia prin intârzierea graduală a momentului cererii către camerele video cu un interval de timp controlat Δ specificat in Fig. 2.1.3.

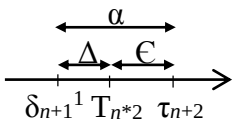


Fig. 2.1.3. Suprapunerea modelelor de achiziţie cu captură la cerere şi fără captură la cerere

Mediul de experimentare utilizează ambele tipuri de camera, video şi infraroşu. Acestea achiziţionează imagini ale unui obiect care emană căldură în timp ce se mişcă cu viteză. Intervalul de timp de întârziere a fost mărit treptat până s-a obţinut o poţitionare similară a obiectului de control atât in imaginile video cât şi în cele infraroşu. După stabilizarea valorii parametrului Δ, cunoscând dinainte valoarea expunerii Є pentru care s-a efectuat experimentul s-a putut calcula constanta α:

$$\alpha = \Delta + Є \tag{2.1.7}$$

Din momentul estimarii perioadei de relaxare α se poate adopta o valoare variabilă $Є_n$ pentru perioada de expunere, care să se modifice la fiecare imagine capturată. În functie de aceasta se va calcula şi o perioadă de întârziere corespunzătoare astfel încât α să se păstreze constant:

$$\Delta_n = \alpha - Є_n \tag{2.1.8}$$

Rezultate experimentale

Sistemul folosit este compus din camera infraroşu FLIR PathfindIR şi două camere video model JAI sincronizate. Camera PathfindIR are distanţa focală de 19 mm şi achiziţionează imagini de rezoluţie 320x240 la frecvenţa de 25Hz. Configuraţia experimentală utilizată pentru a detecta perioada de relaxare α a camerei PathfindIR conţine un ventilator având amplasată pe una din aripi o bandă colorată suprapusă cu un circuit electric ce emană căldură. In acest fel poziţia aripii este clar remarcată atat in imaginile video cât şi in cele infraroşu. Timpul de expunere pentru camerele video a fost fixat, iar ventilatorul a fost rulat la diferite viteze de rotaţie. S-au capturat mai multe imagini pe măsură ce intervalul de intârziere era crescut progresiv. S-au eliminat imaginile pentru care nu s-a obţinut sincronizare. Pe baza imaginilor sincronizate s-a obţinut o perioadă de relaxare de 23ms. Valori similare s-au obţinut şi la repetarea experimentului cu timpi de expunere diferiţi.

2.1.1. b. Sincronizarea achizitiei punctelor 3D furnizate de LiDAR cu imaginile din spectrul Visibil (VIS) si compensarea miscarii vehiculului propriu (motion compensetion)

Sincronizarea punctelor 3D oferite de LiDAR cu imaginile furnizate este importanta intrucat prin contopirea datelor obtinute de la cei doi senzori se poate obtine mai multa informatie legata mediul inconjurator. Sincronizarea punctelor in cadrul proiectului MULTISPECT se face pe baza de timestamp. Un software numit NetTime ajuta la setarea si sincronizarea timpului intre diferite disopzitive. Cu toate ca se trimite un triger sincronizat la camera si LiDAR cand sa vina informatia, avem un delay cauzat de timpul diferit de achizitie la cele doua dispositive. Pentru a avea informatia cea mai corecta se pastreaza time stampul imaginii ca si referinta si se stocheaza frameuri de la LiDAR. Frameul cu diferenta cea mai mica de time stamp din cele achizitionate se considera ca fiind cadrul corespunzator imaginii achizitionate.

Senzorii LiDAR din generația actuală folosesc un număr redus de transcievere laser (pentru a reduce costul) ale căror capabilități sunt augmentate prin montarea acestora pe capuri mobile de scanare, fapt ce permite efectuarea de măsurători cu un câmp de vedere larg – 360°. Datorită acestui design, o scanare completă – ce conține date din întregul câmp de vedere al senzorului – este achiziționată continuu, de-a lungul unui timp de achiziție t_{scan} (e.g. 100*ms*), În cazul în care aparatul de măsură este montat pe o platformă mobilă, acest mod de operare conduce la apariția unor distorsiuni în datele 3D achiziționate (unele puncte apar mult mai aproape decat ar trebui), datorită mișcării sistemului de coordonate al achiziției. Un astfel de exemplu poate fi vazut in de mai jos.

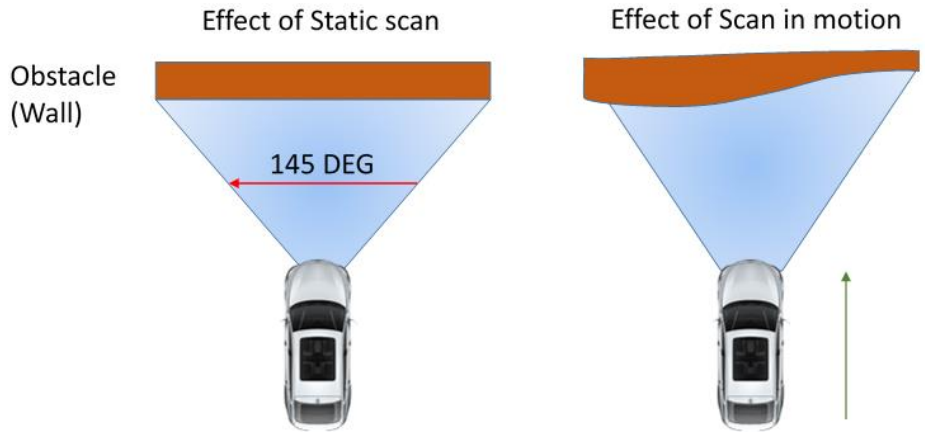


Fig. 2.1.4. Efectului de decompemsare a masuratorii LIDAR-ului datorita miscarii

In continuare se prezinta metodele folosite pentru compensarea masuratorii cu miscarea autevewhicului propriu pentru cei 2 senzori LiDAR (un LiDAR cu 4 canale model Sick AG si un LiDAR cu 16 canale de tip Velodyne care vor fi folositi pe montajul experimental al demoinstratorului.

Compensarea miscarii si integrarea temporala pentru LiDAR-ul cu 4/8 canale

LiDARul cu 4/8 raze este un instrument de masurare care are posibilitatea de a detecta obiectele din scena si distantele la acestea. Aparatul scaneaza mediul folosind mai multe raze laser si primeste ecourile de la fiecare folosind trei fotodiode. Distantele pana la fiecare punct scanat sunt calculate flosind tehnica time of flight [Gal2015] si datele sunt transmise pe interfata Ethernet. Senzorul scaneaza de la stanga la dreapta facand o rotatie de la un unghi de 0 la 145 de grade. Frecventa de achizitie a LiDAR-ului cu 4 canale este de 12.5 Hz, iar datele de la sensor la un moment de timp provin de la 3 straturi, urmand ca senzorul sa faca o baleere in jos dupa perioada de intoarcere si sa continue achizitia de la urmaotarele 3 straturi. Asadar doua dintre straturile senzorului vor avea o frecventa de 25 de Hz fiindca vor proven de la 2 achizitii consecutive. O imagine intuitiva a procesului este ilustrata in figura de mai jos. Cu 1 este reprezentata distanta de 0.8 grade pe verticala intre straturi, cu 2 frecventa de achizitie a unui strat iar cu cifra 3 se indica senzorul laser.

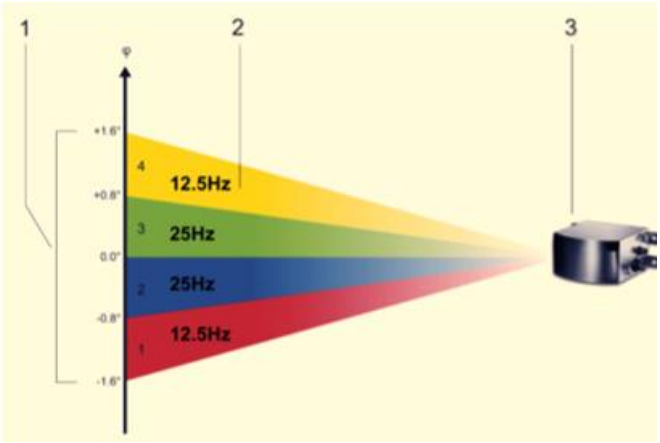


Fig. 2.1.5. Campul vizual vertical vs. frecventa de scanare pentru LiDAR-ul cu 4 canale.

Pentru a corecta informatia de laser, pentru fiecare punct se calculeaza timpul la care a fost acaparat folosind eticheta de timp initiala pe care o ofera senzorul. Expresia de calcul este ilustrata in ecuatia 1 de mai jos.

$$Tp = TI + x * chID + DT \tag{2.1.9}$$

In ecuatia de mai sus Tp reprezinta timpul unui punct oarecare, TI reprezinta timpul initial, x reprezinta durata de scanare a unui punct, chID este id-ul canalului de achizitie si DT reprezinta timpul de intarziere la intoarcere (care e de 16 ms). Stiind timpul fiecarui punct putem alege un moment de timp la care sa fie facuta corectia. Astfel pentru fiecare punct se calculeaza o valoare delta intre timpul punctului si timpul la care dorim sa facem corectia. Se calculeaza o diferenta de distanta folosind viteza vehicolului primita pe magistrala CAN si diferenta de timp. Se modifica valorile punctelor folosind distanta obtinuta precedent pentru fiecare punct.

Pentru a imbunatatii densitatea detectiei, se fuzioneaza 8 cadre, reprezentand 4 scanari de 4 straturi. Un exemplu de fuziune temporala este ilustrata in figura de mai jos.

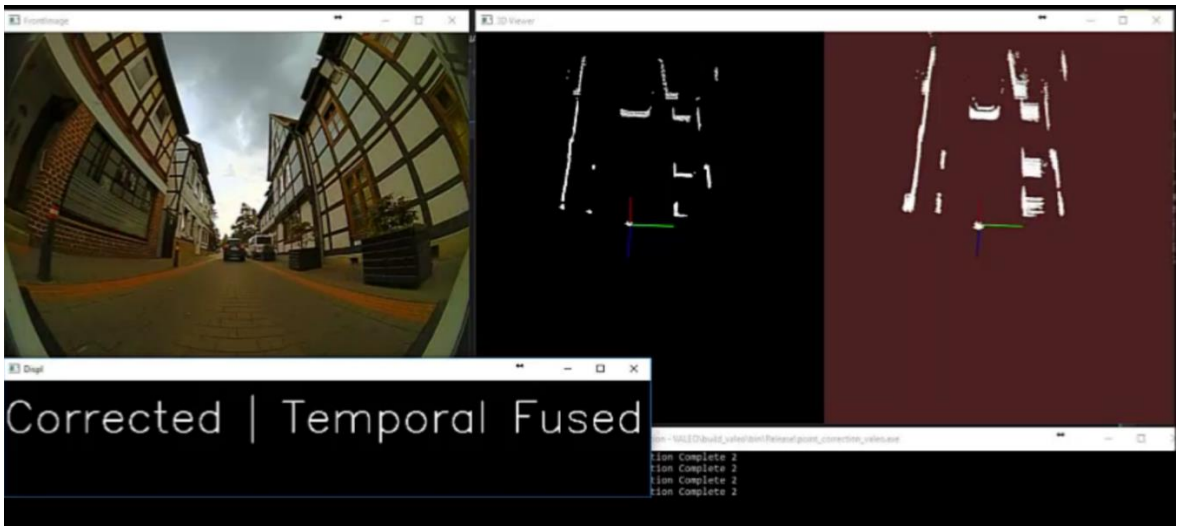


Fig.2.1.6. Exemplu de fuziune a 8 cadre, reprezentand 4 scanari de 4 straturi

Compensarea miscarii si integrarea temporala pentru LiDAR-ul cu 16 canale

Pentru a elimina acest efect din scanările senzorilor Velodyne, se propune un algoritm de compensare al mișcării/aliniere temporală, extinzând [Hon2010]: un senzor specializat montat pe mașină furnizează date privind ego-mișcarea (pe 6 grade de libertate) a vehicolului propriu, ce mai apoi sunt agregate pentru a calcula transformarea de mișcare efectuată de-a lungul achiziției unei scanări complete. Folosind inversa acestei transformări, precum si timpul exact de măsurare al

fiecarui punct 3D din scanare, se poate calcula o transformare corectivă pentru ficare punct individual, eliminând distorsiunile create de mișcarea proprie.

Mai mult, algoritmul poate fi adaptat pentru a realiza această aliniere temporală la un moment arbitrar de timp, precum cel al achiziției imaginilor color/infraroșu. În acest mod, se poate obține o aliniere a datelor 3D – 2D de mai bună calitate, fără necesitatea implementări la nivel hardware al unui sistem de sincronizare între aceste dispozitive.

Concret, transformarea C_i ce trebuie aplicată punctului x_i poate fi calculată folosind următoarea formulă matematică:

$$C_i = Tego^{-\frac{\Delta_i}{\Delta_1}} = exp\left(-\frac{\Delta_i}{\Delta_1} \cdot \log(T_{ego}\right) \tag{2.1.9}$$

Unde: T_{ego} reprezintă transformarea rigidă a vehicolului măsurată de senzori sub formă matricială 4x4, iar Δ_i reprezintă diferența dintre timpul arbitrar de corecție și timpul de măsurare al punctului *i* (Δ_1 făcând referire la diferența asociată primului punct măsurat). Se poate observa aplicarea funcțiilor logaritm și exponențială pe termeni matriciali, operație costisitoare computațional, luând în considerare necesitatea aplicării acesteia pentru fiecare punct individual. Din acest motiv, la nivelul de implementare sunt luate în considerare mai multe metode de optimizare.

2.1.2. Alinierea spatiala a datelor senzoriale

2.1.2.a. Proiectia punctelor 3D-Stereo in spatiul imagine

Pentru o camera de luat vederi ai carei parametrii intrinseci si extrinseci sunt cunoscuti fata de un sistem de coordonate de referinta (ex. sistemul de coordonate al vehiculului propriu), proiectia punctelor 3D observate cu aceasta camera pe planul imagine se poate face cu ajutorul ecuatiei de proiectie [Ned2012].

Pentru punctele 3D furnizate de senzorul de steroviziune nu mai este necesara proiectia punctelor, deorece punctele 3D sunt reconstruite prin stereocorelatia dintre perechi de puncte 2D corespondente din imaginea stanga si dreapta a sistemului stereo.

Pentru proiectia punctelor 3D furnizate de sistemul de steroviziune pe planul imagine al cameri IR (FIR/LWIR) se va face pe baza matricii de proiectie calculata pentru parametrii intrinseci si extrinseci estimati pentru aceasta camera [Ned2017],

$$\mathbf{P}_F = \mathbf{A}_F \cdot \left[\mathbf{R}_W^F \mid \mathbf{T}_W^F\right] \tag{2.1.10}$$

$$s \cdot \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} x_S \\ y_S \\ z_S \end{bmatrix} = \mathbf{P}_F \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{2.1.11}$$

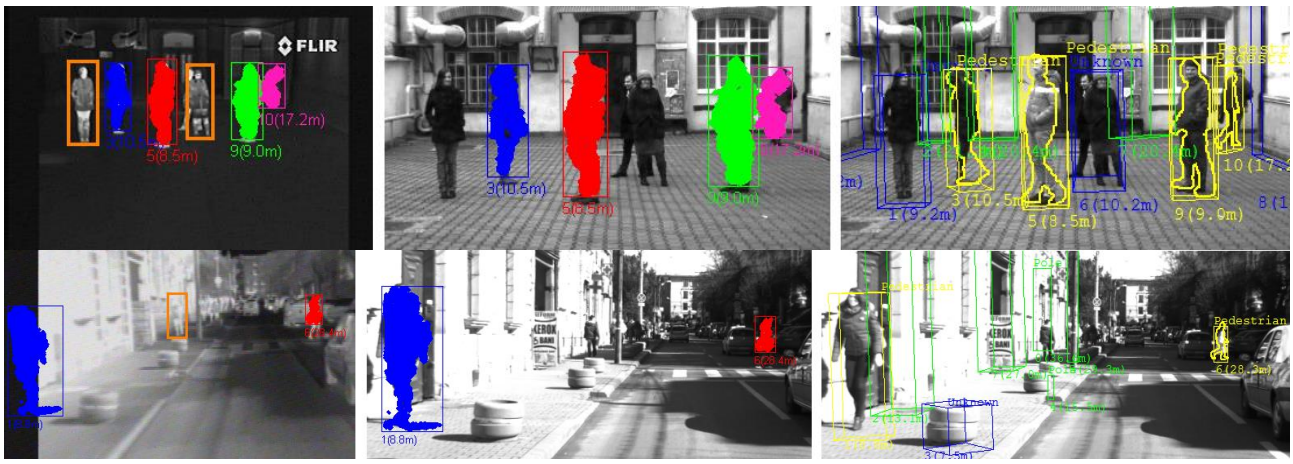
unde: $\mathbf{P}_i(X_i, Y_i, Z_i)$ – coordonatele punctului 3D exprimate in sistemul de coordonate al lumii (vehiculului propriu)

$\mathbf{p_i}[u_i, v_i]$ – coordonatele punctului 3D proiectat pe imaginea IR

$\mathbf{A}_F$ – matricea interna a camerei (include parametrii intrinseci)

$\mathbf{R}_W^F$ – matricea de rotatie dintre sistemul de coordonate al lumii si camera

$\mathbf{R}_W^F$ – vectorul de translatie dintre sistemul de coordonate al lumii si camera



a. b. . c.

Fig. 2.1.7. Ilustrarea procesului de proiectie al punctelor 3D asociate pietonilor detectati de senzorul stereo (c) pe imaginea stanga a senzorului (b) respectiv pe imaginea FIR (a)

2.1.2.b. Proiectia punctelor 3D-LiDAR in spatiul imagine

Fuziunea informaţiei senzoriale de la LiDAR şi camere se poate obţine prin proiecţia punctelor LiDAR în imagine. Pentru aceasta sunt necesari parametrii senzorilor individuali şi transformarea rigidă între ele. Aceşti parametri, numiţi intrinseci şi extrinseci, rezultă din procesul de calibrare.

Fie $\bar{\boldsymbol{X}}_L$ un punct 3D de la un senzor LiDAR dat în sistemul nativ al LiDAR-ului în coordonate omogene 3D:

$$\bar{\boldsymbol{X}}_L = [X \;\; Y \;\; Z \;\; 1]^T \tag{2.1.12}$$

Primul pas este transformarea punctului în sistemul camerei care se poate realiza folosind:

$$\bar{\boldsymbol{X}}_{cam} = T_{cam}^{L} \bar{\boldsymbol{X}}_L \tag{2.1.13}$$

unde: T_{cam}^{L} este matricea de transformare 4x4 în spaţiul omogen 3D şi conţine o matrice de rotaţie şi un vector de translaţie.

$$T_{cam}^{L} = \begin{bmatrix} R_{cam}^{L} & \boldsymbol{t}_{cam}^{L} \\ \boldsymbol{0} & 1 \end{bmatrix} \tag{2.1.14}$$

Având punctul în sistemul de coordonate al camerei trebuie să-l proiectăm în spaţiul imagine. Se vor ignora punctele care sunt în spatele camerei deci au componenta z negativă. Se face distincţia două tipuri principale de proiecţii: se poate proiecta în imaginea originala distorsionată sau în imaginea nedistorsionată.

Dacă se doreşte proiecţia în imaginea originală se va folosi funcţia de proiecţie care rezultă din modelul camerei şi procesul de calibrare. Pentru lentile de tip fisheye se poate utiliza modelul definit de [Gey2000]. Se aplică paşii următori:

- Se consideră punctul în coordonate 3D normale: $\boldsymbol{X}_{cam} = [X \quad Y \quad Z]^T$ (2.1.15)
- Se calculează distanţa de la cameră: $\rho = \sqrt{X^2 + Y^2 + Z^2}$ (2.1.16)
- Se proiectează pe un plan de imagine normalizat (ξ este parametru de calibrare):

$$\boldsymbol{X} = \left[\frac{X}{Z+\xi\rho} \quad \frac{Y}{Z+\xi\rho} \quad 1\right]^T = [x \quad y \quad 1]^T \tag{2.1.17}$$

- Se aplică distorsiunile tangenţiale şi radiale (k_1, k_2, p_1, p_2 sunt parametri de calibrare):

$$x_d = x(1 + k_1 r^2 + k_2 r^4) + 2p_1 xy + p_2(r^2 + 2x^2) \tag{2.1.18}$$

$$y_d = y(1 + k_1 r^2 + k_2 r^4) + 2p_2 xy + p_1(r^2 + 2y^2) \tag{2.1.19}$$

- Se înmulţeşte cu matricea internă K (f_u, f_v, c_u, c_v sunt parametri de calibrare):

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} \tag{2.1.20}$$

- Rezultă coodonatele în spaţiul imagine. Funcţia care mapează punctul original din spaţiul 3D omogen în spaţiul imagine se notează cu:

$$P(\bar{\boldsymbol{X}}_{cam}) = [u\ v\ 1]^T \tag{2.1.21}$$

Dacă se doreşte proiectarea în imaginea nedistorionată se aplică o altă funcţie de proiecţie care returnează intersecţia razei din centrul optic al camerei cu planul virtual folosit pentru undistorsionare. În acest caz avem:

$$[u\ v\ 1]^T = K_{undist} \cdot nonh(\bar{\boldsymbol{X}}_{cam}) \tag{2.1.22}$$

unde: funcţia *nonh* transformă din spaţiul omogen în cel neomogen şi K_{undist} este matricea internă pentru imaginea nedistorsionată şi are în mod tipic următoarea formă dată de rezoluţia orizontală w şi verticală h şi de factorii de scalare t_h şi t_v corespunzători.

$$K_{undist} = \begin{bmatrix} w/(2\ t_h) & 0 & w/2 \\ 0 & h/(2\ t_v) & h/2 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.1.23}$$

Odată ce avem coordonatele imagine $[u\ v]$ pentru un punct LiDAR, putem asocia culoarea sau informaţia semantică de la poziţia respectivă cu punctul original.

Este important de menţionat că senzorii văd mediul din unghiuri diferite. Din această cauză este posibil ca un obiect vizibil de către LiDAR să fie ocluzionat în vederea cameră. În acest caz asocierea făcută va fi greşită şi trebuie corectată. O modalitate ar fi prin estimarea suprafeţei date de obiectele cele mai apropiate de cameră şi eliminarea punctelor care sunt în spatele acestei suprafeţe.

Pentru a densifica numarul de puncte se fuzioneaza norul de puncte de la senzorul de 16 straturi si cel de la senzorul cu 4/8 straturi. Astfel se obtine un nor de puncte care are informatia complementara de la cei doi senzori. LiDARul cu 4/8 straturi are o distanta de lucru mai mare ca cel de 16 straturi. Cu toate acestea acest LiDAR ofera mai putine puncte 3D fata de LiDARul de 16 straturi:

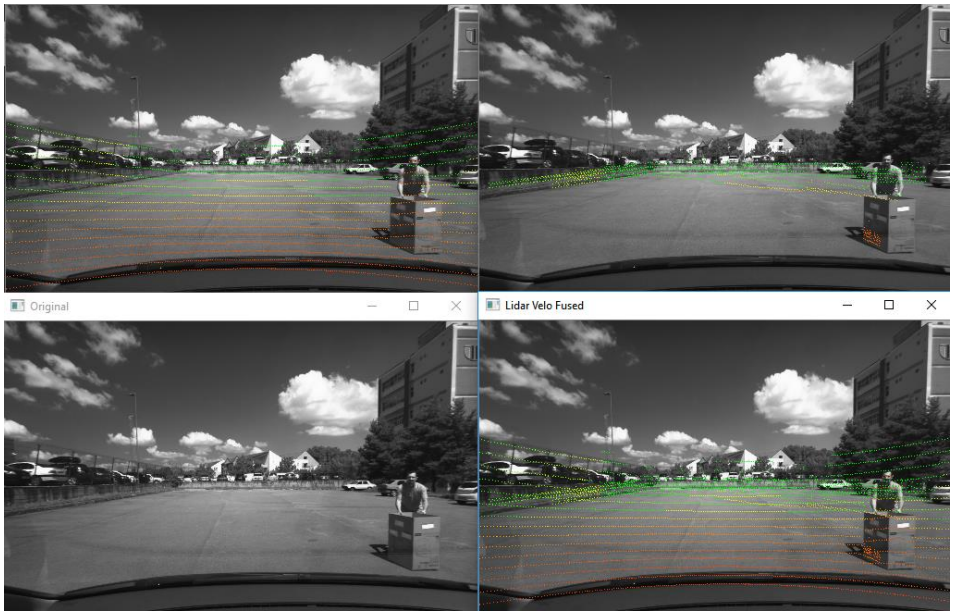


Fig. 2.1.8. Rezultatul proiectiei punctelor 3D furnizate de senzorii LiDAR in imaginea stanga a senzorului stereo: (ss) punctele proiectate de la senzorul Velodyne; (ds) punctele proiectate de la LiDAR-ul cu 4/8 straturi; (dj) proiectia fuzionata a punctelor 3D de la cei doi senzori LiDAR

2.1.2.c. Alinierea spatiala dintre imaginile din spectrul vizibil (VIS) si infrarosu (LWIR)

Alinierea intre imaginile din spectrul vizibil (VIS) si cel in infrarosu indepartat (FIR/termal), sau inregistarea acestora (registration) se poate face prin mai mule abordari, in functie de scenariul de lucru. Daca este disponibila informatia de adancime/disparitate furnizata de senzorul de stereoviziune atunci se pota folosi proectia punctelor 3D pe imaginea FIR pe baza matricii de proiectie (vezi cap. 2.1.2.a) sau abordari echivalent, cum ar fi transferul epipolar sau modelul tensorului-trifocal [Har2003].

Dificultatea problei de inregistrare apare atunci cand lispseste informatia de adancime/disparitate de la senzorul de stereoviziune din diverse motive (iluminare prea puternica (saturatie) sau prea slaba (umbrire puternica, iluminare slaba) a scenei sau a anumitor zone din scena. Problema devine in acset caz destul de dificila de rezolvat din cauza aparentei total diferite dintre intensitatile celor doua tipuri de imagini (VIS vs. FIR). In urma studiului bibliografic si experimentelor efectuate, au fost descoperiti numerosi descriptori si metrici care pot fi aplicati, cu mai mult sau mai put,in succes, ıin procesul de innregistrare a imaginilor multimodale. Dintre acestea, metricile MI (Mutual Information) si LSS (Local Self Similarity) [Tor2011],[Bil2014] ies in evidenta, atat prin popularitatea lor, cat si prin studii comparative realizate (fig. 2.1.9).

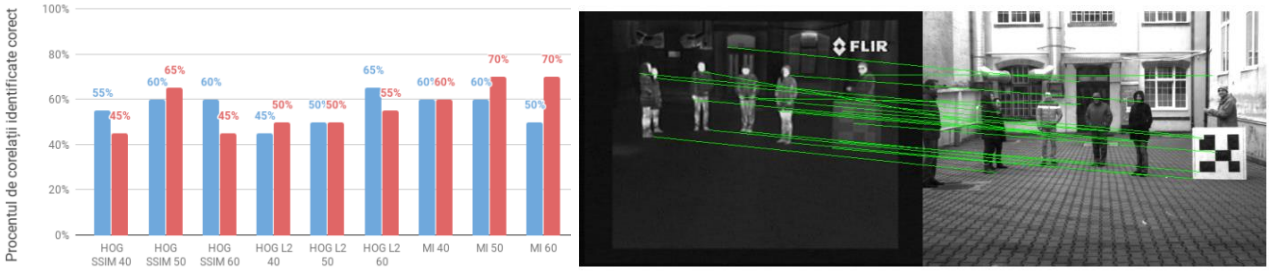


Fig. 2.1.9. (s) Compartia diverselor metrici de corelatie intre punctele din imaginile FIR si VIS; (d) Exemple de corespondente identificate folosind metrica MI.

Abordarea propusa se bazeaza pe folosirea unei astfel demetrici in gasirea corespondentelor dintre imaginile FIR si VIS, Pasi algoritmului de inregistrare propusi sint urmatorii:

1. Se identifica puncte de interes in imaginea FIR folosind un detector consacrat (colturi)
2. Pentru fiecare punct de interes al imaginii se cauta corelatia corespunzatoare de-a lungul dreptei epipolare dedusa din pozitia si oreintarea relativa a celor doua camere (parametrii de calibrare se considera cunoscuti.
3. Se va parcurge linia apipolara cu o ferestra glisanta si se identifica pozitia pentru care metrica de corlatie dintre fereastra asociata punctul de interes din imaginea FIR si fereastra glisanta este maxima
4. Se foloseste metida RANSAC pentru filtrarea corespondentele fals-pozitive si pentru estimarea maparii dintre cele regiunile celor 2 imagini.

Functia de mapare este o transfirmare de persepctiva care se poate aplica unor suprafete plane, si se poate calcula din corespondenta a cel putin 4 puncte (coplanare). In consecinta metoda se va putea aplica pe regiuni cvasi-planare (puncte pt. care varianta asnacimii fata de camera este neglijabila in raport cu adancimea medie pana la camera), cum ar fi poietinii, cu conditia gasirii unui numar suficient de corespondente.

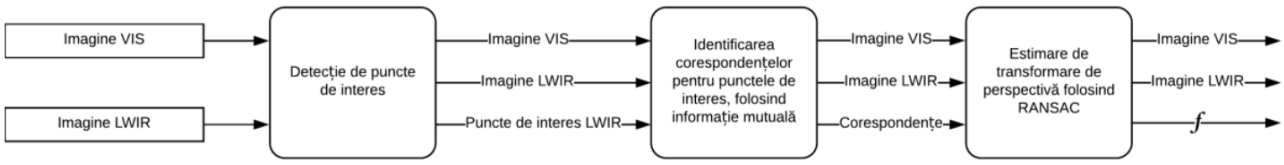


Fig.2.1.10. Diagrama metodei de alinierea a regiunilor dintre imaginii FIR si VIS.

2.1.3. Segmentarea canalelor senzoriale

2.1.3.a. Segmentarea semantica a imaginilor

Segmentarea semantica este sarcina de etichetare a imaginilor cu clase semantice la nivel de pixel. Ofera o reprezentare la nivel inalt si poate avea un rol important in perceptia si intelegerea semantica a mediului. Obiectivul principal este obtinerea unei solutii de acuratete si precizie ridicata la un cost redus de calcul, care sa permita rularea in timp real. Acesta poate fi realizat folosind trasaturi si scheme de clasificare robuste si eficiente din punct de vedere al timpului de executie.

Este importanta exploatarea informatiei multisenzoriale pentru clasificare din color, grayscale, infrarosu sau profunzime din stereoreconstructie sau LiDAR. Propunem extragerea trasaturilor in forma de canale normalizate: canale de trasaturi de baza multimodale filtrate si canale multimodale de context. Aceste canale de trasaturi sunt folosite pentru a esantiona trasaturi de clasificare pentru arbori de decizie invatate prin metoda AdaBoost. Clasificatoarele, care pot fi aplicate la nivel de pixel sau superpixel, permit generarea unor harti de costuri pentru fiecare clasa semantica, care la urma permite generarea segmentarii semantice a imaginii de intrare. Fig. 2.1.11 furnizeaza o vedere de ansamblu a solutiei propuse pentru segmentarea semantica.

In literatura, de obicei, segmentarea semantica foloseste ca si intrare imaginea color. Culoarea poate reprezenta o informatie aditionale fata de imaginea grayscale, dar poate fi si o sursa pentru overfitting. De aceea propunem folosirea si a imaginii grayscale ca si intrare pentru a genera trasaturi de intensitate si gradienti, pentru a avea trasaturi si muchii invariante la culoare. Lipsa informatiei de culoare este compensata prin modalitatiile aditionale de profunzime din stereoreconstructie sau din LiDAR.

Canalele de trasaturi de baza vor capta intensitatea, magnitudinea gradientului si orientarea gradientului din fiecare modalitate. Pentru a imbogatii aceste trasaturi, propunem filtrarea in multirezolutie a canalelor de baza, rezultand in canale filtrate. Pentru filtrarea cananlelor, folosim schema de filtrare ilustrata in Fig. 2.1.11. Canalele de intensitate (grayscale, color LUV, infrarosu, disparitate sau LiDAR interpolat) sunt filtrare de mai multe ori iterativ cu un filtru de mediere de 3 x 3, rezultand in canale multiscala. Canalele rezultate sunt filtrare mai departe cu filtre de tip trece sus, derivate verticale si orizontale. Aceste filtre permit captarea gradientilor la orice orientare si orice scala. Datorita nucleelor de convolutie foarte simple din schema de clasificare, timpul de executie pentru filtrarea canalelor de baza este de doar 1-2 milisecunde folosind o implementare GPU.

Pietonii si obiectele din mediile de trafic sunt limitate de constrangeri spatiale si geometrice. Definim astfel de constrangeri si le incorporam in solutii de perceptie sub forma unor canale aditionale de context pe langa canalele de trasaturi filtrate. In acest fel putem permite clasificatorilor sa invete contextul pentru diferite tipuri de obiecte sau clase semantice. Trasaturile de context pot, de asemenea, accelera procesul de predictie. In cazul arborilor de decizie in cascada-soft (soft-cascade), predictia este oprita daca scorul intermediar de predictie scade sub un anumit prag. Trasaturile de context pot opri predictia intr-un stadiu incipient daca contextul nu este adecvat pentru o clasa semantica. In experimentele noastre pentru segmentarea semantica am folosit urmatoarele contexte:

- 2D spațial: poziția verticală și orizontală in imagine
- 3D spațial: poziție 3D x, y, z, reprezentare densa interpolata din LiDAR sau stereo
- 3D dimensiune: înălțimea și lățimea în 3D a grupărilor de superpixeli

Pentru a putea genera segmentari semantice, antrenam un clasificator binar individual de pixeli pentru fiecare clasa semantica. Propunem utilizarea unor trasaturi de clasificare multi-raza, extrase din canalele de trasaturi. AdaBoost este folosit pentru selectarea trasaturilor si pentru invatarea unor clasificatoare bazate pe boosting.

Clasificam pixelii individuali utilizand valorile pixelilor inconjuratori din canalele de trasaturi. Punctele mai apropiate capteaza structura locala, in timp ce cele indepartate capteaza contextul. In scopul de a capta trasaturile la diferite raze, le extragem cu o grila folosind diferite rate

de esantionare. Utilizam o retea de 13 x 13 in jurul pixelului de interes si aplicam esantionarea utilizand o rata de pas de unu, doi, patru si opt pixeli, rezultand in 676 de puncte din fiecare canal individual. Patru grile cu diferite rate de esantionare acopera o regiune de 13 x 13, pana la 104 x 104 pixeli in imaginea de intrare. Pentru a obtine clasificatoare puternice, se antrenarea 2048 arbori de decizie de 7 nivele pentru fiecare clasa semantica. Dupa antrenarea clasificatoarelor putem genera segmentari semantice. Segmentarea finala este rafinata utilizand un CRF dens la nivel de superpixel.

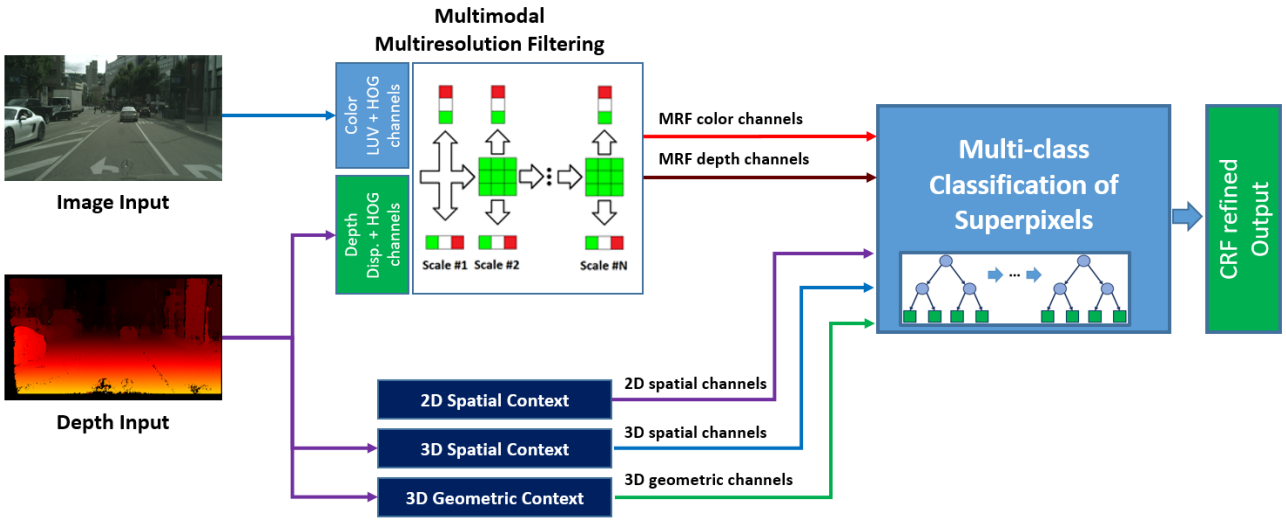


Fig. 2.1.11. Segmentarea semantica folosind canale de trasaturi multimodale filtrate si canale multimodale de context, arbori de decizie invatate prin boosting si rafinare CRF la nivel de superpixel.

Avand in vedere progresele recente in domeniul invatarii profunde abordarile bazate pe retele neuronale au inceput sa domine literatura si in cazul segmentarii semantice. S-au obtinut rezultate semnificative folosind retele neuronale convolutionale profunde. Aceste retele permit invatarea trasaturilor de imagine intr-un mode ierarhic, iar la ultimele nivele se genereaza predicitia finala. Aceeasi retea de baza poate fi folosita pentru mai multe scopuri, cum ar fi clasificare, detectie de obiecte, segmentarea semantica si segmentarea in instante.

Segmentarea in instante poate avea un rol important in interpretarea scenelor. Segmentarea in instante detecteaza fieacre instante a unei clase semantice, furnizand un dreptunghi de incadrare (ca si in cazul detectiei de obiecte) si o masca de instanta care eticheteaza fieacre pixel individual care apartine de obiect. Segmentarea in instante delimiteaza obiectele de fundal si de alte instante de obiecte de aceeasi clasa. Astfel, in cazul proiectarii punctelor 3D in imagine, putem stii exact fiecare punct de ce obiect apartine si putem face rationamente asupra punctelor 3D care apartin de obiecte individuale. In cazul detectiei cu dreptunghiuri de incadrare 2D acest lucru este mai dificil, deoarece pot fi suprapuneri intre dreptunghiuri.

In urmatoarele propunem o arhitectura comuna bazata pe o retea neuronala profunda pentru segmentarea semantica si pentru segmentarea in instante. Ca si arhitectura de baza pornim de la solutia Mask R-CNN [He2017], o arhitectura bazata pe o retea piramidala de trasaturi (Feature Pyramid Network - FPN) [Lin2016] si ResNet [He2016] pentru detectie de obiecte si segmentare in instante. Segmentarea semantica este obtinuta prin extinderea acestei arhitecturi cu un head de segmentare. Aceasta extensie imbina trasaturile multiscala ale piramidei de trasaturi, fiind o varianta mai eficienta din punct de vedere al costului de calcul in compartie cu arhitecturile bazate pe convolutii dilatate, care necesita mai multa memorie.

Iesirea FPN-ului consta in 256 de canale de trasaturi la 4 scale (1/4, 1/8, 1/16 si 1/32). Aceste canale de trasaturi sunt folosite ca si intrate pentru headuri de predictie pentru detectie de obiecte:

clasificare si regresie de dreptunghi de incadrare. Pentru o utilizare eficienta a trasaturilor, refolosim trasaturile din FPN si pentru segmentarea semantica, extinzand arhitectura cu un head de predictie pentru segmentare. Arhictectura este ilustrata in Fig.2.1.12. Headul de segmentare utilieaza trasaturile din toate cele patru nivele ale FPN-ului. Aceste trasaturi sunt rafinate prin straturi aditionale de convolutie. Convolutiile dilatate (atrous) au un rol important in extragerea trasaturilor de context si de raza lunga. De aceea aplicam o piramida spatiala atrous (ASP) pentru headul de segmentare la nivelele 1/32 si 1/16, folosdin o convolutie 1 x 1 si trei convolutii dilatate 3 x 3 avand ratele de dilatare 6, 12 si 18. Hartile de trasaturi rezultate sunt comprimate la 128 de canale folosind convolutii 1 x 1 si sunt marite la o scala comuna de 1/4. Concatenarea lor rezulta in 512 canale la scala de 1/4. In final, aceste canale sunt folosite pentru a genera predictia finala costurile de clasificare pentru fiecare clasa, folosind convolutii 1 x 1. Costurile de clasificare la nivel de pixel permit generarea segmentarilor semantice. Arhitectura neuronala comuna permite detectia de obiecte, segmentare in instante si segmentarea semantica. O vedere de ansamblu poate fi vazut in Fig.2.1.12.

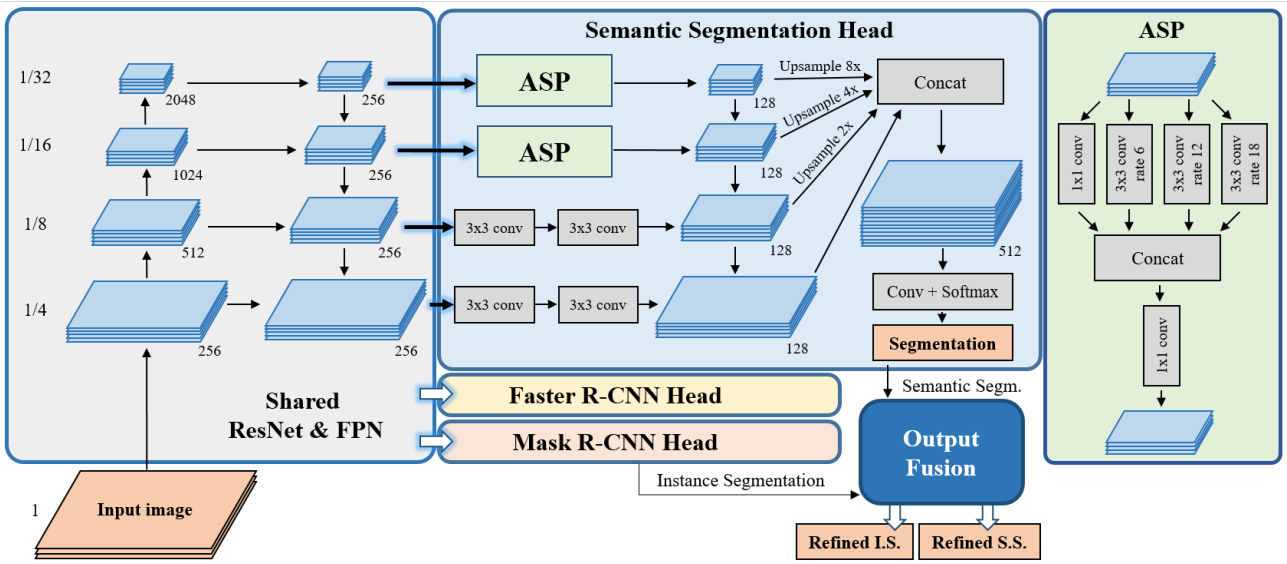


Fig. 2.1.12. Retea neuronala pentru detectie, segmentare in instante si segmentare semantica. Retea comuna pentru calcularea trasaturilor de baza folosind o retea convolutionala reziduala, ResNet si extensie cu piramida de trasaturi (FPN). Reteau este extinsa cu 3 retele de predictie: detectie (Faster R-CNN), segmentare in instante (Mask R-CNN) si segmentare semantica (retea propusa).

2.1.3.b. Segmentarea spatiului 3D

Setul de puncte 3D va fi segmentat in obiecte. Pe langa caracteristicile spatiale (pozitia 3D), punctele 3D vor avea proprietatile de clasa, respectiv instanta semantică.

In proiectarea algoritmului de segmentare a spatiului 3D se tine cont, la diferite niveluri, de aceste proprietati derivate din imagini. Prezenta informatiilor aditionale va ajuta algoritmul de segmentare (mai ales) in situatii limita unde conditia de conectivitate 3D este greu de luat doar pe baza trasaturilor 3D.

Segmentarea 3D se va face dupa pasii standard de (1) detectie a suprafetei drumului si (2) gruparea punctelor de obstacol pentru identificarea obstacolelor.

1. Pentru detectia suprafetei drumului, informatia semantica ajuta mult la reducerea spatiului de cautare. O abordare standard pentru detectarea suprafetei drumului ar implica determinarea zonelor din scena unde suprafata se supune anumitor constrangeri: panta locala sub un anumit prag, sau verificarea unui model global de suprafata (planar, polinomial, etc.). O mare parte din masuratorile 3D nu apartin drumului, dar ele intra in procesul de determinare a suprafetei in abordarea standard. Acest neajuns se elimina daca din setul de puncte se folosesc doar cele care

au clasa de drum. Astfel, modelul drumului se va estima direct din puncte care au o mare probabilitate sa fie drum. Se va reduce timpul de procesare si de asemenea va creste robustetea, deoarece scade probabilitatea ca puncte de obstacol sa fie incluse in procesul de detectie a drumului.

2. Dupa determinarea suprafetei drumului, punctele 3D sunt separate in puncte de drum, respectiv puncte de obstacol, in functie de pozitia lor relativ la suprafata. Punctele de drum intra intr-un proces de grupare, folosind o reprezentare intermediara sub forma de voxeli, in care proprietatea de adiacenta se va stabili nu doar pe baza distantei in spatiul 3D, ci si folosind proprietatile de clasa, respectiv instanta sematica. Proprietatea de clasa semantica va asigura o prima imbunatatire a gruparii: puncte apartinand de clase diferite nu se vor grupa in acelasi obstacol. Proprietatea de instanta va rezolva situatii dificile, cand obiecte din aceeasi clasa sunt situate foarte aproape in spatiul 3D si sunt greu de grupat individual. De exemplu un grup de pietoni, sau un sir de vehicule parcate aproape unul de celalalt, vor fi mai usor de segmentat 3D folosind proprietatea de instanta.

2.1.4. Generearea modelului de aliniere si reprezentare multi-senzorial si multi-spectral STMAR (Spatio-Temporal Multispectral Appearance-based Representation)

In urma alinierii spatio-temporale a datlor senzoriale primare si a segmentarii acestora se propune un model de fuziune (low-level) a acestor date, obtinandu-se o reprezentare originala denumita STMAR (Spatio-Temporal Multispectral Appearance-based Representation). Peste aceasta reprezentare s-au dezvoltat toate metodele de preceptie avansata a mediului dintre care amintim:

- detectia de pietoni si alti participanti vulnerabili la atrafic

- detectiae vehicule

- detectia structurilor/obiectlor statice din scena (stalpi, copaci, borduri, parapeti)

Entitatile detectate/clasificate de metodele de perceptie avansata sunt in final fuzionate cu informatia 3D segmentata in forma de obiecte generice pentru a infera mai departe o reprezentare completa a scenei in care fiecare obiect va putea avea asociat date de localizare, dimensiune, viteza/oientare si identitate (clasa).

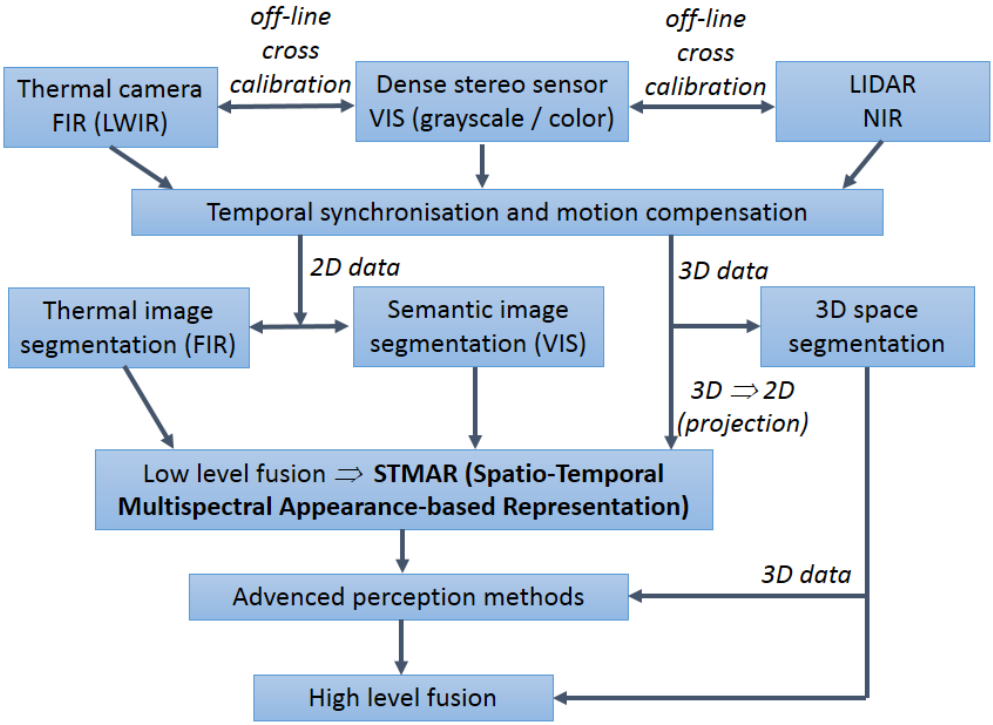


Fig. 2.1.14. Schema de perceptie avansata a mediului pe baza modelului propus: STMAR (Spatio-Temporal Multispectral Appearance-based Representation).

2.2. Dezvoltarea generatoarelor de regiuni de interes pentru obiecte (SO3.1-2)

Regiunile de interes pentru obiecte precum pietoni, maşini, stâlpi, alte obstacole au fost generate astfel: (1) prin obținerea obiectelor 3D din steroviziune proiectate apoi pe imaginile de intensitate și clasificate pe clase prin metode de Boosting și (2) utilizând mecanisme de analiză și clasificare a amprentei termale corespunzătoare pietonilor în imaginea FIR/LWIR.

Arhitectura propusă este descrisă în figura 2.2.1:

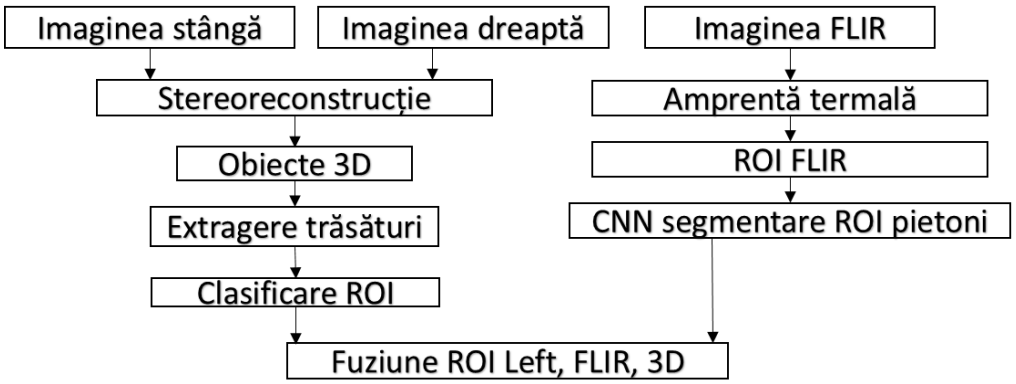


Fig.2.2.1. Arhitectura propusă pentru generatoarele de regiuni de interes

Sistemul de stereoviziune achiziționează imagini grayscale cu o rezoluție scalata la 512x383 pixeli. Stereo-reconstrucția este realizată utilizând un algorutm de tip semiglobal matching (SGM) implementat cu o implementare optimala pe procesorul grafic (GPU) cu ajutorul mediului CUDA (nVidia). Rezultatul este o hartă de adâncimi care completează informația 2D cu o hartă de puncta 3D. Harta de adâncimi codifică pentru fiecare punct distanța față de sistemul de achiziție stereo. Modulul de detecție foloseste informația 3D și grupează punctele aflate la distanțe asemanătoare în obiecte 3D. Este propus și un mecanism de urmărire a obstacolelor prin flux optic.

Pentru fiecare obiect 3D se calculează proiecția punctelor sale pe imaginea stângă și astfel se obțin regiuni de interes din imaginea grayscale corespunzătoare obiectului 3D. Pentru a determina ce tip de obiect corespunde fiecărei regiuni de interes se calculează trăsături precum HOG (Histograma orientării gradientului), LBP (Local Binary Features) și textoni. Diferiți algoritmi de învățare pot fi antrenați pe aceste trăsături pentru a genera o decriere semantică a fiecărei regiuni. S-au experimentat clasificatori de tip Joint Boosting pentru a împărți regiunile de interes în 4 clase: pietoni, maşini, stâlpi și alte obstacole (care nu pot fi recunoscute).

Experimentele preliminare au considerat peste 20000 de instanțe din fiecare clasă. Evaluarea s-a făcut prin crossvalidation cu 10 iterații. Performanța modelului este arătată în Tabelul 1:

Clasa	TP Rate	FP Rate
Maşină	0.95	0.07
Pieton	0.96	0.06
Stâlp/Copac	0.93	0.08
Alte obstacole	0.82	0.14

Tabel 2.2.1: ROI pe imaginea stângă

Clasa	Acuratețe segmentare
Pieton vizibil	72%
Pieton parțial vizibil	69%
Grup pietoni vizibil	76%
Grup pietoni parțial vizibil	72%

Tabel 2.2.2: ROI pentru FIR

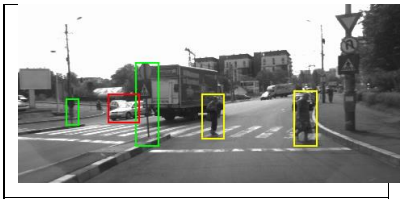


Figura 2.2.2: ROI pe imaginea stângă

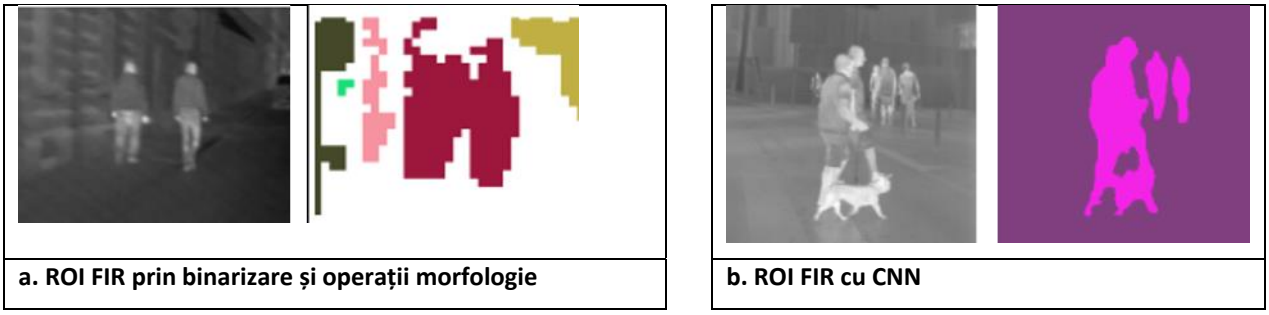
Fig.2.2.2. prezintă rezultate parțiale pe imaginea stângă. Regiunile de interes sunt marcate cu culori diferite pentru fiecare clasă.

Rezultatele pe mulțimea definită pentru experimentele preliminare sunt satisfăcătoare. Dar considerând numărul mare de obstacole care se obțin pentru o rulare în timp real este nevoie de îmbunătățirea modelului pentru a ajunge la o acuratețe mai mare.

Pentru generarea regiunilor de interes în imagini FLIR se propun două abordări:

(1) Bazată pe amprenta termală – care cuprinde binarizare și operații morfologice
(2) Bazată pe rețele neuronale convoluționale care realizează segmentarea informației termale.

La generarea regiunilor de interes pe baza amprentei termale se propune filtrarea punctelor de muchie verticală în imgini completată de operații morfologice de închidere cu filtre rectangulare de diferite dimensiuni. Pentru această abordare s-au realizat experimente pentru generarea regiunilor de interes pentru pietoni și grupuri de pietoni complet sau partial vizibili. Rezultatele experimentale sunt prezentate în Tabelul 2.2.2 și rezultatele vizuale sunt prezentate în figura 2.2.3.a:



a. ROI FIR prin binarizare și operații morfologie

b. ROI FIR cu CNN

Fig. 2.2.3. Ilustrarea rezultatelor pentru cele 2 abordari propuse pentru generatoarele ROI

A doua abordare proiectată se bazează pe rețele neuronale de segmentare. S-a proiectat un model bazat pe arhitectura ERFNet [Rom2018] care se bazează pe un encoder care micșorează hărțile de trăsături urmat de un decoder care realizează segmentarea și scalarea la rezoluția originală a hărților de trăsături. În experimentele realizate s-a măsurat metrica Intersection Over Union și acuratețea cea mai bună a fost de 80%. Rezultatele vizuale pe imagini FIR sunt prezentate în figura 2.2.3.b.

2.3. Dezvoltarea mecanismului de clasificare redunant multi-senzorial si multi-spectral al obiectelor (SO3.3-2)

Mecanismul de clasificare multiredundant si multispectral al obiectelor va permite detectia obiectelor de interes din scenariile de trafic pe baza reprezentarii scenei obtinuta prin fuziunea de tip low-level a datelor multisenzoriale primare (STMAR) peste care se vor aplica inferente specifice fiecarei categorii de obiect care se doreste a fi detectat. Redundanta senzoriala va permite o detectie robusta a categoriilor de obiecte chiar si in conditiile in care unul dintre senzori nu este capabil sa furnizeze date (conditii de vizibilitate deficitara etc.) sau va mari gradul de confidenta al rezultatelor detectiilor. In continuare se va exempifica mecanismul propus pentru detectia unor categorii de obiecte relevante din trafic, atat dinamice (pietonii) cat si statice (structurile verticale - stalpi, structuri orizontale – borduri).

2.3.1. Detectia pietonilor in spatiul 3D fuzionat multispectral

Pentru detectia pietonilor in spatiul fuzionat s-a dezvoltat un mecanism conform celui prezentat in figura 2.3.1:

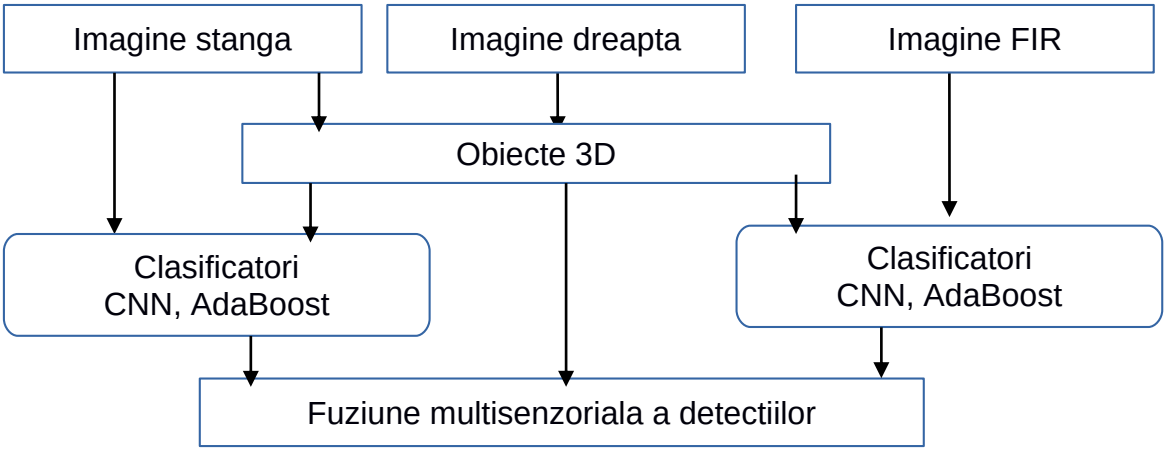


Fig. 2.3.1. Arhitectura metodei de detectie a pietonilor in spatiul 3D fuzionat multispectral

S-au realizat experimente pentru detectia pietonilor in spatiul infrarosu si in cel grayscale. Pentru aceasta au fost antrenati clasificatori de tip AdaBoost cu canale agregate de informatie (Aggregated Channel Features)[Dol2014] si se propune o implementare a detectiei utilizand clasificatori bazati pe retele neuronale convolutionale pornind de la arhitectura retelei YOLO[Red2018]. Pentru antrenarea ACF s-au extras trasaturi de histograma orientarii gradientului si sabloane locale binare (LBP) atat pe imaginea stanga cat si pe imaginea FIR. S-au antrenat clasificatori specializati pentru fiecare tip de imagine (grayscale si infrarosu). Fuziunea multisenzoriala a detectiilor combina detectiile celor doi clasificatori cu ipotezele 3D ale obiectelor pentru a obtine detectii multisenzoriale si multimodale. Experimentele preliminare pentru detectia pietonilor au dovedit o acuratete satisfacatoare a detectiilor pe cele doua imagini, aceasta acuratete putand fi imbunatatita prin introducerea unor clasificatori mai puternici cum sunt cei bazati pe retele neuronale. In acest sens s-a proiectat o arhitectura de retea neuronala convolutionala completa (fully convolutional neural network) [Bre2018]. Arhitectura YOLO a fost modificata pentru a invata trasaturi relevante pentru imaginile infrarosu dar si pentru imaginile grayscale. In experimentele preliminare s-a obtinut o medie a intersectiei pe reuniune (Intersection over union) de aproape 70% pe multimea de antrenare. Reteaua a fost antrenata pe imaginile din multimea de date rezultata dupa adnotare. Aceasta metrica va fi imbunatatita prin furnizarea unor imagini mai variate si mai multe (cateva mii) in multimea de antrenare.

2.3.2. Detectia structurilor verticale

Într-un scenariu urban se întâlnesc o varietate de structuri asemănătoare stâlpilor, precum stâlpi pentru semnele din trafic, semafoare, stâlpi utilitari, de iluminare sau de semnalizare. Acestea reprezintă repere robuste care pot ajuta în cazul unor probleme cum ar fi localizarea mașinii, crearea unei hărți a mediului sau navigare autonomă. Prin metoda pe care o propunem [Bla2018], se rezolvă problema extragerii stâlpilor având ca intrare măsurătorile de la camera stereo, chiar si inconditii de iluminare variabila/slaba.

Metodologia pentru detecția stâlpilor

1. ***Extragerea candidaților:*** Din imaginea stângă de intensitate, convertită în tonuri de gri, se extrag muchiile cu ajutorul algoritmului lui Canny. S-a aplicat un pas de normalizare a gradienților care a îmbunătățit detecția în diferite cazuri de iluminare și mai ales în zone cu umbre proeminente.Pe imaginea care conține muchii s-a aplicat algoritmul Probabilistic Hough Transform pentru detectarea liniilor. Unghiul acestora a fost limitat între 60 și 120 de grade. Mai apoi, acestea au

fost grupate în perechi, începând din partea stângă. S-a impus o limită de distanță între două linii, pentru a limita grosimea în pixeli a detecțiilor. Regiunile de interes sunt dreptunghiuri care s-au aflat prin calcularea minimului și maximului pozițiilor pe axele x și y ale perechiilor de linii.

2. ***Selectarea stâlpilor:*** În acest pas, s-a folosit informația de disparitate pentru a crea o hartă 3D a mediului înconjurător. Aceasta conține informații importante legate de poziția și adâncimea stâlpilor. Pentru a afla poziția, vom proiecta punctele mai întâi pe orizontală și extragem regiunile de maxim. Mai apoi, proiectăm punctele pe verticală, pentru a obține planul drumului. Combinând aceste două informații, suntem capabili să extragem poziția de start a stâlpilor, în imagine. În fig. 2.3.2 se observă marcat cu verde, regiunile de interes, iar cu albastru, zonele propuse pentru poziția de start a stâlpilor. Prin combinarea acestor două informații, suntem capabili să grupăm regiunile de interes deasupra planului drumului și să extragem regiunea finală a reperului dorit.

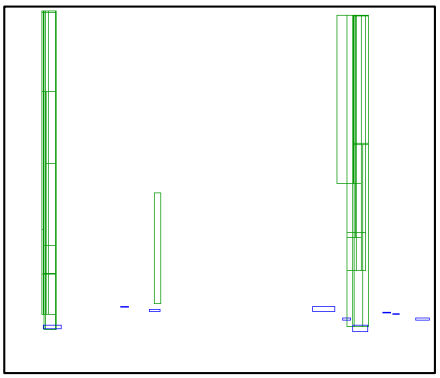


Fig. 2.3.2. Regiunile de interes și pozițiile de start

3. ***Eliminarea detecțiilor false:*** Deoarece există detecții greșite care apar pe suprafețe plane, cum ar fi clădirile, aplicăm o serie de filtre pentru a le elimina. De exemplu, prin analiza histogramei disparităților, dacă există o variație mare a acestora, detecția este anulată.

Îmbunătățirea imaginilor pe timp de noapte

Folosind metoda propusă în [Shi2018], aplicăm o serie de pași pentru a detecta lumina atmosferică A și funcția de transfer a luminii în mediu $t(x)$, pentru a corecta iluminarea astfel:

$$J(x) = \frac{I(x)-A}{t(x)} + A \tag{2.3.1}$$

Poziționarea stâlpilor

În ultima etapă, folosind informația de adâncime oferită de camera stereo și poziția GPS a mașinii, putem determina latitudinea și longitudinea stâlpilor detectați folosind formulele:

$$P_{lat} = lat + \frac{180}{\pi} \cdot \frac{P(Y)}{63813} \tag{2.3.2}$$

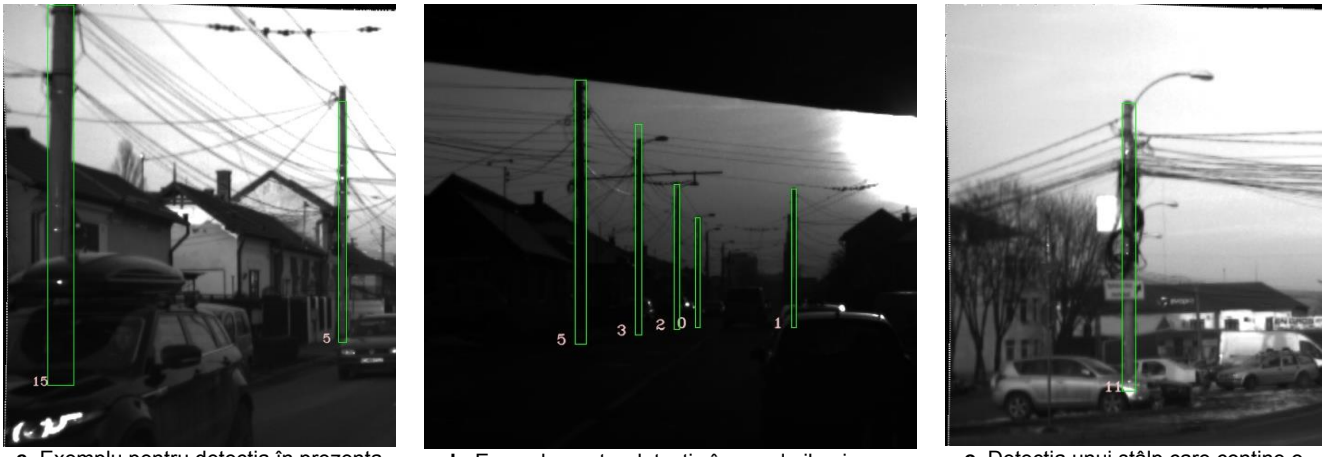
$$P_{long} = long + \frac{180}{\pi} \cdot \frac{P(X)}{63813} \cdot \frac{1}{\cos(lat)} \tag{2.3.3}$$

Rezultate

Pentru a testa algoritmul, am analizat patru secvențe. Calitatea detecțiilor este rezumată în Tabel , unde capetele tabelului au următoarele semnificații: **F** este numărul de cadre dintr-o secvență, **T** – numărul total de stâlpi aflați în secvență, **D** – numărul total al detecțiilor folosind metodologia propusă, **TP** – numărul detecțiilor corecte, **FP** – numărul detecțiilor incorecte, **P** – precizie, **R** – reamintire (recall), **C** – numărul mediu al cadrelor consecutive în care este detectat același reper. În figurile 2.3.3.b si 2.3.3.c, se pot observa cazurile mai dificile în care metoda noastră e capabilă să detecteze cu succes stâlpii.

Tabel 2.3.2. Acuratețea metodologiei pentru detecția stîlpilor

Set	F	T	D	TP	FP	P	R	C
1)	830	31	40	27	13	*67.5%*	*87%*	15
2)	1,268	58	72	58	14	*80.5%*	*100%*	20
3)	2,765	84	103	78	25	*75.7%*	*92.8%*	20
4)	1,325	48	59	45	14	*76.2%*	*93.7%*	15



a. Exemplu pentru detecția în prezența obstacolelor

b. Examplu pentru detecție în caz de iluminare scăzută

c. Detecția unui stâlp care conține o aglomerare de alte obiecte

2.3.3. Rezultate experimentale aferente detectiei structuirilor verticale

2.3.3. Detectia bordurilor

Detectia bordurilor este un subiect foarte important pentru masinile autonome sau cele cu sisteme ADAS. Este un pas vital pentru detectarea suprafetei navigabile a drumului, planificarea traiectoriei, parcarea masinii cat si pentru o localizare mai exacta a masinii pe o harta a mediului cunoscuta. Metoda propusa pentru detectia bordurilor se bazează pe fuziunea dintre imaginile segmentate semantic obținute utilizând o rețea neuronală convoluțională aplicata pe imaginile venite de la camerele masinii și un nor de puncte 3D venit de la senzorul LiDAR. Pipeline-ul siste,ului propus este vizibil in figura 2.3.4.

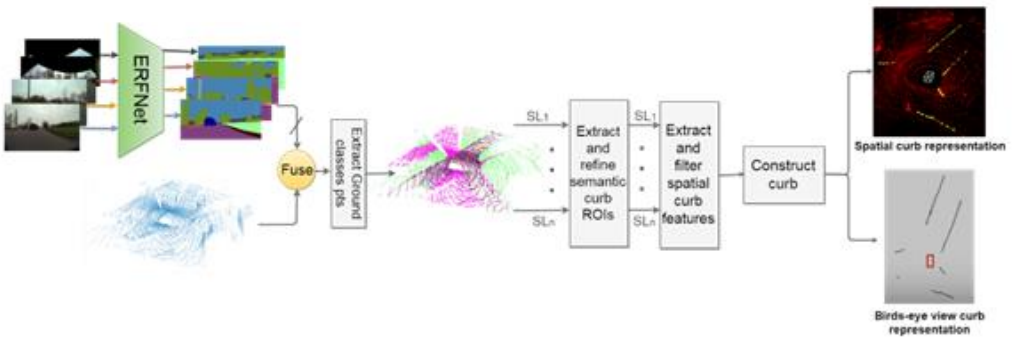


Fig. 2.3.4. Pipeline-ul metodei de detctie a bordurilor

Informațiile semantice vor fi utilizate pentru a exploata contextul pentru detectia bordurilor din mediul urban. Folosind doar etichetele semantice asociate punctelor 3D, vom defini un set de regiuni de interes (ROI) în care bordurile sunt cel mai probabil localizate, reducând astfel spațiul de căutare. Astfel, un ROI pentru bordura se va afla acolo unde informatia semantica a detectat o regiune de puncte continuu etichetate ca bordura sau in dreptul unei tranzitii dintre regiunea de drum si cea a unei suprafete adiacente (de ex: zona cu vegetatie, zona pietonala) pentru care sunt sanse mari sa existe o bordura. Fiecare ROI este definit pe o linie a scanerului laser, avand o pozitie de inceput, corespunzatoare regiunii dinspre drum si o pozitie de sfarsit, corespunzatoare regiunii dinspre suprafata adiacenta drumului. Pentru a rafina ROI-urile, pozitiile de inceput si sfarsit ale acestora se vor mari cu o fereastra de marime exact definita (1.5m-3m) pentru a capta cat mai bine structura bordurii.

O metodă tradițională de detectie a bordurilor inspirata din literature de specialitate tipica senzorului LiDAR este utilizată în continuare pentru a corecta ROI-urile obținute anterior. In cadrul fiecarui ROI, caracteristicile spațiale ale celei mai proeminente regiuni monotona ascendenta sunt extrase și filtrate utilizând măsurătorile de înaltă precizie ale LiDAR-ului. Caracteristicile utilizate vor fi:

1. diferenta de elevatie dintre extremitatile zonei montona ascendenta [Zha2018], [Yao2012], [Che2015], [Yan2014]
2. unghiul dintre pozitia de inceput a ROI-ului si a pozitiei de inceput (unde se gaseste punctul cu elevatia cea mai mica) si de sfarsit (unde se gaseste punctul cu elevatia cea mai mare) a zonei monotona ascendenta [Zha2018]

Ambele caracteristici trebuie sa fie incluse intr-un interval exact definit pentru ca ROI-ul sa fie considerat ca avand intradevar o bordura in interiorul sau: [2cm,25cm] pentru diferenta de elevatie si [80 grade, 150 grade] pentru unghi. Daca un ROI a trecut ambele teste, punctul de bordura din interiorul lui este ales ca fiind pozitia de inceput a zonei monotona ascendenta.

In continuare, reconstructia bordurii finale este obtinuta prin trasarea poliliniilor intre punctele de bordura extrase anterior din ROI-uri valide invecinate orientate in directii opuse.

Dintre avantajele metodei propuse amintim:

- Flexibilitatea: Metoda propusa detecteaza borduri de inaltimi variate incepand de la 2 cm pana la 25-30 de cm.
- Adaptabilitatea: Metoda propusa detecteaza borduri pe orice tip de drum dintr-o scena urbana: drum drept, drum curbat, intersectie in T cat si intersectii mai complexe cu sensuri giratorii incluse.

Metoda are si cateva dezavantaje::

- Daca segmentarea semantica eticheteaza nu eticheteaza bordura sau o eticheteaza deplasat (ex. In mijlocul drumului) pentru un frame, bordura nu va fi detectata chiar daca aceasta exista.
- Outlieri pot fi depistati in zonele cu vegetatie pe teren apropiate de drum si etichetate semantic incorect.
- Distanta la care sunt detectate bordurile este - max 30m, intrucat la o distanta mai mare numarul outlerilor creste intrucat precizia etichetelor semantice scade.

Cateva rezultate obtinute pot fi vizualizate in figura 2.3.5:

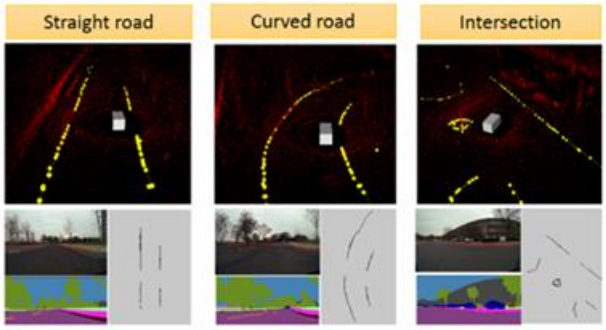


Fig. 2.3.5. Rezultate obtinute pe diverse scenarii

2.4. Dezvoltarea uneltei de adnotare si construirea seturilor de date pentru antrenare şi testare (SO3.2-2)

Pentru adnotarea imaginilor FIR și a imaginilor de intensitate s-a experimentat unealta de adnotare de la Caltech [Dol2012]. Aceasta permite definirea de etichete, propagarea adnotării pe mai multe cadre prin aplicarea de interpolare biliniară, corectarea sau ștergerea adnotărilor.

Deoarece această unealtă nu poate fi adaptată pentru a adnota imagini FIR și de intensitate în paralel s-a dezvoltat o unealtă de adnotare proprie. Într-o primă fază s-au adnotat pietoni complet vizibili și pietoni partial vizibili. După cum se poate observa în Figura 5 unealta dezvoltată permite navigarea printr-o secvență, adnotarea pietonilor, a mașinilor și a grupurilor de pietoni din imagini FIR și imaginea stângă. Unealta permite selectarea de culori diferite pentru instanțele fiecărei clase și asigură faptul unitatea de culoare între adnotările din FIR și adnotările din imaginea stângă. Se permite astfel realizarea adnotărilor unice pe instanță, adică o instanță de pieton va fi adnotat cu aceeași culoare în imaginea FIR și în imaginea stângă. Unealta de adnotare a fost dezvoltată în Visual Studio 2015 utilizând OpenCV. La pornire este nevoie de transmiterea directorului unde se află imaginile de adnotat sau imaginile deja adnotate. Adnotările se salvează sub forma unor fișiere .txt care au același nume cu imaginea adnotată (diferă doar extensia). Fișierul de adnotări asociat unei imagini cuprinde câte o adnotare pe fiecare linie. Adnotările sunt de forma:

TipImagine, NumeClasă, CuloareAdnotare, x,y, width, height, unde:

- *TipImagine* este 1 – dacă adnotarea se aplică la imaginea stângă, 7 dacă adntarea se aplică la imaginea FIR
- *NumeClasă* este numele clasei referite de adnotare. Poate fi Pedestrian, Car, Group.
- *CuloareAdnotare* este o pereche de 3 numere care referă codul Blue, Green, Red cu care s-a colorat adnotarea.
- *x,y, width, height* reprezintă coordonatele dreptunghiului care circumscrie adnotarea.

Dacă există imagini care au deja adnotări acestea vor fi afișate automat pe imaginea FIR și pe imaginea stângă. Pentru modificarea unei adnotări se face click pe dreptunghiul care mărgineste adnotarea. Aceasta permite modificare dimensiunii adnotării sau ștergerea ei dacă se apasă tasta “d”.

Fig. 2.4.1: Adnotare pietoni și mașini

Pentru experimentele preliminare s-au adnotat pietoni complet sau partial vizibili. Mulțimea adnotată cuprinde 2350 de cadre FIR și cele 2350 cadre corespunzătoare din imaginea stângă. Adnotările conțin cam 1000 instanțe de pietoni.

Ca și dezvoltări ulterioare se propun adăugarea unor butoane pentru o interfață grafică mai intuitivă, cum ar fi buton pentru rularea continuă a secvenței, oprirea la un cadru, navigarea la primul cadru sau la ultimul cadru din secvență. Se va studia și dacă este necesară introducerea unei funcții care realizează interpolarea între cadre successive și dacă este necesară introducerea unei ferestre pentru vizualizarea punctelor 3D.

2.5. Proiectarea demonstratorului (SO4.1-1)

Pentru experimentarea metodelor/algoritmilor dezvoltati in cadrul proiectului se va folosi platforma mobila de cercetare aflata in dotarea centrului IPPPRC. Sistemul hardware este compus din pachetul principal de senzori: Sistem Stereo, Camera LWIR, LiDAR 4 Canale și Velodyne PUCK; împreună cu unitatea centrală de calcul și echipamentul auxiliar – switch Ethernet pentru conectarea senzorilor, GPS Velodyne pentru sincronizarea ceasului Velodyne PUCK, interfața CAN pentru date despre ego-mișcare, etc.

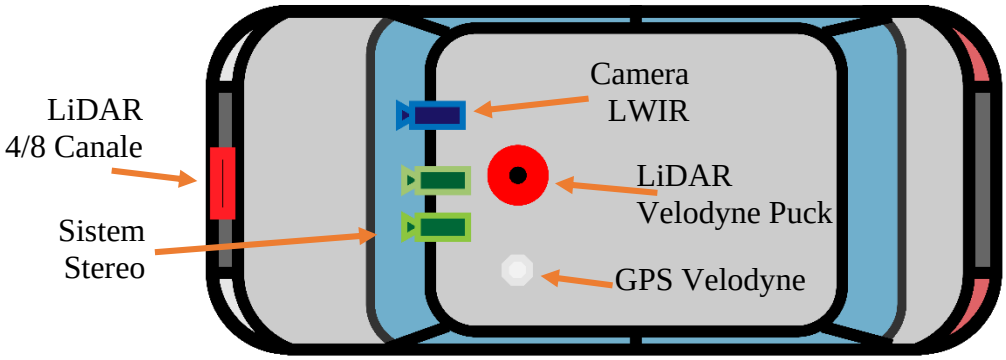


Fig. 2.5.1. Schema de montaj a senzorilor pe platforma mobila de cercetare.

Pentru a demonstra capabilitățiile componentelor software ale demonstratorului, acestea vor fi implementat sub forma unor module compatibile cu unul dintre framework-urile standard în industrie, EB Assist ADTF (Automotive Data and Time-Triggered Framework). Acesta permite operarea unui sistem definit sub forma unor fișiere de configurare atât în mod ”direct” (utilizând date capturate în timp real de la senzori) cât și în mod ”redare”, în care operarea senzorilor este simulată folosind datele capturate în urma unui proces de înregistrare. Figura de mai jos prezinta principalele module software, precum și conexiunile dintre acestea.

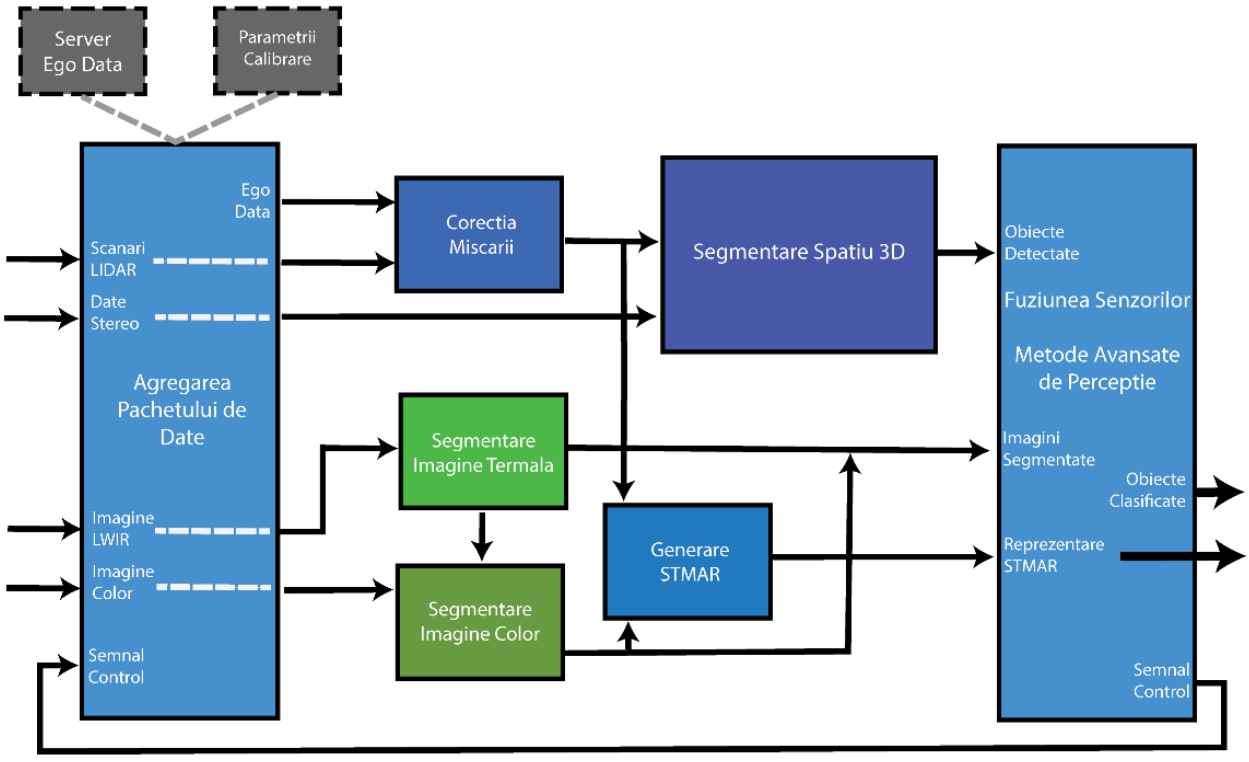


Fig. 2.5.2. Arhitectura software a demonstratorului

Primul modul din lanțul de procesare are rolul de a pregăti pachete noi de măsurători, de a lansa execuția stagiilor de procesare și de a monitoriza starea sistemului până la furnizarea rezultatelor. Modulele software constituente, precum cele de segmentare, generare a reprezentării STMAR, etc. sunt decuplate de modul de operare (direct/redare) al sistemului la rulare, fapt ce ușurează designul acestora. Un punct de interes important luat în considerare este reprezentat de performanța sistemului din puctul de vedere al timpului de procesare. Din acest motiv, arhitectura modulelor contine optimizări în acest sens.

2.6. Diseminare rezultate (SO4.2-1)

Diseminarea rezultatlor s-a materializat prin prezentarea si publicarea a 8 lucrari la conferinte internationale:

- 2018 IEEE Intelligent Vehicles Symposium (IV),26-30 June 2018, Changshu, China;
- 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 6-8 Sept. 2018, Cluj-Napoca, Romania
- 2018 IEEE Intelligent Transportation Systems Conference (ITSC), 4-7 Nov. 2018, Maui, Hawaii, USA.

Mentionam ca toate volumele in care s-au publicat lucrarile sunt sau vor fi indexate BDI (IEEE Xplore Digital Library) si vor fi indexate in Clarivate Analytics Web of Science (fosta ISI Proceedings)

1. V. Miclea, S. Nedevschi, "Real-time Semantic Segmentation-based Depth Upsampling using Deep Learning", 2018 IEEE Intelligent Vehicles Symposium (IV), 26-30 June 2018, Changshu, China, p. 300-306, DOI: 10.1109/IVS.2018.8500555.
2. R. Brehar, F.I. Vancea, T. Mariţa, S. Nedevschi. "A Deep Learning Approach For Pedestrian Segmentation In Infrared Images", 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 6-8 Sept. 2018, Cluj-Napoca, Romania, p. 253-258, DOI: 10.1109/ICCP.2018.8516630
3. B.C.Z. Blaga, S. Nedevschi, "A Method for Automatic Pole Detection from Urban Video Scenes using Stereo Vision", 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 6-8 Sept. 2018, Cluj-Napoca, Romania, p. 293-300, DOI: 10.1109/ICCP.2018.8516640.
4. H. Florea, R. Varga, S. Nedevschi, "Environment Perception Architecture using Images and 3D Data", 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 6-8 Sept. 2018, Cluj-Napoca, Romania, p. 223-228. DOI: 10.1109/ICCP.2018.8516581
5. M.P. Muresan, S. Nedevschi, "Multimodal sparse LiDAR object tracking in clutter", 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 6-8 Sept. 2018, Cluj-Napoca, Romania, p. 215-221, DOI: 10.1109/ICCP.2018.8516646
6. F.I. Vancea, S. Nedevschi, "Semantic information based vehicle relative orientation and taillight detection". 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 6-8 Sept. 2018, Cluj-Napoca, Romania, p. 259-264, DOI: 10.1109/ICCP.2018.8516631
7. A.D. Costea, A. Petrovai, S. Nedevschi, "Fusion Scheme for Semantic and Instance-Level Segmentation", 2018 IEEE Intelligent Transportation Systems Conference (ITSC), 4-7 Nov. 2018, Maui, Hawaii, pp. 3469-3475 , DOI: 10.1109/ITSC.2018.8570006

8. V. Miclea, S. Nedevschi, L. Miclea, "Real-Time Stereo Reconstruction Failure Detection and Correction Using Deep Learning", 2018 IEEE Intelligent Transportation Systems Conference (ITSC), 4-7 Nov. 2018, Maui, Hawaii, pp. 1095-1102, DOI: 10.1109/ITSC.2018.8569928
9. Cristian-C. Vancea, Sergiu Nedevschi, "Image Rectification and Matching Solutions Optimized for Dedidated Hardware Accelerators", Automation Computers Applied Mathematics (ACAM), vol. 27, nr. 1, pg. 13-23, 2018, ISSN 1221-437X

Referinte bibliografice

[Bil2014] G.-A. Bilodeau, A. Torabi, P.-L. St-Charles, and D. Riahi, \Thermal{visible registration of human silhouettes: A similarity measure performance evaluation,' Infrared Physics & Technology, vol. 64, pp. 79{86, 2014.

[Bre2018] R. Brehar, F.I. Vancea, T. Mariţa, S. Nedevschi. "A Deep Learning Approach For Pedestrian Segmentation In Infrared Images", 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 6-8 Sept. 2018, Cluj-Napoca, Romania, p. 253-258.

[Che2015] T. Chen, B. Dai, D. Liu, J. Song, Z. Liu, "Velodyne-based curb detection up to 50 meters away",Intelligent Vehicles Symposium (IV), pp. 241-248, 2015.

[Dol2012] P. Dollar, C. Wojek, B. Schiele and P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 4, pp. 743-761, April 2012.

[Dol2014] P. Dollár, R. Appel, S. Belongie and P. Perona, "Fast Feature Pyramids for Object Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 8, pp. 1532-1545, Aug. 2014. doi: 10.1109/TPAMI.2014.2300479

[Gal2015] R. Galatus, N. Puscas, T. Marita**,** *Senzori Optici: concepte fundamentale si aplicatii*, Editura Casa Cartii de Stiinta, Cluj-Napoca, 2015, ISBN 978-606-17-0748-5.

[Gey2000] Geyer, C., & Daniilidis, K. (2000, June). A unifying theory for central panoramic systems and practical implications. In *European conference on computer vision* (pp. 445-461). Springer, Berlin, Heidelberg.

[Har2003] R Hartley, A Zisserman, Multiple View Geometry in Computer Vision, 2-nd ed,Cambridge University Press,2003

[He2016] K. He, X. Zhang, S. Ren, J. Sun. "Deep residual learning for image recognition". IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[He2017] K. He, G. Gkioxari, P. Dollar, R. Girshick. "Mask R-CNN". IEEE International Conference on Computer Vision (ICCV), 2017.

[Hon2010] S. Hong, H. Ko, and J. Kim, "Vicp: Velocity updating iterative closest point algorithm" in Robotics and Automation (ICRA), 2010 IEEE International Conference on. IEEE, 2010, pp. 1893-1898.

[Lin2017] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie. "Feature pyramid networks for object detection". IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[Ned2012] S. Nedevschi, R. Danescu, F. Oniga, T. Marita, Tehnici de viziune artificiala aplicate în conducerea automata a autovehiculelor, Editura U.T. Press, Cluj-Napoca, 2012.

[Ned2017] S.Nedevschi, et. al, “Perceptia multispectrala a mediului prin fuziunea datelor senzoriale 2d si 3d din spectrul vizibil si infra-rosu (MULTISPECT)”, raport stiintific si tehnic (RST), dec. 2017

[Red2018] Redmon, Joseph and Ali Farhadi. “YOLOv3: An Incremental Improvement.” CoRR abs/1804.02767 (2018): n. pag.

[Rom2018] E. Romera, J. M. lvarez, L. M. Bergasa, and R. Arroyo, “Erfnet: Efficient residual factorized convnet for real-time semantic segmentation,” IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 1, pp. 263–272, Jan 2018.

[Shi2018] Z. Shi, M. M. Zhu, B. Guo, M. Zhao, and C. Zhang, "Nighttime low illumination image enhancement with single image using bright/dark channel prior," EURASIP Journal on Image and Video Processing, vol. 2018, p. 13, February 14 2018.

[Tor2011] A. Torabi, M. Najafianrazavi, and G.-A. Bilodeau, A comparative evaluation of multimodal dense stereo correspondence measures,’ in Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on. IEEE, 2011, pp. 143{148.

[Yan2014] B. Yang, L. Fang, J. Li. "Semi-automated extraction and delineation of 3D roads of street scene from mobile laser scanning point clouds", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 79, pp. 80-93, 2013.

[Yao2012] W. Yao, Z. Deng, L. Zhou. "Road curb detection using 3D LiDAR and integral laser points for intelligent vehicles”, In Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), pp. 100-105, 2012.

[Zha2018] Y. Zhang, J. Wang, X. Wang, J. Dolan, et al. "Road-segmentation-based curb detection method for self-driving via a 3D-LiDAR sensor." IEEE Transactions on Intelligent Transportation Systems, vol. 99, pp. 1-11, 2018.