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 3 / 2019

Director proiect:

Prof. dr. ing. Sergiu Nedevschi

Colectiv:

Conf. dr. ing. Tiberiu Marita
Sef lucrari dr. ing. Raluca Brehar
Sef lucrari dr. ing. Robert Varga
Doctorand ing. Cristian Vancea
Doctorand ing. Arthur Costea
Doctorand ing. Mircea Muresan
Doctorand ing. Flaviu Vancea
Masterand ing. Horatiu Florea
Masterand ing. Zelia Blaga
Masterand ing. Selma Goga (cas. Deac)
Masterand ing. Barbu Florin-Alexandru

Cluj-Napoca, Decembrie 2019

Raport stiintific

privind implementarea proiectului in perioada: ianuarie – decembrie 2019

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

Denumire etapa 3: Implementarea, testarea, evaluarea si optimizarea metodelor originale pentru perceptia multispectrala si multisenzoriala a mediului

Activitati etapei 3/2019

3.1 Implementarea, testarea, evaluarea si optimizarea generatoarelor de regiuni de interes pentru obiecte din imagini (SO3.1-3)

3.2. Implementarea, testarea, evaluarea si optimizarea mecanismului de clasificare redundant multi-senzorial multi-spectral al obiectelor (SO3.3-3)

3.3. Implementarea, testarea, evaluarea si optimizarea demonstratorului (SO4.1-2)

3.4. Diseminare rezultate (SO4.2-2)

3.5. Implementarea, testarea, evaluarea si optimizarea metodelor originale pentru alinierea si reprezentarea 3D spatio-temporala si bazata pe aparente a datelor multi-sensoriale si multi-spectrale primare (SO2.1-3, SO2.2-3)

Cuprins

3.1 Implementarea, testarea, evaluarea si optimizarea generatoarelor de regiuni de interes pentru obiecte din imagini (SO3.1-3)

Regiunile de interes pentru obiecte precum pietoni, maşini, stâlpi, alte obstacole au fost generate astfel:

(1) - prin segmentarea norului de puncta 3D obținute prin stereoviziune sau de la senzorii LiDAR in obiecte de interes
- proiectarea obiectelor 3D pe imaginile de intensitate
- clasificarea zonelor din imagine selectate prin metode de Boosting

(2) - utilizând mecanisme de analiză și clasificare a amprentei termale corespunzătoare pietonilor în imaginea FIR/LWIR.

Descrierea datelor utilizate în evaluare:

În experimentele realizate pentru a valida eficiența modelului implementat s-au utilizat mai multe multimi de date:

1) Datele proprii – contin imagini adnotate ale secventelor de trafic colectate pe timp de iarna și pe timpul toamnei. Mulțimea conține 1924 de pietoni cu înălțime în contextul imaginii intre 12 și 150 de pixeli, și factor de aspect intre (0.3, 0.4, 0.5.0.6). Au fost adnotati pietoni complet vizibili, pietoni partial acoperiți, grupuri de pietoni.

2) Datele puse la dispoziție de producătorul senzorilor FLIR:

Colectia de imagini, FLIR_ADAS [FLIR2018] conține imagini color și imagini termale sincronizate temporal. Pentru aceasta mulțime de date nu se cunosc parametrii camerelor și s-a propus în cadrul acestui proiect un algoritm de identificare a pietonilor utilizand retele neuronale convolutionale și cel mai apropiat vecin într-un spațiu normalizat pentru cele 2 imagini.

Clasa	Numar adnotari in multimea de antrenare	Numar adnotari in multimea de test
Pietoni	21924	5779
Masini	40711	5432
Biciclete	3581	471
Caini	226	14

3.1.1. Generarea regiunilor de interes prin proiectia punctelor 3D pe imaginile de intensitate

Punctele 3D sunt obtinute prin procesul de stereo-reconstructie. Pentru un obiect 3D se calculează proiecția punctelor sale pe imaginea stângă și astfel se obțin regiuni de interes din imaginea stanga corespunzătoare obiectului 3D.

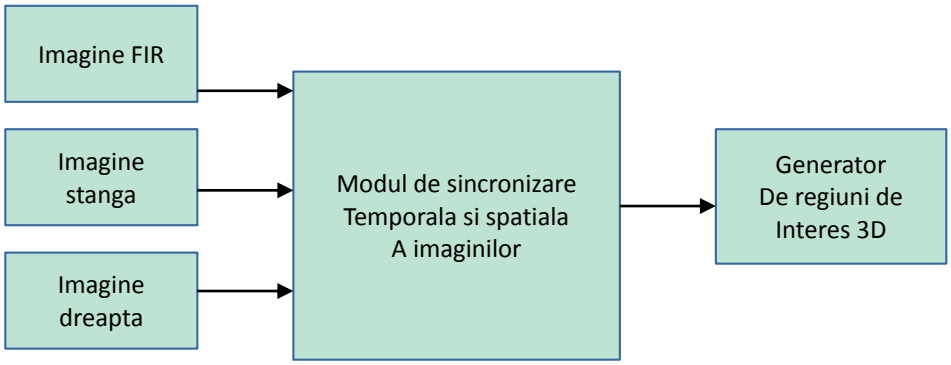

Fig. 3.1.1. Principalele componente sunt modulul de sincronizare și generatorul de regiuni de interes.

Experimentele realizate au considerat peste 20000 de instanțe din fiecare clasă (pieton, mașina, stalp). S-a obținut o rata de segmentare corecta de 90% pentru setul de date propriu, pe imaginile de intensitate și pe imaginile FIR o rata de segmentare corecta de 70%.

3.1.2. Generarea regiunilor de interes prin analiza amprentei termale corespunzatoare pietonilor în imaginea FIR/LWIR

S-au dezvoltata 2 metode una bazata pe o abordare traditional si a doua bazata pe metode de invatare.

Prima abordare de generare a regiunilor de interes prin analiza amprentei termale se bazeaza pe estimarea densitatii muchiilor verticale. Diagrama de componente a modelului este reprezentata in Figura :

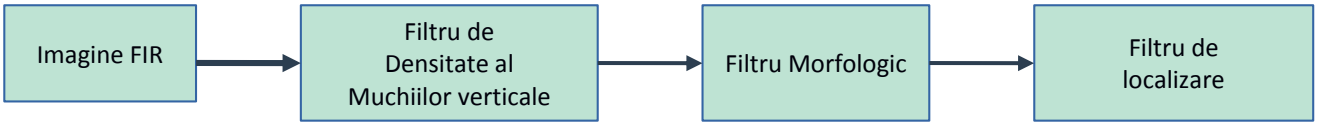

Fig. 3.1.2. Diagrama de componente a metodei bazate pe o abordare traditionala

Componenta de filtrare a densitatii muchiilor verticale foloseste operatorul Scharr pentru a determina punctele de muchie verticala. Imaginea este apoi impartita in celule de dimensiune 32x32 cu un factor de suprapunere de 16x16. Pentru fiecare celula se calculeaza densitatea de muchii. Se elimina celulele cu o densitate mica (sub 10%) sau foarte mare (peste 70%).

Componenta de filtrare morfologica realizeaza mai multe operatii de inchidere pe niveluri de gri cu elemente structurale de dimensiune (3x7), (8x5), (13x30), (18x38).

Filtrul de localizare utilizeaza informatia data de pozitia posibila a unui pieton in imagine, cunoscand parametrii senzorului infrarosu cu care au fost captate imaginile.

Evaluarea generatorului de regiuni de interes astfel obtinut se face utilizand masuratoarea definita astfel: acuratete_roi = arie(intersectie dintre ROI si pieton) / arie (pieton)

Rezultatele sunt prezentate in Tabelul de mai jos:

Clasa	acuratete_roi
Pieton	72%
Pieton vizibil partial	69.6%
Grup de pietoni	76%
Grup de pietoni partial vizibil	72%

Al doilea model de generare a regiunilor de interes pentru imaginile FIR/LWIR bazat pe amprenta termala este realizat prin segmentarea semantica utilizand metode de Deep Learning. Figura de mai sus prezinta principalele componente ale modelului propus. S-au testat diferite metode de imbunatatire a calitatii imaginilor si s-a evaluat performanta segmentarii.

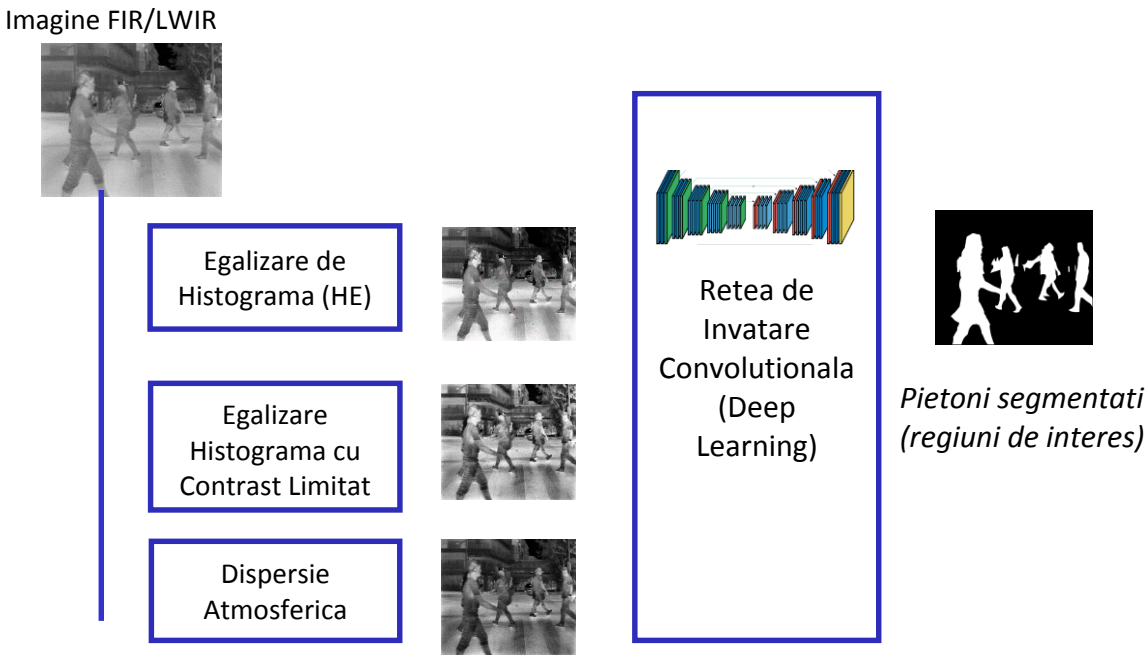

Fig. 3.1.3. Componentele generatorului de regiuni de interes generate prin metode de invatare profunda (CNN)

Componenta de egalizare a histogramei – mapeaza nivelurile de gri ale imaginii FIR la o distributie uniforma cu un spectru mai larg. In calcularea histogramei egalizate se utilizeaza histograma cumulata scalata astfel incat maximul ei sa fie 255.

Componenta de egalizare a histogramei cu contrast limitat adaptiv (CLAHE) avantajeaza imaginile care contin nuante de gri extreme (regiuni foarte inchise sau foarte deschise ale imaginii) pentru care algoritmul de egalizare a histogramei nu va evidentia suficient contrastul. Pentru a imbunatatii rezultatele, pentru fiecare pixel se realizeaza o egalizare de histograma locala. Transformarea aplicata unui pixel este proportionala cu distributia cumulativa a nivelelor de intensitate in regiunea vecina.

Pentru a imbunatati calitatea imaginii utilizand dispersia atmosferica s-a implementat algoritmul prezentat de [Li2016]. Componentele implementate au urmatoarele functii:
- pentru o imagine de intrare I se aplica un filtru de medie si se calculeaza rata de transmisie.
- calcularea imaginii imbunatatite folosind un model optic
- re-maparea nivelurilor de gri utilizand histograma egalizata cu limitare adaptiva de contrast.

Pentru generarea regiunilor de interes s-a adaptat o retea neurolana convolutionala la modelul propus. S-a implementat componenta de invatare de tip ERFNet [Rom2018].

Arhitectura retelei este prezentata in figura de mai jos:

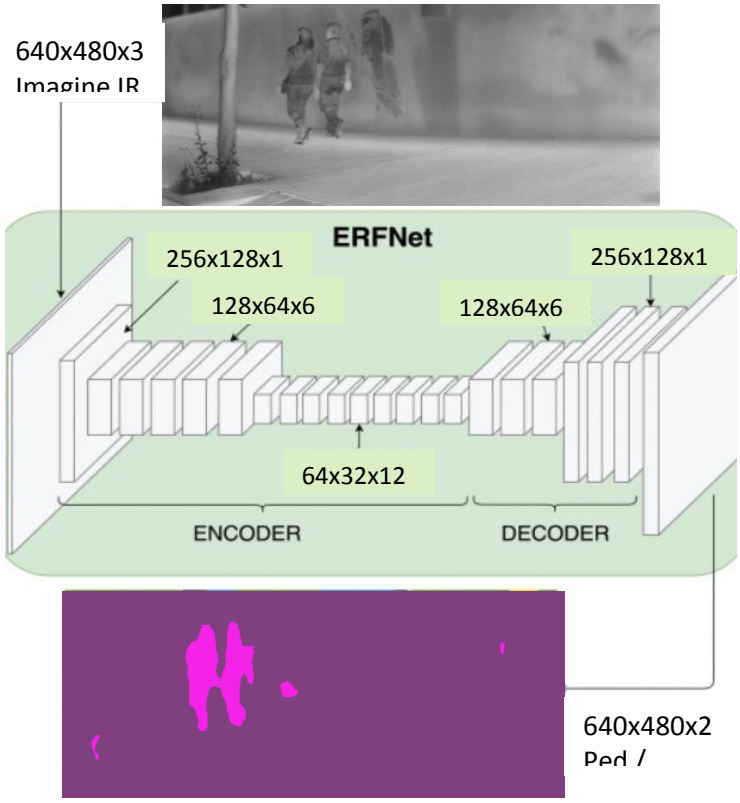

Fig. 3.1.4. Arhitectura retelei neuronale

Encoderul contine 16 nivele care contin blocuri reziduale si blocuri de down-sampling dar si convolutii dilatate. Decodificatorul (decoder) contine nivelele de la 17 la 23 si acestea au rolul de a mari hartile de trasaturi pentru a se potrivi cu dimensiunea imaginii de intrare. Totodata decodificatorul contine si nivele de deconvolutie care simplifica calculele si optimizeaza utilizarea memoriei.

Pentru evaluarea rezultatelor s-a utilizat metrica de Intersectie pe Uniune (Intersection over Union – IoU) calculata pe setul de date proprii.

Metoda de imbunatatire	**IoU**
Niciuna	75.4
Egalizare de histograma	64.6
CLAHE	63.8
Dispersie atmosferica	69.1

Se observa din tabelul rezultatelor ca media intersectiei cu uniunea este pe la 70% si ca se obtine cel mai bun rezultat de 75.4% daca asupra imaginii nu se aplica nicio imbunatatire. Timpul de executie al generarii regiunilor de interes este de 13 milisecunde, deci se obtine o viteza de procesare de 76fps. Exemple de rezultate obtinute se pot observa in figura de mai jos:

Fig. 3.1.5. Exemple de rezultate obtinute

Din analiza rezultatelor se observa ca pentru pietoni se obtine o rata de Intersection over Union de 75%, mai buna decat in cazul metodei bazate pe filtrarea amprentei termale unde s-a obtinut o acuratete maxima de 72%.

3.2. Implementarea, testarea, evaluarea si optimizarea mecanismului de clasificare redundant multi-senzorial multi-spectral al obiectelor (SO3.3-3)

Mecanismul redundant multi-senzorial multi –spectral de clasificare se bazeaza pe fuziunea la nivel de pixel sau la nivel de obiecte a informatiilor provenind de la diversii senzori utilizati. Informatia 3D primara la nivel de nor de puncte provine de la senzorul de stereoviziune sau de la senzorii LiDAR. Aceasta informatie poate fi imbunatatita prin proiectia punctelor 3D in imaginile video si infra rosu preluindu-se informatia de culoare si temperatura. In cazul segmentarii semantice a imaginilor video informatia semantica poate fi asociata punctelor 3D. Informatiile multi-senzoriale si multi-spectrale se pot colecta in oricare din aceste spatii. Fuziunea de nivel jos permite dezvoltarea unor algoritmi robusti de clasificare dar necesita un timp de procesare ridicat.

Cresterea vitezei de procesare poate fi obtinuta prin fuziunea la nivel de obiecte. In acest caz se lucreaza majoritar doar cu cuboidele s-au dreptunghiurile delimitand obiectele 3D sau 2D.

In continuare se prezinta cate un exemplu din cele 2 abordari posibile.

3.2.1 Detectia participantilor vulnerabili si a vehiculelor in spatiul fuzionat multispectral

Spatiul fuzionat multispectral contine imaginile din spectrul vizibil si imaginea infrarosu, precum si metode de trecere dintr-un spatiu in celalalt. In acest spatiu s-a implementat o metoda

robusta de detectie si clasificare a obiectelor bazata pe fuziunea la nivel de obiecte intre detectiile din spatial FIR si cele din spatiul vizibil color.

Componentele modelului de detectie si clasificare sunt prezentate in figura urmatoare:

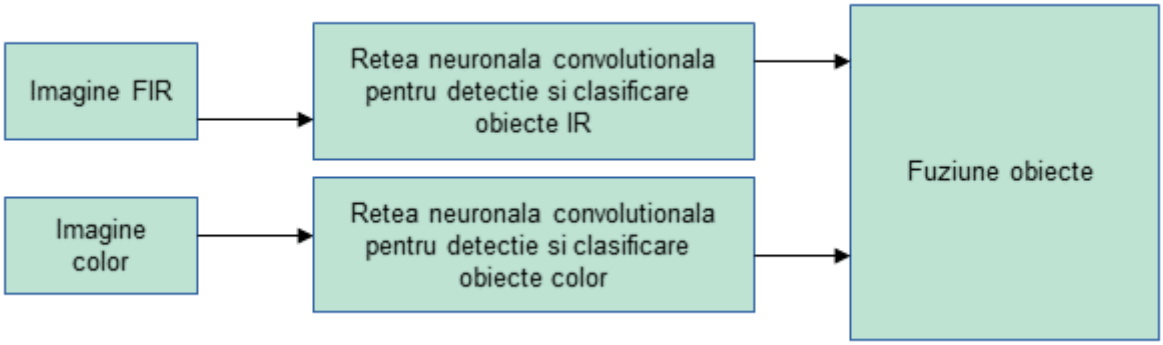

Fig. 3.2.1. Componentele modelului de detectie si clasificare

Pentru detectia din imaginile infrarosu s-a modificat o retea neuronala de tip YOLO [Red2018] si antrenarea s-a facut pe setul de date FLIR_ADAS. Pentru imaginile din multimea de antrenare s-au calculat ancorele dreptunghiurilor care marginesc obiectele utilizand algoritmul k-means. Reteaua primeste ca intrare o imagine care este impartita in mai multe regiuni si pentru fiecare regiune se realizeaza predictii utilizand ancorele pre-calculate. Pentru fiecare posibil dreptunghi care incadreaza un obiect se prezice scorul de obiect (eng. Objectness score) prin regresie logistica. Functia de entropie incrucisata binara (binary cross entropy) se utilizeaza pentru a calcula pierderea si pentru a prezice clasele pe care le poate contine un dreptunghi care incadreaza un obiect.

Reteaua neuronala convolutionala s-a antrenat pentru a detecta participantii vulnerabili cum ar fi pietoni si biciclisti dar si pentru a detecta vehicule [Bre2019a].

Rezultatele obtinute la nivel de precizie au fost urmatoarele:

Metoda	Persoana	Vehicul	Bicicleta
Yolo-v3	78.68%	84.92%	66.27%
Yolo-v3-spp	82.05%	85.78%	66.27%
RefineDetect5120	79.4%	85%	58%

Metoda RefineDetect5120 este raportata de autorii multimii adnotate FLIR_ADAS pentru ca cercetatorii sa poata compara rezultatele obtinute cu o metoda benchmark. Se poate observa ca Yolov3-spp (o retea care augmenteaza Yolov3 cu niste nivele de extragere spatiala piramidala), modificata in cadrul acestui proiect pentru a functiona cu imagini infrarosu, obtine rezultate mai bune.

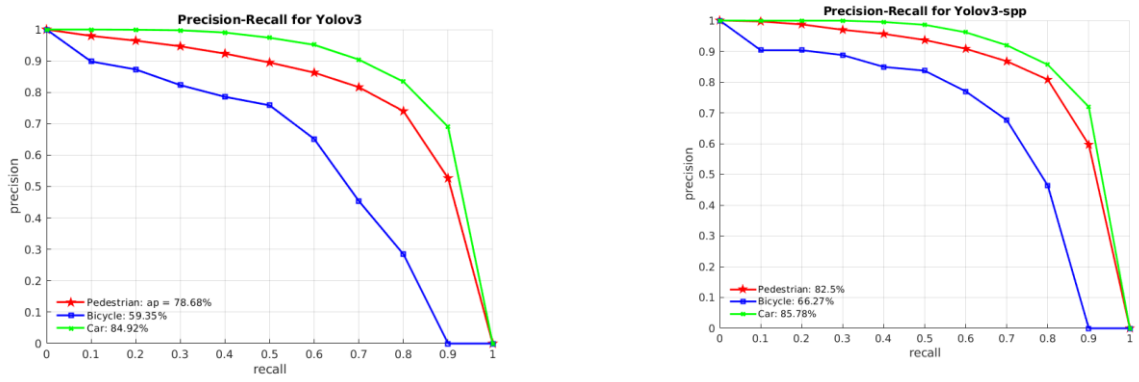

Fig. 3.2.2. Graficele de precizie pot fi observate in figurile de mai sus.

Acelasi algoitm s-a folosit si pentru detectia si clasificarea obiectelor de interes din imagini color.

Pentru a proiecta detectiile din spatiul imaginii FLIR in spatiul imaginii color s-au dezvoltat metode de proiectie bazate pe puncte cheie, vezi cap. 3.5.2. Fuziunea preia informatia mai credibila pe baza probabilitatii clasificarilor.

3.2.2. Detectia structurilor statice din scena

Detectia bordurilor

S-a implementat o metoda robsta de detectie a bordurilor bazata pe fuziunea de nivel jos dintre imagini segmentate semantic si norul de puncte 3D funrnizat de senzorii LiDAR. Pasii algoritmului implementat sunt ilustrati in figura de mai jos [Dea2019]:

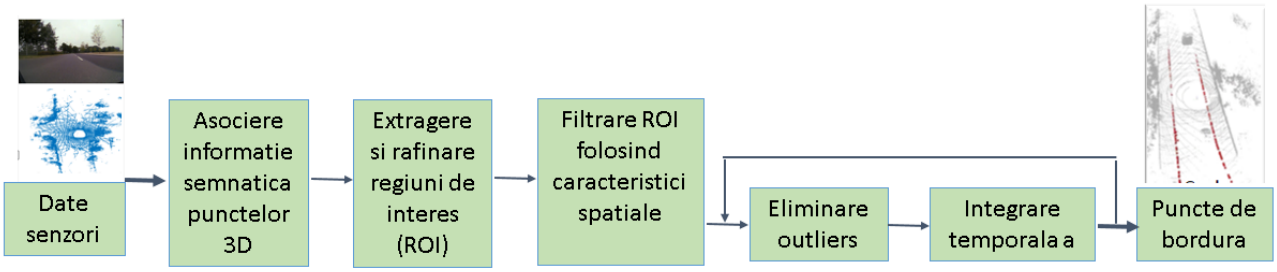

Fig. 3.2.3. Pasii algoritmului de detectie a bordurilor

1. Asocierea informatiei semantice punctelor 3D – se face printr-o fuziune de nivel jos intre punctele 3D furnizate de LiDAR si imaginile segmentate semantic [Cos2018], prin proiectia punctelor 3D furnizate de LiDAR pe planul imagine si asociera clasei semantice a pixelului 2D din imagine la punctul/voxelul 3D.

2. Extragerea si rafinarea regiunilor de interes (ROI) – regiunile de interes aferente bordurilor sunt detectate pe baza etichetei semantice asociate punctelor furnizate de LiDAR. ROI sunt extinse pentru a mari sansele de a include punctele de pe muchiile lor inferioare/superioare

3. Filtrarea regiunilor de iteres folosind caracteristici spatiale – marginile regiunilor de interes sunt micsorate pentru potrivire cu punctele de muchie ale bordurilor candidat.

4. Eliminare outliers – sunt identificate punctele candidat de bordura care descriu cel mai bine forma reala a bordurii

5. Integrarea temporala a punctelor de bordura – proprietatea de peristenta a obectelor de tip bordura este folosita pentru densificarea punctelor de curbura detectate.

Se obtine la iesire un model non-parametric de bordura sub forma unei liste de puncte. Aceste puncte corespund punctelor cu inaltimea cea mai mica (tangente cu suprafata drumului) ale bordurilor detectate.

Pentru evaluarea cantitativa a metodei s-au considerat 2 metrici: distanta medie (AvgD a punctelor de bordura detectate la o harta de precizie a scenei si precizia (PPV) punctelor de bordura detectate, obtinandu-se valori cuprinse intre 0.2...0.34 pentru AvgD si 63.84.2% pentru PPV in functie de complexitatea scenei:

Scenariu	**AvgD**	**PPV**
Drum drept	0.20 m	79.4 %
Drum cu curbura orizontala	0.24 m	63.0 %
Intersectie complexa	0.32 m	84.2 %
Intersectie de tip T’	0.24 m	81.0 %
Sens giratoriu	0.34 m	80.6 %

Eroarea medie si precizia de localizare a punctelor de bordura

3.3. Implementarea, testarea, evaluarea si optimizarea demonstratorului (SO4.1-2)

3.3.1. Componentele hardware ale demosntratorului

Demsonstratorul a fost implementat pe platforma mobila de cercetare din dotarea IPPPRC (https://erris.gov.ro/Image-Processing-and-Pattern). Sistemul hardware este compus din:

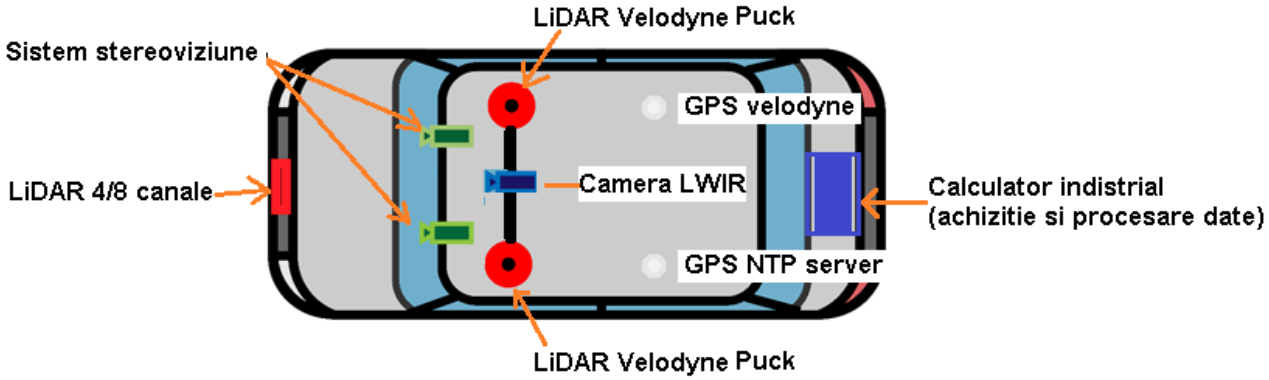

Fig. 3.3.1. Arhitectura sistemului senzorial

Sistmul senzorial utliziat pentru perceptie este compus din:

- sistem de sterovizіune alcatuit din 2 camere color Jai-Pulnix TMC-1325-CL cu senzor color (1392 x 1040 pixeli), interfata Camera Link si lentile cu distanta focala de 6.5 mm, camp vizual de 68°H x 56°V grade, dispuse cu un deplasament orizontal (baseline) de 190 mm, ceea ce asigura o adancime de detectie viabila intre 0 si cca. 45 m adancime.
- camera LWIR (LongWaveleghth Ifra Red), model “Pathfinder IR” produsa de FLIR [Fli2017], dedicata special pentru viziunea nocturna in aplicatiii automotive, resolutia native a senzorului: 320 x 240 [pixeli], extrapolabila la 640x 480 [pixeli], camp vizual 36°H x 27°V, banda spectrala: 8000 … 14000 nm cu interfata proprietara.
- doua LiDARe Velodyne Puck cu 16 raze cu interfata Ethernet.
- un LiDAR 4/8 canale model Sick AG cu interfata Ethernet.
- senzori de odometrie ai platformei mobile (senzor yaw rate sensor si viteza disponibili pe interfata CAN a platformei mobile).

Unitatea de achiztie si procesare a datelor este implemenata pe un calculator industrial MXC-6301D (fanless embedded computer) cu alimentare 12V DC de la bateria platformei mobile avand conectate urmatoarele dispozitive periferice folosite la interfatarea senzorilor si sincronizarea datelor senzoriale:

- interfata la magistrala CAN a platformei mobile Softinng CANpro USB
- frame-grabber digital SISO Menable4 pentru achizitia sincrona a imaginilor de la sistemul de stereoviziune
- switch Ethernet pentru achizitia datelor de la senzorii LiDAR
- server NTP NetBurner cu modul GPS pentru sincornizarea temporala a senzorilor

Fig. 3.3.2. Montajul fizic al senzorilor LiDAR (margini) si LWIR (mijloc) pe acoperisul platformei mobile

Fig. 3.3.3. Montajul fizic al senzorului de steoviziune color (mijloc) sub plafonul platformei mobile

Fig. 3.3.4. Unitatea centrala pentru achizitia si procesarea datelor senzoriale si sistemul electric de alimentare

Pentru a putea alinia spatial datele senzoriale relativ la un sistem unic de referinta (sistemul de coordonate asociat platformei mobile) acestia au fost calibrati prin proceduri off-line astfel:

- Parametrii intrinseci ai camerelor sistemului de steroviziune s-au calibrat printr-o metoda proprie imbunatatita a uneltei Camera calibration Toolbox [Bou2015]
- Parametrii extrinseci ai sistemului de steroviziune s-au calibrat printr-o varianta optimizata a metodei originale proprii [Mar2006]

- Parametrii extrinseci ai camerei InfarRosu (LWIR), relativi la sistemul de coordonate al senzorului de steroviziune, au fost calibrati prin metoda proprie propusa in [Ned2017]).
- Parametrii extrinseci ai senzorilor LiDAR, relativi la sistemul de coordonate al senzorului de steroviziune au fost calibrati prin metoda proprie propusa in [Ned2017], [Bla2017].
- Parametrii extrinseci absoluti ai camerei LWIR si ai senzorilor LIDAR relativi la sistemul de coordonate al platformei mobile (ego-vehicle) au fost obtinuti prin transformarile:

$$K_F^W = K_C^W \cdot K_F^C$$

$$K_L^W = K_C^W \cdot K_L^C$$

unde:

K este matricea omogena care contine matricea de rotatie si translatie relativa dintre senzorul cu indicele subscript la senzorul cu indicele superscript [Ned2017]

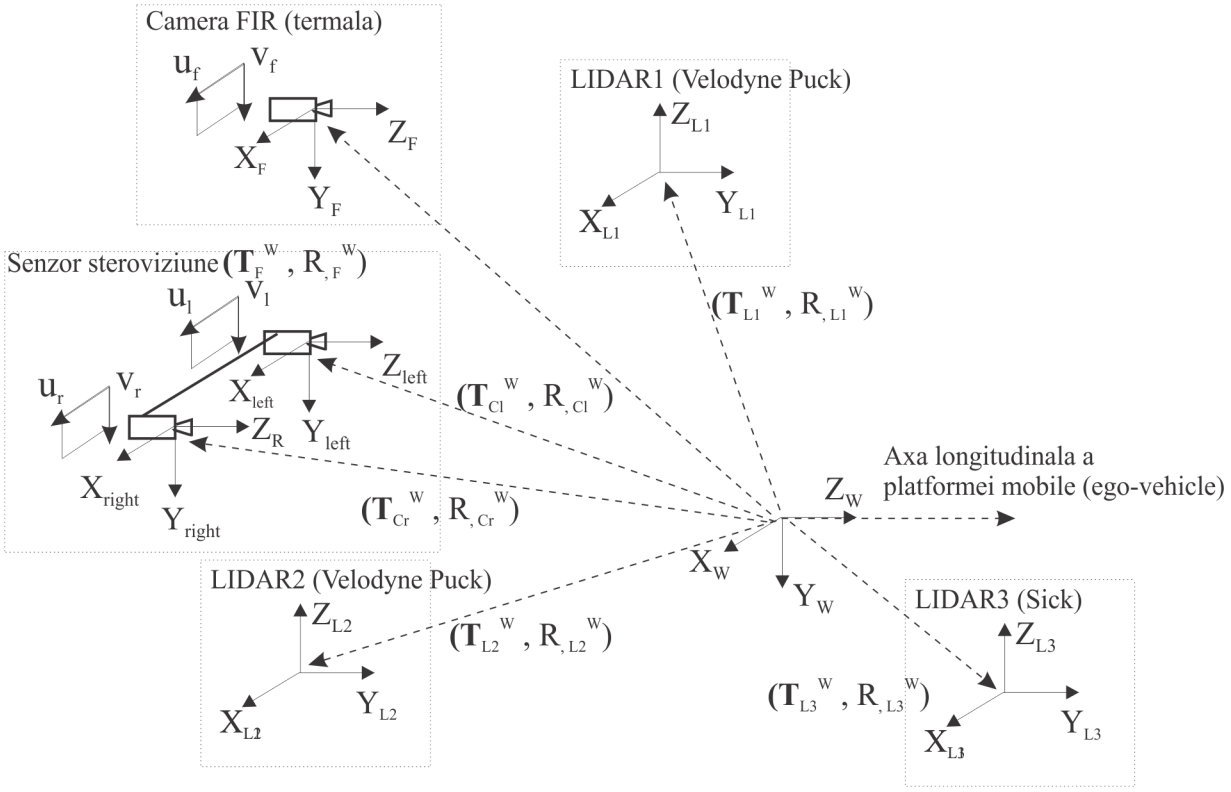

Fig. 3.3.5. Modelul geometric al sistemului senzorial

3.3.2. Aplicatia software pentru achizitia si procesarea datelor senzoriale

Aplicația este concepută ca o structură de clase și containere de tip *struct* care facilitează adaptarea funcționalității astfel încât să permită achiziția online de date și salvarea pe disc sau încărcarea unei secvențe salvată în prealabil, pentru procesare offline si testare. Diagrama de clase utilizate în cadrul aplicației este prezentată în Fig. 1 și conține o parte din principale clase utilizate la nivel înalt.

Aplicația dezvoltată are la bază o structură SequenceState care înglobează pe de o parte funcționalitatea grafică specifică interfeței utilizator și interacțiunii cu acesta, iar pe de altă parte modulul de achiziție.

Clasa OpenCVToolbarWindow din cadrul structurii de bază SequenceState conține setul de butoane de control al aplicației. Butoanele grafice sunt acționate prin click de mouse și au asociate

următoarele acțiuni: rulare continuă, rulare secvență cu secvență, oprire rulare, întoarcere la imaginea anterioara, salt la începutul secvenței, salt la sfârșitul secvenței, oprirea rulării pentru încărcarea unei alte secvențe, ieșirea din aplicație. S-a definit un set prestabilit de butoane care poate fi ajustat. La rularea aplicației in mod online o parte din aceste funcționalități nu au sens (salturi sau revenire timp) și ca atare apăsarea butoanelor asociate nu va avea efect sau poate fi dezactivată.

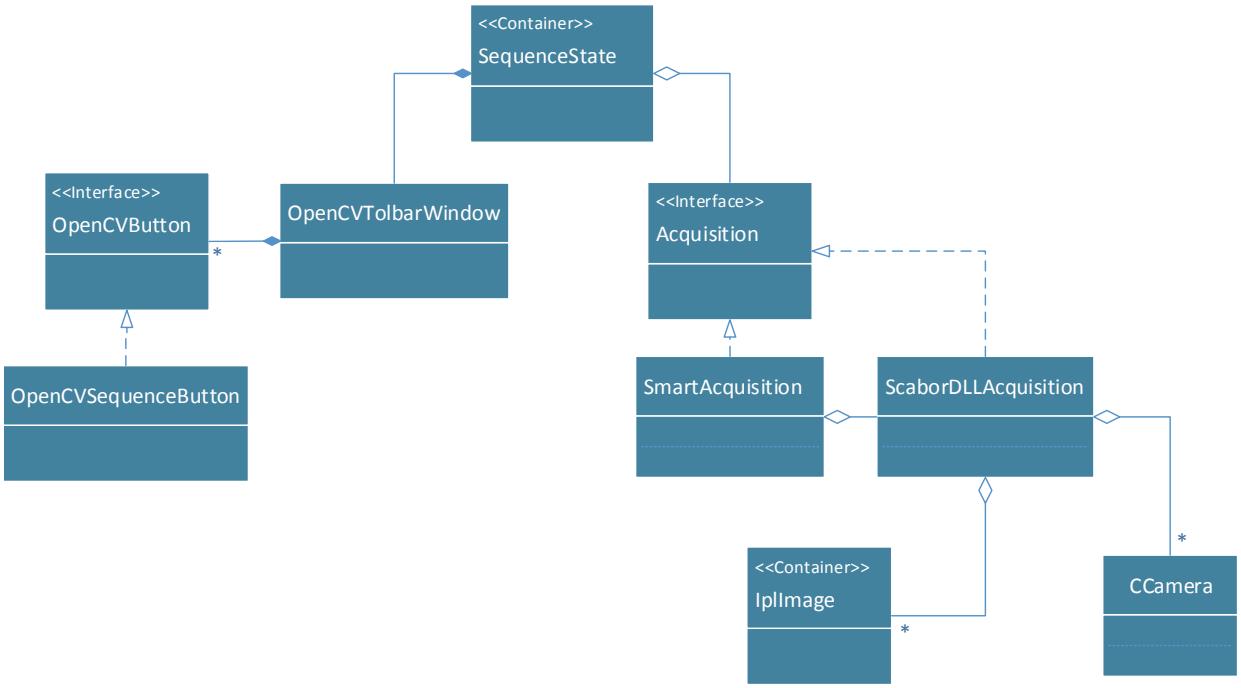

Fig. 3.3.6. Diagrama de clase pentru aplicația de achiziție date, pocesare si vizualizare/testare

Butoanele grafice sunt implementate in OpenCV fiind definite de interfața OpenCVButton și implementate în clasa OpenCVSequenceButton. Implementarea are în vedere afișajul grafic pe ecran, modificarea aspectului în funcție de acțiunea întreprinsă de utilizator precum și managementul evenimentelor specifice fiecărui buton cu funcții de tip *listener*. În firul principal de execuție se interoghează acțiunea întreprinsă la nivelul interfeței pentru a se acționa în consecință la nivelul modulelor de achiziție și a procesării datelor oferite de senzori.

Partea de achiziție este definită la nivelului interfeței Acquisition și implementată folosind o topologie pe mai multe nivele. Nivele definesc gradul de procesare și de informație cu care se lucrează. Astfel are loc o dezvoltare progresivă începând cu nivelul de bază unde se gestionează imagini și pixeli până la niveluri superioare la care se extrage și gestionează informația 3D sau alte informații primite de la vehicul.

La nivelul de bază clasa ScaborDLLAcquisition deține referințe către imaginile oferite de senzori și către camerele care furnizează aceste imagini. Folosim noțiunea de canal ca fiind asocierea dintre o cameră caracterizată de parametrii acesteia și imaginile care le furnizează. Din acest punct de vedere aplicația poate gestiona un număr predefinit de canale definite în funcție de numărul efectiv de senzori utilizați. De exemplu un stereohead cu 2 camere video poate furniza informații pe 8 canale: 2 pentru imaginile color (dacă este vorba de camere color), 2 pentru imaginile corespunzătoare grayscale, 2 pentru harta de profunzime pe regiunea de interes și 2 pentru imaginile rectificate și redimensionate la dimensiunea hărții profunzime. Un alt exemplu, o camera mono infraroșu se poate gestiona pe 4 canale, din care 1 pentru imaginea originală, 2 pentru harta de profunzime (dacă e cazul) și 1 pentru imaginea rectificată și redimensionată după cea de profunzime. Datele de la un sistem LiDAR se pot stoca pe 2 canale care reprezintă harta de profunzime. Motivul pentru care harta de

profunzime se folosesc 2 canale la îl reprezintă faptul că distanțele sunt reprezentate pe 16 biți, iar canalele utilizate au ca suport imagini pe 8 biți.

Informațiile de sistem (ex. CAN, timpi de achiziție etc.) se stochează la nivelul imaginii asociate primului canal care reprezintă camera din partea stângă a stereohead-ului cu vedere spre direcția de mers. Aceste informații se stochează peste pixelii din partea de jos – ocupând 1 sau 2 rânduri – fără a compromite datele relevante din imagine, având în vedere faptul că acestea se află de regulă în partea imediat superioară.

Interacțiunea dintre clasa ScaborDLLAcquisition și API-ul specific senzorului se face prin biblioteci de funcții incluse în librarii dinamice dll. Aceste librării dețin funcții care la inițializare pornesc fire de execuție paralelă cu firul principal și care au ca principal scop achiziția de date și sincronizarea lor pentru ca în final să fie puse la dispoziția aplicației.

Clasa de bază CCamera stochează parametrii specifici unei camere – intrinseci și extrinseci – și pune la dispoziție funcții de încărcare-salvare a acestora și de proiecție bidirecțională între planul imagine, sistemul de referință al camerei și cel al lumii.

La nivel superior, clasa SmartAcquisition extrage și interpretează informațiile de sistem stocate în imaginea primului canal și informațiile 3D din canalele rezervate imaginii de profunzime. În cazul lipsei informației 3D se pot aplica diverși algoritmi de reconstrucție 3D folosind imaginile stereo furnizate pe canale pereche. Specific acestei clase este faptul că dacă sunt prezente informațiile 3D de la un anumit senzor atunci se calculează automat și se pun la dispoziție hărțile de disparitate și de profunzime pentru toate canalele aplicației, chiar dacă senzorii asociați nu furnizau informație 3D. Această funcționalitate se realizează folosind operațiile puse la dispoziție de clasa CCamera și ușurează conversia informației de nivel înalt între canalele asociate diferiților senzori de percepție ai sistemului utilizat.

3.4. Diseminare rezultate (SO4.2-2)

S-a elaborat 1 articol de jurnal acceptat spre publicare in IEEE Transactions on Intelligent Transportation Systems ISI (IF 5.744 / 2018, clasifict in zona Q1/rosie), indexat/publicat in IEEE Xplore (Early Access) si in curs de indexare in WoS:

1. V. Miclea, S. Nedevschi, Real-Time Semantic Segmentation-Based Stereo Reconstruction, IEEE Transactions on Intelligent Transportation Systems (Early Access), pp. 1-11, 2019, DOI: 10.1109/TITS.2019.2913883

Diseminarea rezultatlor s-a materializat de asemenea prin prezentarea si publicarea a 6 lucrari in volumele unor conferinte internationale indexate sau care vor fi indexate BDI (IEEE Xplore Digital Library) si /sau in Clarivate Analytics Web of Science (fosta ISI Proceedings):

- 2019 IEEE Intelligent Vehicles Symposium (IV), 9-12 June 2019, Paris, France;
- 2019 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 5-7 Sept. 2019, Cluj-Napoca, Romania
- 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 27-30 Oct. 2019, Auckland, New Zeeland.
- 2019 2nd International Joint Conference on Computer Vision and Pattern Recognition (CCVPR), 22-24 Nov., 2019, Prague, Czech Republic

Lista detaliata a lucrarilor prezentate la aceste conferinte si publicate in volumele asociate este prezentata mai jos:

1. A. Petrovai, S. Nedevschi, Efficient Instance and Semantic Segmentation for Automated Driving, 2019 IEEE Intelligent Vehicles Symposium (IV), 9-12 June 2019, Paris, France, pp. 2575-2581, DOI: 10.1109/IVS.2019.8814177 (indexat BDI: IEEE Xplore | Sopus, in curs de indexare WoS)
2. R. Brehar, F. Vancea, T. Marita, C. Vancea, S. Nedevschi, Object Detection in Monocular Infrared Images Using Classification – Regresion Deep Learning Architectures, 2019 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 5-7 Sept. 2019, Cluj-Napoca, Romania, ISBN 978-1-7281-4914-1. (in curs de indexare BDI: IEEE Xplore | Scopus si WoS)
3. M.P. Muresan, S. Nedevschi, Multi-Object tracking of 3D cuboids using aggregated features, 2019 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 5-7 Sept. 2019, Cluj-Napoca, Romania, ISBN 978-1-7281-4914-1. (in curs de indexare BDI: IEEE Xplore | Scopus si WoS)
4. A. Petrovai, S. Nedevschi, Multi-task Network for Panoptic Segmentation in Automated Driving, 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 27-30 October, Auckland, New Zeeland, pp. p. 2394-2401, DOI: 10.1109/ITSC.2019.8917422 (indexat BDI: IEEE Xplore, in curs de indexare Scopus si WoS)
5. S.E.C. Deac, I. Giosan, S. Nedevschi, Curb detection in urban traffic scenarios using LiDARs point cloud and semantically segmented color images, 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 27-30 October, Auckland, New Zeeland, pp. 3433-3440, DOI: 10.1109/ITSC.2019.8917020. (indexat BDI: IEEE Xplore, in curs de indexare Scopus si WoS)
6. R. Brehar, T. Marița, M. Negru, S. Nedevschi, Pedestrian Identification in Infrared and Visible Images Based on Pose Keypoints Matching, 2019 2nd International Joint Conference on Computer Vision and Pattern Recognition (CCVPR 2019), Nov. 22-24, 2019, Prague, Czech Republic. (in curs de indexare BDI: Scopus si WoS)

3.5. Implementarea, testarea, evaluarea si optimizarea metodelor originale pentru alinierea si reprezentarea 3D spatio-temporala si bazata pe aparente a datelor multi-sensoriale si multi-spectrale primare (SO2.1-3, SO2.2-3)

3.5.1. Sincronizarea spatio-temporala a datelor senzoriale

Procedura de sincronizare temporala dintre imaginile de la senzorul infraroşu (LWIR – Long Wavelength Iinfra-Red) si senzorul de sterovizіune a fost detaliata in [Ned2018] cap. 2.1.1.a. Camera in infraroşu funcţionează la o frecvenţă de achiziţie constantă fără a oferi facilităţi de captură la cerere. Spre deosebire, sistemul stereo ofera imagini la cerere. Sincronizarea are loc prin corelarea cereri de captura a imaginiilor stereo cu camera LWIR.

Sincronizarea temporala a datelor achizitionate de senzorii LiDAR si a imaginilor video si infrarosu este implementata pe baza semnaturilor temporale (“time stamps”). Pentru a avea informatia cea mai corecta se pastreaza time stampul imaginii ca si referinta si se stocheaza “chunk-urile” furnizate de LiDAR. Setul de “Chunk-uri” acoperind 360 de grade cu diferenta temporala cea mai mica din cele achizitionate fata de referinta de achizitie a imaginilor se considera ca fiind cadrul corespunzator imaginii achizitionate. Cunoscand viteza si directia de deplasare (odometria) vehicolului propriu, datele provenite de la LiDAR se pot corecta pentru a corespunde momentului achizitiei imaginilor printr-o operatie de compensare a miscarii.

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

Pentru a elimina acest efect din datele senzorilor LiDAR (Velodyne Puck 16) s-a implementat un algoritm de compensare al mișcării/aliniere temporală, extinzând [Hon2010]: senzorii de odometrie (yaw rate si viteza) furnizează date privind ego-mișcarea vehicolului propriu, care ulterior sunt agregate pentru a calcula transformarea de mișcare efectuată de-a lungul achiziției unei scanări

complete (360 grade). Folosind inversa acestei transformări, precum si timpul exact de măsurare al fiecarui punct 3D din scanare, se calculeaza o transformare corectivă pentru ficare punct individual, eliminând distorsiunile create de mișcarea proprie.

Mai mult, algoritmul este 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 bună calitate, fără necesitatea implementări la nivel hardware al unui sistem de sincronizare între aceste dispozitive. In figura de mai jos sunt ilustrate rezultatele procesului de compensare a miscarii pentru LiDAR-ele cu 16 canale:

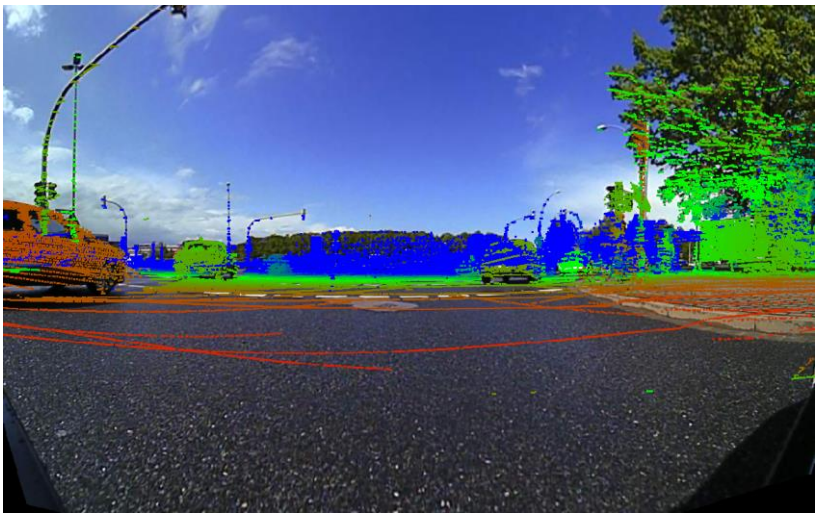

Fig. 3.5.1. Punctele 3D furnizate de LiDAR, corectate prin algoritmul de compensare a miscarii sunt proiectate pe imaginea din spectrul vizibil

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

Înregistrarea între imagini multispectrale VIS-LWIR este o sarcină foarte dificilă datorită relației neliniare dintre intensitățile pixelilor corespondenți: intensitatea pixelilor imaginii LWIR este proporțională cu temperatura obiectelor în timp ce intensitatea pixelilotr din VIS este dată de reflectanta sau culoarea obiectelor. Mai mult decât atât, imaginile LWIR prezintă un contrast mai mic, gradienți de intensitate mai mici și detalii(textură) mai slabe, în special datorită efectului de difuzie termică. Există două abordări principale pentru înregistrarea (alinierea) unor astfel de imagini: bazate pe sterioviziune si bazate pe viziune monoculară.

Dacă sunt disponibile informații de profunzime asociate imaginilor din VIS obtinute prin stereoviziune și parametrii extrinseci relativi dintre camera stânga a sistemului de stereoviziune și camera LWIR sunt cunoscuti prin calibrare, corespondența de pixeli între camera VIS stângă și camera LWIR poate fi calculată prin utilizarea matricei de proiecție [Ned2018] sau a teoriei tensorului trifocal [Che2019].

În cazul monocular (nu există informații de profunzime), o relație omografică între câmpiile de imagine VIS și LWIR poate fi stabilită doar în unele cazuri particulare care sunt rareori întâlnite în practică: rotație pură, toate punctele din scenă sunt coplanare sau sunt la infinit [Har2003]. Soluția în acest caz este să efectuam înregistrarea (alinierea) numai pe unele obiecte / patch-uri care pot fi aproximate ca suprafețe plane și apoi să corelam pixelii între aceste obiecte, calculând o transformare omografică, folosind câteva puncte de interes (de obicei colturi) detectate si potrivite in ambele imagini.

In acest context s-aau implementat 2 metode de mapare a regiunilor de interses dintre imaginile VIS-LWIR bazate pe puncte cheie (keyponts):

Maparea bazata pe puncte cheie de tip colţ si constrangeri epipolare

In abordarea prezentata in [Bae2019] se pleaca de la o regiune de interses din una dintre cele doua perechi de imagini (VIS/LWIR), se calculeaza puncte de interes (colturi), se calculeaza corespondentul acestora in celealta imagine folsindu-se, la alegere, 2 metrici de corelatie (Informatia Mutula (MI) sau Autosimilaritatea Locala (LSS)), intr-un spatiu de cautare redus, dat de dreapta epipolara (dedusa din parametrii extrinseci) si se estimeaza apoi o functie de mapare (omografie) H intre cele 2 patchuri de imagine corespondente folsoind metoda RANSAC pentru eliminarea corespondentelor false (outliers).

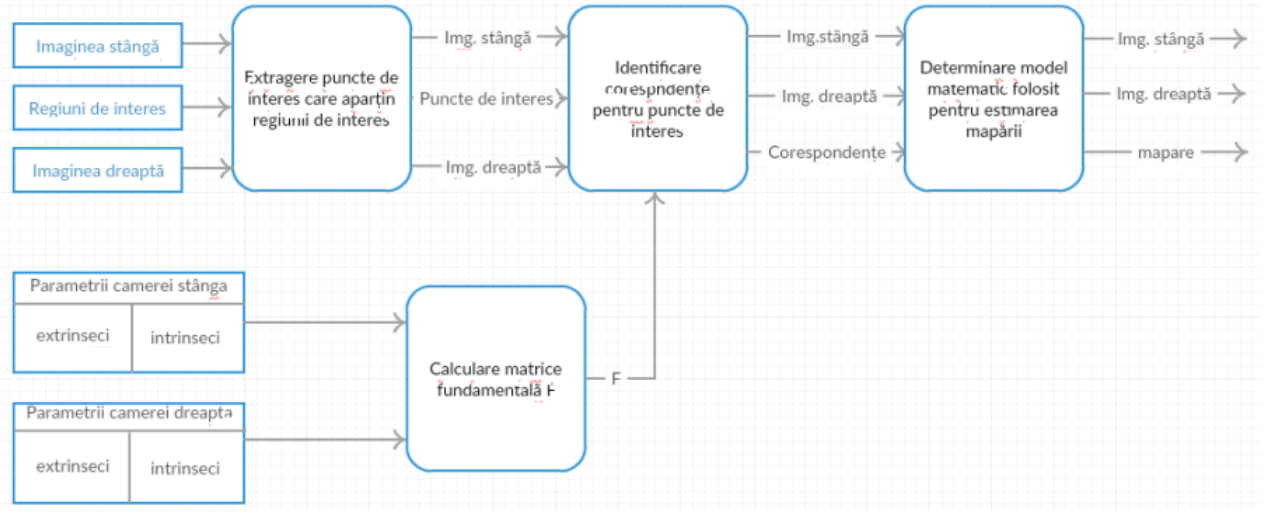

Fig. 3.5.2. Ilustrarea procesului de mapare a regiunilor de interes intre cele 2 imagini (VIS<->LWIR) unde imaginie stamga si dreapta sunt nume generice date celor doua tipuri de imagini multispectrale (VIS sau LWIR)

In figurile de mai jos se prezinta rezultatele experimentale ale procesului de mapare pentru regiuni de interes corespondente unor pietoni.

a) Corelarea unei regiuni de interes folosind sistemul bazat pe MI

b)Corelarea unei regiuni de interes folosind sistemul bazat pe LSS

Fig. 3.5.3. Ilustrari ale maparii de ROI VIS->LWIR folosind cele 2 tipuri de metrici de corealatie a punctelor de interes ($ROI_{R(LWIR)} = H^{*} \cdot ROI_{R(VIS)}$)

a)

b)

Fig. 3.5.4. Ilustrari ale maparii de puncte de interes LWIR->VIS folosind cele 2 tipuri de metrici de corealatie (a-MI, b-LSS) a punctelor de interes ($ROI_{R(VIS)} = H^{*} \cdot ROI_{L(LWIR)}$)

Nr. Imagine Stângă	Tip Imagine Stângă	Rezultate Stereocorelație	Rezultate MI	Rezultate LSS
161	VIS	416 90 40 85	414 91 44 74	418 87 44 82
161	LWIR	523 194 51 106	507 210 49 102	515 193 25 61
0	VIS	90 51 33 82	82 94 23 90	90 52 31 80
0	LWIR	160 120 54 141	156 136 59 132	158 122 59 149

Exemple numerice ale maparilor regiunilor de interes (bidirectionale) pe cele doua scene (161 si 0)sub forma coordonatelor x_top, y_top, width si height ale ROI

a b c

Fig. 3.5.5. Ilustrarea maparilor aferente randurilor 1(a), 2(b) si 3(c) din tabelul de mai sus.

Dupa cum se poate observa din rezultatele experimentale prezentate mai sus, maparea LWIR->VIS este mai putin precisa, datorita erorilor de detectie ale punctelor de interes initiale (in imaginile LWIR) ale metodei “GoodFeaturesto track” (detectorul de puncte de interes/colturi din OpenCV) care nu este optimizat pentru imaginile LWIR (cu contrast mai mic, gradienți de intensitate mai mici și detalii/textura mai putine).

Maparea bazata pe puncte cheie ale modelului uman de tip skeleton

Deorece detectia si maparea punctelor cheie de tip colt intre imaginile multispectrale este dificila datorita aparentei vizuale diferite a celor 2 tipuri de imagini, s-a propus o metoda alternativa de potrivire bazata pe puncte cheie ale modelului scheletal al obiectelor de tip pieton [Bre2019b].

Pentru fiecare pereche de imagini (color si LWIR) se aplica o metoda de detectie a ipotezelor corespunzatoare pietonilor bazata pe arhitecura YOLO [Red2018]. Pentru fiecare instanta de pieton se aplica o medoda de estimare a modelulul scheletal [Xiu2019] care combina rezultatele a doua retele neuronale convolutionale: STN (Spatial Transformer Network) si SPPE (Single Person Pose Estimation), retinandu-se 12 puncte cheie care corespund articulatiilor umerilor, coatelor, incheieturilor mainilor, soldurilor, genunchilor, calcaielor si picioarelor si a unui punct suplimentar corespunzator mijlocului segmentului care uneste umerii.

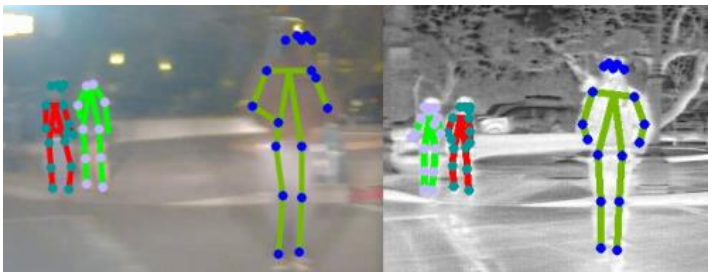

Fig. 3.5.6. Modelele scheletale obtinute prin metoda [Xiu2019] pe imaginile color (stanga) si LWIR (dreapta) ale aceleiasi scene

Pentru realizarea corespondentei dintre modelele scheletale corespunzatoare aceleiasi persoane din perechile de imagini multimodale s-a propus un algoritm original de potrivire bazat pe metoda vecinului celui mai apropiat aplicata intr-un spatiu euclidian normalizat:

Se extrage regiunea de interes rectangulara care circumscrie punctele cheie selectate/detectate avand parametrii: (x_t , y_t w, h)

Pentru ca potrivirea sa fie invarianta la scala obiectelor se aplica urmatoarea normalizare: $x_n = x/width$; $y_n = y/height$

Vectorul de trasaturi asociat punctelor cheie selectate in spatiul de trasaturi normalizat va fi: $K^m = \{(x^m_1, y^m_1), (x^m_2, y^m_2) ...(x^m_{13}, y^m_{13}), width, height\}$ unde m reprezinta modalitatea (color/LWIR)

Se limiteaza spatiul de cautare al potrivirilor punctelor chie la distante mai mici decat o valoare de prag (in spatiul normalizat) pentru a evita potrivirile false (potrivirile in cele doua imagini multimodale trebuie sa fie in regiuni apropiate spatial)

Corespondentele multimodale intre vectorii de trasaturi ale modelelor scheletale care apartin aceluiasi pieton sa detemina prin minimizarea distantei euclidiene dintre doua instante multimodale ale vectorilor de trasaturi: $|| K^{color} - K^{LWIR}||$

Validarea metodelor s-a realizat pe setul de date FLIR-ADAS [Fli2018] care contine imagini multimodale adnotate in spatiul LWIR. Adnotarile au fost extinse si pentru imaginile color iar fiecarei perosane din setul de date i s-a atribuit un identificator unic. Adnotarile au fost impartite in 3 clase: pietoni mari (cu inaltimea mai mare dacat ½ din inaltimea imaginilor, pietoni medii cu inaltimea intre ¼ si ½ din inaltimea imaginii si pietoni mici cu inaltimea mai mica decat ½ din inaltimea imaginii. In tabelul de mai jos, in primele 2 coloane se prezinta acuratetea medie a metodei de detcetie a pietonilor (YOLO-v3) pe imaginile LWIR si color. O detectie este considerata corecta daca metrica IoU (Intersection over Union) intre ipoteza detectata si cea de referinta depaseste 50% si cel putin 80% din corpul pietonului este vizibil (rata de ocluzie este mai mica de 20%). Acurateetea de corespondenta obtinuta a variat intre 67% pentru pietonii de dimensiune mica de pietoni si 76% pentru pietonii de dimensiune medie.

3.5.3. Segmentarea canalelor senzoriale

Segmentarea semantica a imaginilor

Segmentarea semantica a imaginilor este realizata prin implementarea arhitecturii propuse in [Cost2018]. Arhitectura are la baza o retea unificata ResNet&FPN (Feature Pyramid Network) si este flolosita pentru 3 task-uri: detectia si clasificarea de obiecte (“head-ul” Faster-RCNN), segmentarea la nivel de instante (“head-ul” Mask-RCNN), segmentarea semantica la nivel de pixeli (“head-ul” de segmentare semantica imbunatatit cu Atrous Spatial Pyramid). Extensia arhitecurii cu “head-ul” de segmentare semantica ASP 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.

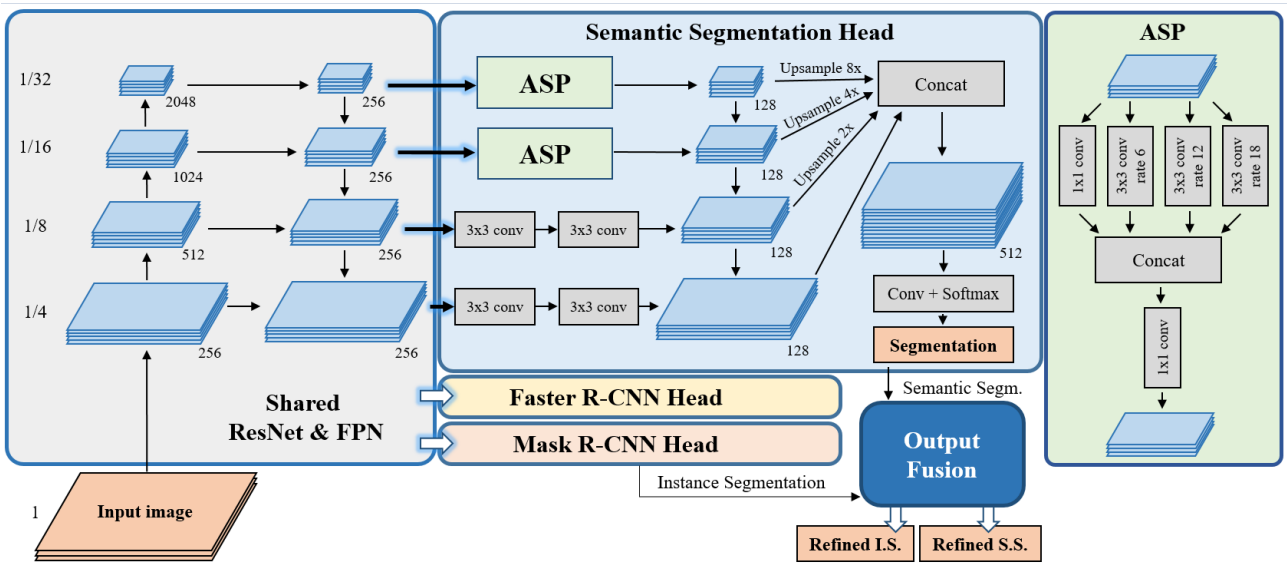

Fig. 3.5.7. Retea neuronala pentru detectie, segmentare in instante si segmentare semantica. Reteaua comuna calculeaza trasaturile de baza folosind o retea convolutionala reziduala, ResNet si o extensie cu piramida de trasaturi (FPN). Reteau este extinsa cu 3 retele de predictie: detectie (Faster R-CNN), segmentare in instante (Mask R-CNN) si segmentare semantica (reteaua propusa).

Pentru rafinarea iesirilor furnizate de segmentarea semantica si segmentarea de instante a fost implementata o metoda de fuziune bazata pe impartirea pixelilor in 2 clase: infrastructura (fundal) si utilizatori ai drumului (obiect) pe baza segmentarii semnatice, urmata de o euristica de fuziune a segmentarilor semantice si de instanta obtinandu-se una din primele implementari ale segmentare panoptice.

Metoda a fost evaluata pe setul de date CityScapes (19 clase semantice + 8 clase pentru instante) constand din 5000 de imagini de trafic cu o rezoutie de 2048x1024 adnotate la nivel de pixel. Evaluarea segmentarii semantice a fost realizata pe baza valorii medii a metricii Intersection-over-Union (IoU) iar a segmentarii de instante pe baza preciziei medii (AP).

Method	backbone	AP mask	mIoU
Mask-RCNN	ResNet50	36.4	-
PSPNet	ResNet50-dilated	-	71.7
Unified baseline	ResNet50-FPN	37.0	71.6
+ ASP and RP	ResNet50-FPN	37.2	72.9
+ fusion	ResNet50-FPN	37.3	76.0

Rezultatele segmentarii la nivel de instanta si semnatice pe setul de validare Cityscapes cu o rezolutie de 2048x1024 pixeli

Method	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorbike	bike	mIoU
ResNet50-FPN+ASP	97.7	82.3	91.2	48.6	51.3	56.9	66.9	73.1	91.5	61.8	93.1	80.1	59.8	93.1	63.0	77.5	64.1	59.7	75.3	72.9
ResNet50-FPN+ASP+fusion	97.7	82.3	91.2	48.6	51.3	56.9	66.9	73.1	91.5	61.8	93.1	**81.0**	**65.6**	**94.0**	**81.6**	**89.8**	**80.1**	**61.9**	**75.6**	76.0

Rezultatele segmentarii semnatice pe setul de validare Cityscapes fara/cu schema de fuziune

Din al doile tabel se poate observa ca rezultatele segmentarii semnatice pentru clasele de utilizatori ai drumului (cele din dreapta) sunt imbunatatite datorita unei clasificari mai robuste la nivel de obiect si datorita folosirii unei etichete unice pe instant rezultand o crestere de la 71.6 la 76 a parametrului mIoU.

Pentru setul de date propriu s-a obtinut urmatoarele rezultate:

Solutia de procesare si rezolutia de antrenare	AP mask	mIoU	Timp de executie pe GPU Nvidia GTX 1080Ti
ResNet50-FPN 2048x1024	36.4	72.9	178ms
ResNet50-FPN 1924x512	28.6	65.5	65ms

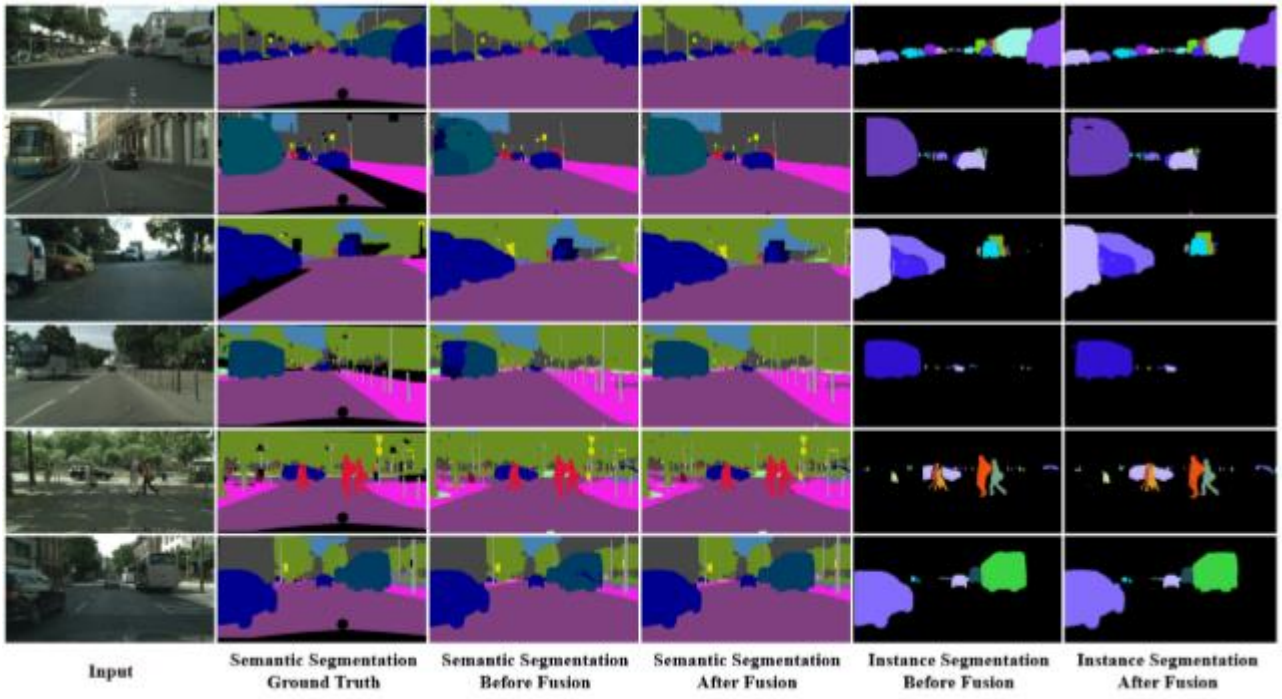

Fig. 3.5.8. Rezultatele segementarii semantice si a segmentarii la nivel de instant rafinata prin fuziune

In [Mic2019] este prezentata o abordare in care utilizarea segmentarii semnatice a imginilor poate imbunatati calitatea reconstructiei dense prin steroviziune. Solutia propusa urmeaza pipeline-ul de executie al steroviziunii clasice, fiind imbunatatit prin informatia aditionala furnizata de segmentarea semantica (obtinuta cu ajutorul unei retele ERFNet adaptate). Imbunatatirile sunt aduse la nivelul calculelor, agregarii si optimizarii prin adaptarea tehnicilor existente astfel incat sa integreze informatia semantica a fiecarei clase. Pentru pasii de calcul ai functiei de cost si optimizare

se propune utilizarea algoritmilor genetici care pot ajusta paremetrii in mod incremental pentru optimizarea solutiei. Se propune de asemenea o tehnica se filtrare post-procesare sensitiva la muchii bazata pe o arhitectura CNN pentru rafinarea disparitatilor. Metoda implementata are un timp de rulare/procesare de 30 fps pe un GPU conventional. O ilustrare comparativa a erorilor medii de reconctructie si a timpilor de procesare relativ la alte metode, aplicate pe setul de date KITTI 2015 este prezentata in tabelul de mai jos:

Method	D1-all	D1-bg	D1-fg	Speed (ms)	Platform	CNN
DispNetC [31]	4.05%	4.11%	3.72%	60	Nvidia GTX Titan X (Caffe)	Yes
DeepCostAggr [26]	5.61%	4.82%	10.11%	30	GPU @ 2.5 Ghz (C/C++)	Yes
AABM [12]	6.26%	4.49%	15.22%	80	1 core @ 3.0 Ghz (C/C++)	No
SNCC [11]	6.69%	5.00%	15.21%	80	1 core @ 3.0 Ghz (C/C++)	No
CSCT+SGM+MF [17]	6.56%	5.37%	12.58%	6.4	Nvidia GTX Titan X (CUDA)	No
PCOF + ACTF [10]	8.03%	5.98%	18.40%	80	1 core @ 3.0 Ghz (C/C++)	No
Proposed	**5.67%**	**4.83%**	**10.75%**	**34 (17+17)**	**Nvidia GTX 1080 (CUDA)**	**Yes**

Segmentarea spatiului 3D pentru detectia obiectelor din scena

Pentru detecția obiectelor din scena 3D s-a folosit o metodă bazată pe învățare profundă. Această metodă se numește Frustum PointNets [Qi2018], care este o retea neuronala capabila sa detecteze obiecte 3D sub forma de cuboide, plecand de la detectii 2D si un nor de puncte 3D. Acerastă metodă se bazează pe PointNet [Qi2017a], o rețea neuronala creată pentru a procesa scene 3D. Rețeaua Frustum PointNets este formată din trei componente.

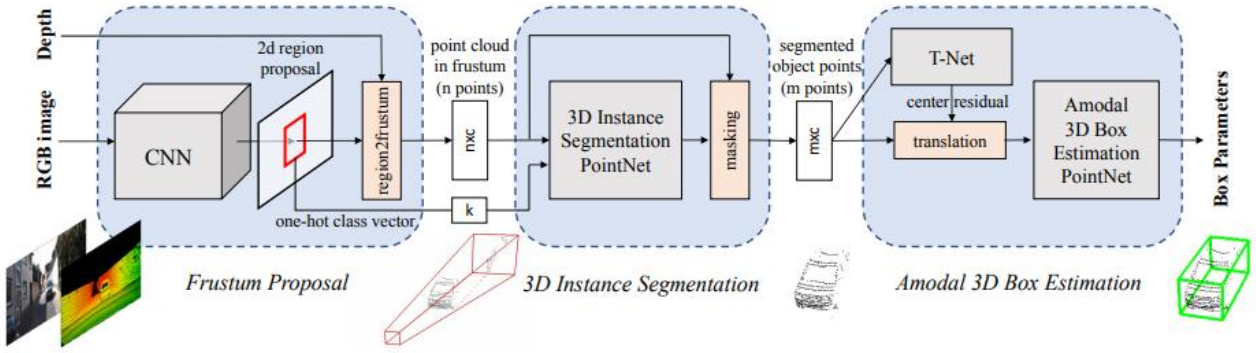

Fig. 3.5.9. Arhitectura retelei Frustrum PointNet [Qi2018]

Prima componentă (Frustrum Proposal) are rolul de a propune trunchiuri de piramidă (frustum-uri), și de a colecta toate punctele 3D din fiecare trunchi. Aceasta componentă folosește rezultatul unui detector de obiecte 2D. Toate punctele 3D sunt proiectate în imagine, și pentru fiecare detecție 2D se selectează punctele care sunt proiectate în interiorul acelei detecții. Pentru ca algoritmul să fie invariant la rotație, toate punctele dintr-un trunchi de piramidă sunt rotite astfel încât axa centrală a trunchiului să fie ortogonală cu planul imaginii.

A doua componentă (3D Instance Segmentation) are ca scop segmentarea 3D a punctelor dintr-un trunchi de piramidă. Această segmentare păstrează doar punctele care fac parte din obiect, eliminând punctele de fundal. Pentru a îmbunătății invarianța la translație, norul de puncte segmentat se translateaza astfel încât media aritmetică a coordonatelor norului sa fie egală cu originea.

A treia componentă are ca scop estimarea unor cuboide 3D de incadrare pentru fiecare obiect. În acestă componentă sunt folosite două rețele de tip PointNet: *T-Net* și *Amodal 3D Box Estimation PointNet*. T-Net are ca scop estimarea centrului adevărat al unui obiect detectat, deoarece este posibil ca centrul estimat în componenta precedentă să nu fie adevăratul centru al obiectului, și translatarea punctelor astfel încat centrul nou estimat să fie originea. Rețeaua Amodal 3D Box Estimation

PointNet are ca scop estimarea unui nou centru pentru obiect, dimensiunea lui (înălțime, lățime și lungime) și orientarea obiectului.

În final, centrul unui obiect se reconstruieste folosind ecuația:

$$C\_pred = C\_seg + \Delta C\_tnet + \Delta C\_box\_net$$

unde:

C_seg este media coordonatelor punctelor segmentate de primul modul,
ΔC_tnet este centrul estimat de *T-Net*
ΔC_box_net este centrul estimat de *Amodal 3D Box Estimation PointNet.*

Există două versiuni ale rețelei FrustumPointNets, versiunea 1 fiind bazată pe PointNet [Qi2017a], si versiunea 2 pe PointNet++[Qi2017b]. In figura de mai jos sunt prezentate rezultatele experimentale ale detecției folosind un model de tipul FrustumPointNets versiunea 1, obtinut prin antrenare pe imagini si seturi de puncte 3D furnizate de senzori LIDAR din setul KITTI [Gei2012].

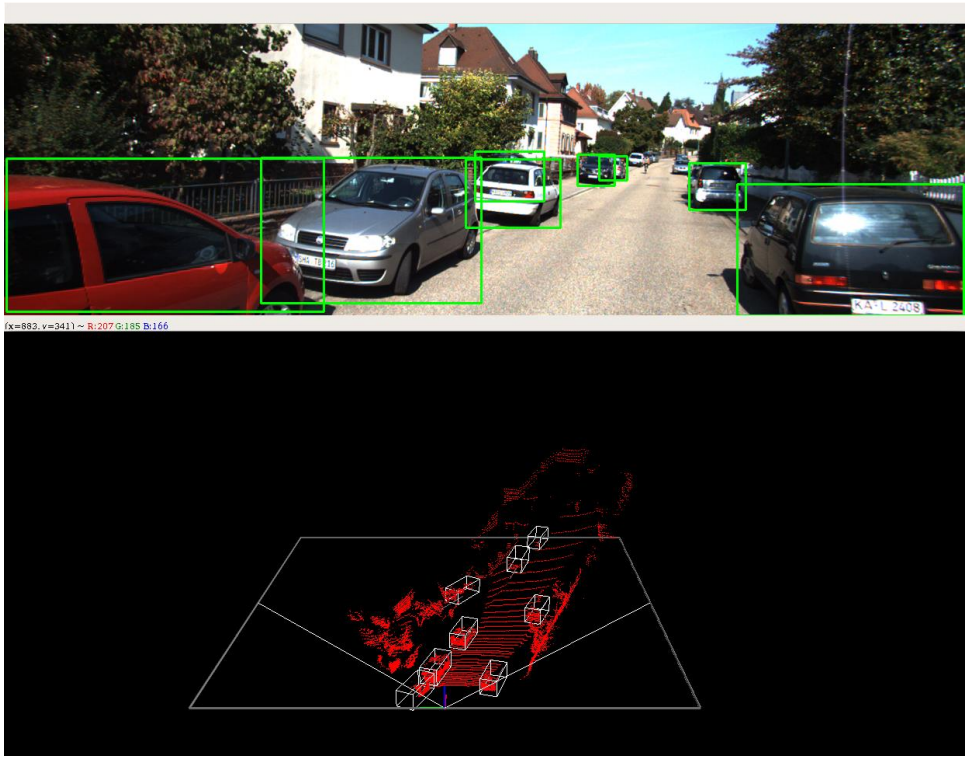

Fig. 3.5.10. Rezultatele segmentarii spatiului 3D pentru detectia obiectelor din scena reprezentate sub forma de cuboide pe setul de date KITTI [Gei2012]

Modelele preantrenate pe setul de date KITTI genereaza erori notabile pe setul de date propriu. Pentru imbunatatirea rezultatelor, modelul retelei FrustumPointNets v1 a fost rafinat pe o secventa proprie de 1610 cadre. Pentru generearea datelor de antrenare, ambele modele preantrenate FrustumPointNets v1 si v2 au fost utilizate pentru generarea de cuboide 3D dintre care cele mai bune cubioide 3D au fost selectate manual.

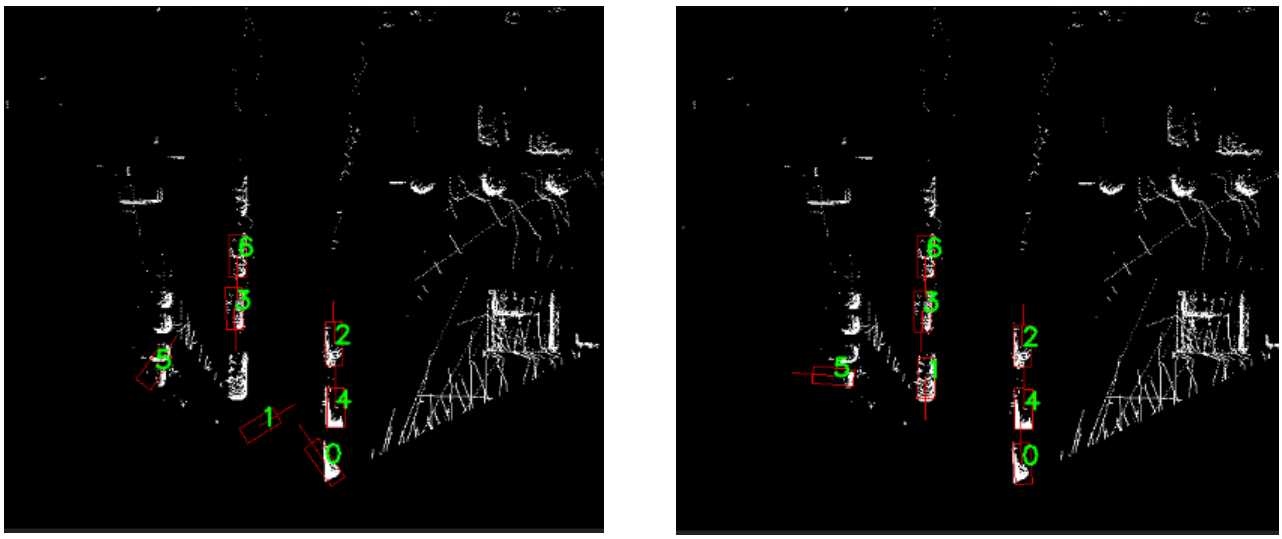

Fig. 3.5.11. Rezultatele detectiei obiectelor (proiectia top-view) inainte de rafinare (stanga) si dupa rafinare (dreapta) pe setul de date propriu

In figura de mai jos este prezentat rezultatul comparativ al metodei de detcetie a obiectelor 2D folosind Yolo-v3 [Red2018] antrenata pe setul CityScapes (stanga sus) vs. proiectia pe imagine a cuboidelor 3D detectate cu modelul rafinat FrustumPointNets v1 antrenat pe setul de date propriu. In imaginea de jos se prezinta proiectia top-view a cuboidelor 3D detectate cu modelul rafinat.

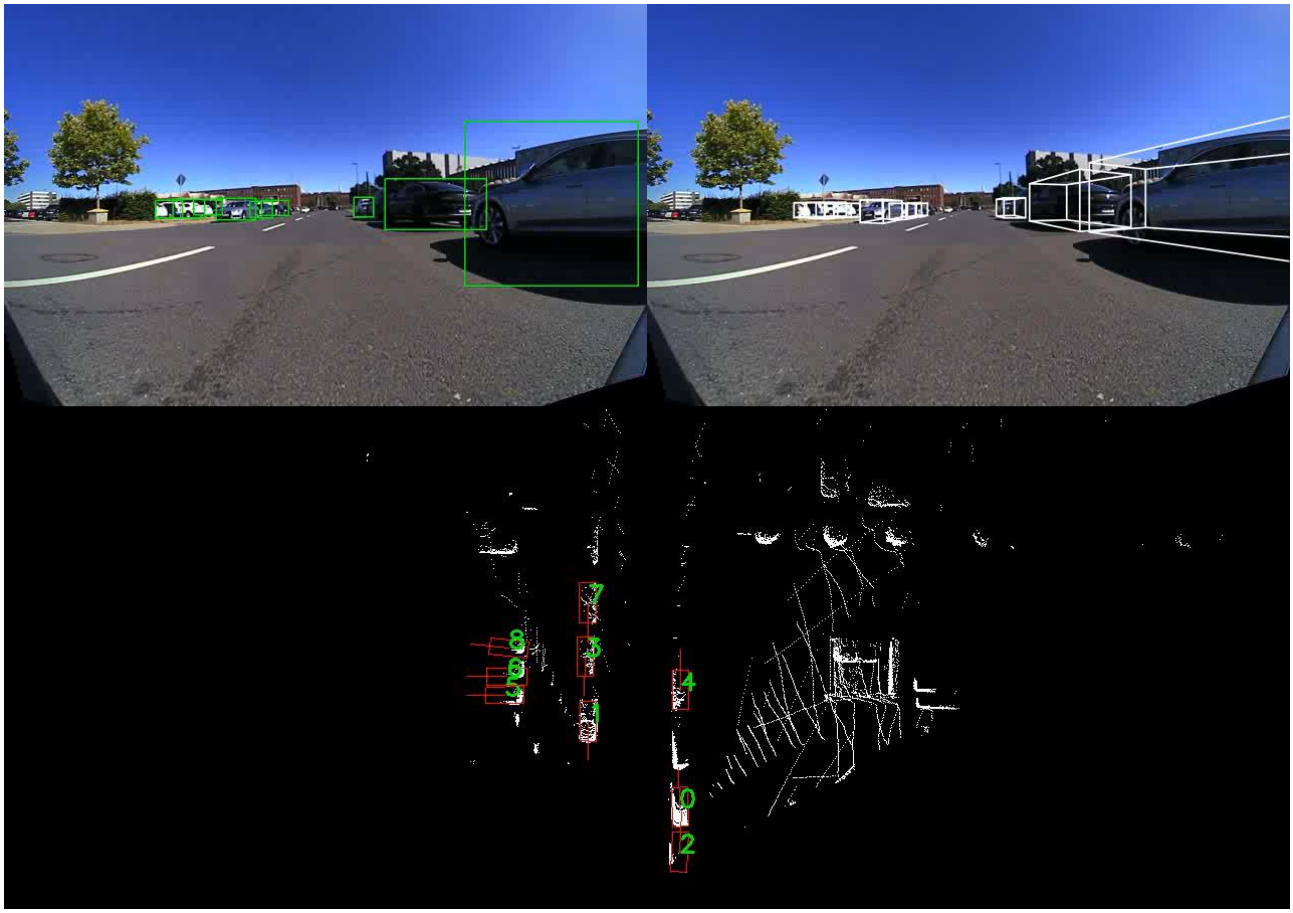

Fig. 3.5.12. Rezultatele detectiei obiectelor: YOLO-v3 (stanga sus) vs. FrustumPointNets v1 rafinat (dreapta sus si mijloc-jos) pe setul de date propriu

3.5.4. Concluzii

S-a propus si implementat un model de fuziune (low-lewel) a datelor senzoriale primare obtinut prin prin alinierea spatio-temporal si segmentarea acestora (vezi modelul STMAR propus: [Ned2018] - fig. 2.1.4). Aceasta fuziune se poate realiza prin 3 modalitati distincte in functie de disponibilitatea datelor senzoriale:

- Daca este diponibila harta densa de diparitati sau puncte 3D furnizata de senzorul de steroviziune, este posibila fuziunea (corespondenta) multispectrala densa dintre punctele 2D din imaginile din spectrul vizibil (camera stanga a sistemului de steroviziune) si a punctelor 2D din imaginea termala (LWIR) prin intermediul maparii intemdiate de matricea de proiectie ([Ned2018] - cap. 2.1.2.a) dedusa din parametrii de calibrare al sistemului senzorial vizual.
- In lipsa informatiei 3D furnizata de senzorul de steroviziune sunt disponibile urmatoarele alternative:
  - maparea la nivel de obiecte 2D (patchuri de imagine considerate planare) intermediata de determinarea unor corespondente intre puncte cheie apartinand acestor obiecte/patchuri
  - daca este disponibil un nor de puncte 3D furnizat de senzorii LiDAR se poate realiza maparea intre detectiile 2D de obiecte din imagini si punctele 3D asociate din norul de puncte furnizat de LiDAR prin aplicarea metodei Frustum PointNets. Avand corepondenta obiecte/puncte 3D <-> puncte/obiecte 2D pentru unul dintre canalele spectrale de imagine (ex. vixibil), maparea/fuziunea cu canalul spectral alternativ (ex. LWIR) se face prin intermediul matricei de proiectie asociate senzorilor LiDAR aplicata norului de puncte segmentat ([Ned2018]- cap. 2.1.2.b).

Referinte bibliografice

[Bar2019] Barbu Florin-Alexandru, Unelte pentru înregistrarea imaginilor multispectrale, Lucrare de licenta, Departamentul de Calculatoare, Universitatea Tehnica din Cluj-Napoca, iulie 2019.

[Bou2015] J.Y. Bouguet, Camera Calibration Toolbox for Matlab, Computational Vision Group, California Institute of Technology, Pasadena, California, http://www.vision.caltech.edu/bouguetj/calib_doc/index.html (ver. 2015)

[Bre2019a] R. Brehar, F. Vancea, T. Marita, C. Vancea, S. Nedevschi, Object Detection in Monocular Infrared Images Using Classification Regresion Deep Learning Architectures, 2019 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), 5-7 Sept. 2019, Cluj-Napoca, Romania

[Bre2019b] R. Brehar, T. Mariţa, M. Negru, S. Nedevschi, Pedestrian Identification in Infrared and Visible Images Based on Pose Keypoints Matching, 2019 2nd International Joint Conference on Computer Vision and Pattern Recognition (CCVPR 2019), Nov. 22-24, 2019, Prague, Czech Republic

[Che2019] Chen, Z., & Huang, X. Pedestrian Detection for Autonomous Vehicle Using Multi-Spectral Cameras. *IEEE Transactions on Intelligent Vehicles*, *4*(2), 211-219, 2019

[Cor2016] M. Cordts et al.,The cityscapes dataset for semantic urban scene understanding. In CVPR, 2016.

[Cost2018] 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

[Dea2019] S.E.C. Deac, I. Giosan, S. Nedevschi, Curb detection in urban traffic scenarios using LiDARs point cloud and semantically segmented color images, 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 27-30 October, Auckland, New Zeeland, pp. 3433-3440.

[Fli2017] FLIR, Pathfinder IR - User manual, (citat 2017): http://www.flir.com/uploadedFiles/UserManual_PathFindIR.pdf
[Fli2018] FLIR. Flir thermal datasets for algorithm training. https://www.flir.com/oem/adas/dataset/.
[Li2016] Y. Li, Y. Zhang, A. Geng, L. Cao, and J. Chen, "Infrared image enhancement based on atmospheric scattering model and histogram equalization," Optics and Laser Technology, vol. 83, pp. 99 – 107, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030399216000268
[Gei2012] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite, 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3354-3361.
[Har2003] R Hartley, A Zisserman, Multiple View Geometry in Computer Vision, 2-nd ed.,Cambridge University Press, 2003.
[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.
[Mar2006] T. Marita**,** F. Oniga, S. Nedevschi, T. Graf, R. Schmidt, Camera Calibration Method for Far Range Stereovision Sensors Used in Vehicles, Proceedings of IEEE Intelligent Vehicles Symposium, *(*IV2006), June 13-15, 2006, Tokyo, Japan, p. 356-363,
[Mic2019] V. Miclea, S. Nedevschi, Real-Time Semantic Segmentation-Based Stereo Reconstruction, IEEE Transactions on Intelligent Transportation Systems (Early Access), pp. 1-11, 2019, DOI: 10.1109/TITS.2019.2913883
[Ned2008] S. Nedevschi, R. Danescu, T. Marita, F. Oniga, C. Pocol, S. Bota and C. Vancea, A Sensor for Urban Driving Assistance Systems Based on Dense Stereovision, book chapter in Stereo Vision editor A. Bhatti, published by InTech Education and Publishing, Vienna, 2008, pages 235-272, ISBN: 978-953-7619-22-0. (download)
[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 privind implementarea proiectului in perioada: iulie – decembrie 2017 (etapa 1), decembrie 2017.
[Ned2018] S. Nedevschi, et al. "Perceptia multispectrala a mediului prin fuziunea datelor senzoriale 2D si 3D din spectrul vizibil si infra-rosu (MULTISPECT)" – Raport stintific Raport stiintific privind implementarea proiectului in perioada: ianuarie – decembrie 2018 (etapa 2), decembrie 2018.
[Pet2019] A. Petrovai, S. Nedevschi, Efficient Instance and Semantic Segmentation for Automated Driving, 2019 IEEE Intelligent Vehicles Symposium (IV), 9-12 June 2019, Paris, France, pp. 2575-2581.
[Qi2017a] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 652-660.
[Qi2017b] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space, Advances in neural information processing systems, 2017, pp. 5099-5108, arXiv preprint arXiv:1706.02413, 2017.
[Qi2018] C. R. Qi, Liu Wei, Wu Chenxia, Su Hao, Guibas Leonidas J (2017). Frustum PointNets for 3D Object Detection from RGB-D Data, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 918-927, arXiv preprint arXiv:1711.08488.
[Red 2018] J. Redmon, A. Farhadi. Yolov3: An incremental improvement, arXiv:1804.02767, 2018.
[Rom2018] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation, IEEE Transactions on Intelligent Transportation Systems, 19(1):263–272, 2018.
[Xiu2019] Xiu, Y., Li, J., Wang, H., Fang, Y., & Lu, C., Pose Flow: Efficient Online Pose Tracking. In British Machine Vision Conference, arXiv:1802.00977, 2018