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

# Raportul științific aferent perioadei 2017 – septembrie 2019

Director proiect:

Prof. dr. ing. Sergiu Nedevschi

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 1 / 2017

### **Director proiect:**

Prof. dr. ing. Sergiu Nedevschi **Colectiv:** Conf. dr. ing. Tiberiu Marita, Sef lucrari dr. ing. Raluca Brehar Asistent dr. ing. Robert Varga Doctorand ing. Arthur Costea Doctorand ing. Mircea Muresan Masterand ing. Zelia Blaga Masterand ing. Horatiu Florea Masterand ing. Selma Goga

Cluj-Napoca, Decembrie 2017

# **Raport stiintific**

privind implementarea proiectului in perioada: iulie - decembrie 2017

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

**Denumire etapa:** Studiu comparativ teoretic si experimental al metodelor de ultima ora existente in perceptia multispectrala si multisenzoriala a mediului; specificatii; calibrarea parametrilor extrinseci relativi ai senzorilor

### Activitatile etapei 2017:

- A.1.1. Calibrarea parametrilor extrinseci relativi ai senzorilor: studiul metodelor existente, specificarea modelelor matematice, dezvoltarea si implementarea de metode originale (SO1.1, SO1.2)
- A.1.2. Studiul metodelor de de aliniere si reprezentarea 3D spatio-temporala si bazata pe aparente a datelor multi-sensoriale si multi-spectrale primare (SO2.1-1, SO2.2-1). Elaborare specificatii
- A.1.3. Studiul metodelor de detectie si clasificare a obstacolelor si de reprezentare si fuziune a datelor senzoriale de nivel inalt (SO3.1-1, SO3.2-1, SO3.3-1). Elaborare specificatii.

### **Cuprins:**

1.1. Calibrarea parametrilor extrinseci relativi ai senzorilor: studiul metodelor existente, specificarea modelelor matematice, dezvoltarea si implementarea de metode originale (SO1.1, SO1.2)

1.1.1. Modelul sensorial propus

- 1.1.2. Calibrarea parametrilor extrinseci ai camerei FIR
- 1.1.3. Calibrare parametrilor extrinseci ai senzorului LiDAR

1.2. Studiul metodelor de de aliniere si reprezentarea 3D spatio-temporala si bazata pe aparente a datelor multi-sensoriale si multi-spectrale primare (SO2.1-1, SO2.2-1). Elaborare specificatii

1.2.1. Alinierea spatiala dintre imaginile din spectru vizibil si infra-rosu indepartat (FIR)

1.2.2. Alinierea dintre imaginile 2D si punctele 3D pentru construirea unei reprezentari spatio-temporale si bazata pe aparente a datelor sensoriale primare

1.3. Studiul metodelor de detecție si clasificare a obstacolelor si de reprezentare si fuziune a datelor senzoriale de nivel inalt (SO3.1-1, SO3.2-1, SO3.3-1). Elaborare specificatii

1.3.1. Definirea regiunilor de interes

1.3.2. Construirea unei colectii adnotate de imagini si reprezentarea obiectelor relevante

1.3.3. Detectia si clasificarea multi-senzoriala redundanta si fuziunea datelor la nivel de obiect Referinte bibliografice

Lista lucrarilor publicate

#### Indeplinirea activitatilor prevazute in proiect

# **1.1.** Calibrarea parametrilor extrinseci relativi ai senzorilor: studiul metodelor existente, specificarea modelelor matematice, dezvoltarea si implementarea de metode originale (SO1.1, SO1.2)

Scopul obiectivului (O1) este da estima cu precizie ridicata parametrii extrinseci ai fiecarui senzor reativ la un sistem de coordinate unic al lumii (care coincide cu sistemul de coordinate al vehiculului propriu).

Sistemul de steroviziune va fi folosit ca si sistem de referinta, metoda lui de calibrare de inalta precizie si distanta mare pentru scenarii specific automotive fiind deja prezentata de autori in [Mar06], [Mar07]. Metoda se bazeaza pe detectia in planul imagine (cu precizie de sub-pixel) a unor puncte de control corespunzatoare unor tinte 3D amplasate si masurate in scena de calibrare si minimzarea erori de proiectie a acestor puncte pe planul imagine. Metoda perimite estimarea pozitiei camerelor cu erori de ordinum mm, a orientarii cu erori de ordinal sutimilor de grad, necesare pentru un process de stereo-corelatie correct. Parametri calculati permit o reconstructia densa a scenei cu o erori relative mai mici de 5%. Desi procesului de construire a scenei de calibrare este laborios, datorita montajului rigid al camerelor de steroviziune relativ la sasiul autovehiculului, metoda trebuie aplicata doar la montajul camerelor.

In ceea ce priveste estimarea parametrilor extrinseci ai celorlalti senzori (camera FIR (Far Infra Red) si senzorul LIDAR) se propune dezvoltarea unor metode semiautomate care sa simplifice procedura de calibrare si sa permita recalibrarea usoara a fiecarui senzor la montari/demontari repetate (FIR si LIDAR) respetiv sa monitorizeze si sa corectere on-line acuratetea parametrilor (LIDAR).

#### 1.1.1. Modelul sensorial propus

Calibrarea parametrilor extrinseci ai senzorilor presupune estimarea pozitiei si orientarii relative a senzorilor fata de sistemul de coordinate unic al lumii.

Parametrii extrinseci sunt necesari pentru a transforma coordonatele unui punct 3D din sistemul de coordonate al lumii în sistemul de coordonate al senzorului și invers. Acești parametrii sunt:

- vectorul de translație  $\mathbf{T} = [Tx, Ty, Tz]^{\mathrm{T}};$
- vectorul de rotație  $\mathbf{r} = [Rx, Ry, Rz]^{T}$  sau matricea de rotație echivalentă  $\mathbf{R}$ .

Pentru sistemul propus, sistemul de coordinate al lumii este cel asociat autovehiculului propriu (fig. 1). Pozitia si orientarea fiecarui senzor (notat generic cu S) relativa la acest sistem va fi determinata de un vector de translatie  $Ts^W$ , si o matrice de rotatie  $Rs^W$ .

Pentru a putea exprima un punct arbitrar P (având coordonate  $XX_W = [X_W, Y_W, Z_W]^T$  în sistemul de coordonate al lumii) în sistemul de coordonate al senzorului  $XX_S = [X, Y, Z]^T$  avem nevoie de o transformare constând dintr-o rotație urmată de o translație. Această transformare poate fi codificată prin următoarea ecuație matriceală:

$$XX_{s} = R_{W}^{s} \cdot XX_{w} + T_{W}^{s}$$

$$\tag{1.1}$$

unde:

 $T_W^S$  – vectorul de translație din sistemul lumii în sistemul senzorului;

 $R_W^S$  – matricea de rotație din sistemul lumii în sistemul senzorului:

$$\mathbf{R}_{W}^{s} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$
(1.2)

Rotația poate fi exprimată prin cele trei unghiuri Euler  $R_x$  (pitch),  $R_Y$  (yaw) și  $R_z$  (tilt / roll), care definesc o secvență de trei rotații elementare ale sistemului de coordonate al lumii în jurul axelor x, y și respectiv z. Rotațiile sunt realizate mai întâi în jurul axei x, apoi în jurul axei y care deja a fost odată rotită și în final în jurul axei z care a fost rotită de două ori în precedentele două etape. Semnul unghiurilor de rotație este stabilit după regula mâinii stângi (valori pozitive vor corespund rotațiilor în sensul acelor de ceasornic).



Fig. 1.1. Modelul sistemului multisenzorial propus

Pentru a putea exprima un punct arbitrar P (având coordonate  $XX_S = [X, Y, Z]^T$  în sistemul de coordonate al senzorului) în sistemul de coordonate al lumii  $XX_W = [X_W, Y_W, Z_W]^T$  avem nevoie de o transformare constând dintr-o rotație urmată de o translație. Această transformare poate fi codificată prin următoarea ecuație matriceală:

$$XX_{W} = R_{W}^{s^{-1}} \cdot (XX_{s} - T_{W}^{s})$$
(1.3)

Deoarece matricea de rotație este ortogonală [Tru1998], inversa matricei R este egală cu transpusa ei:

$$R \cdot R^{T} = R^{T} \cdot R = 1 \Longrightarrow R^{T} = R^{-1}$$
(1.4)

$$\mathbf{X}\mathbf{X}_{W} = \mathbf{R}_{W}^{S^{T}} \cdot (\mathbf{X}\mathbf{X}_{S} - \mathbf{T}_{W}^{S}) = \mathbf{R}_{S}^{W} \cdot \mathbf{X}\mathbf{X}_{S} + \mathbf{T}_{S}^{W}$$
(1.5)

unde:

 $\mathbf{T}_{s}^{W}$  – vectorul de translație din sistemul camerei în sistemul lumii:  $\mathbf{T}_{s}^{W} = -\mathbf{R}_{w}^{s}^{T}\mathbf{T}_{w}^{s}$ 

(pozitia centrului sistemului de coordonate al sensorului in sistemul de coordonate al lumii);

 $\mathbf{R}_{S}^{W}$  – matricea de rotație din sistemul camerei în sistemul lumii:

$$\mathbf{R}_{s}^{W} = \mathbf{R}_{W}^{s}^{T}, \ r_{s}^{W} = -r_{W}^{s}$$
(1.6)

$$\mathbf{R}_{s}^{W} = \mathbf{R}_{w}^{s^{T}} = \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix}$$
(1.7)

Parametri extrinseci relativi descriu poziția și orientarea relativa a doi senzori (poziția și orientarea sistemului de coordonate al senzorului S2 în sistemul de coordonate al senzorului S1), și pot fi deduși din parametrii extrinseci absoluți (vectorul  $T_{rel}$  este un vector cu originea în sistemul de coordonate al senzorului S1):

$$\begin{cases} \mathbf{T}_{rel} = \mathbf{R}_{S1}^{W^{T}} (\mathbf{T}_{S2}^{W} - \mathbf{T}_{S1}^{W}) \\ \mathbf{R}_{rel} = \mathbf{R}_{S1}^{W^{T}} \cdot \mathbf{R}_{S2}^{W} \end{cases}$$
(1.8)

Pentru senzorii de viziune (camera) mai trebuie luati in considerare inca un set de parametrii care descriu geomertia interna a camerei: parametrii intrinseci. Acestia descriu caracteristicile optice și geometrice ale camerei (caracteristicile interne ale camerei). Ei depind in exclusivitate de caracteristicile constructive ale camerei (parametrii obiectivului / ansamblului de lentile, poziția centrului optic al obiectivului fata de elementul senzorial). Sunt folosiți în procesul de transformare / proiecție geometrica a unui punct 3D din sistemul de coordonate asociat camerei în planul imagine (punct 2D) și învers. Acești parametrii sunt:

- *distanța focală* distanța dintre centrul optic al sistemului de lentile al obiectivului și planul imagine: *f* [mm] sau (*fx*, *fy*) [pixeli];
- *punctul principal* coordonatele centrului real al imaginii (intersecția axei optice a obiectivului cu planul imagine) in coordonate imagine: (**u**<sub>0</sub>,**v**<sub>0</sub>) [pixeli];
- coeficienții de distorsiune radială și tangențială ai lentilelor obiectivului.

Distanta focala si punctul principal sunt inglobati in matricea interna a camerei:

$$A = \begin{bmatrix} f_{X} & 0 & u_{0} \\ 0 & f_{Y} & v_{0} \\ 0 & 0 & 1 \end{bmatrix}$$
(1.9)

Matricea de proiecție permite proiecția / maparea directa a unui punct 3D din sistemul de coordonate al lumii  $XX_W = [X_W, Y_W, Z_W]^T$  în planul imagine  $\mathbf{p}[u,v]$ . Pentru aceasta ea integrează matricea interna a camerei A necesară proiecției din sistemul de coordonate al camerei in planul imagine, si parametrii extrinseci necesari transformării de coordonate din sistemul lumii in sistemul camerei (1.2):

$$\mathbf{P} = \mathbf{A} \cdot \left[ \mathbf{R}_{W}^{C} \mid \mathbf{T}_{W}^{C} \right]$$
(1.10).

Pentru a scrie ecuația de proiecție a punctului  $XX_W$  cu ajutorul matricei de proiecție, coordonatele acestuia se vor scrie in formă omogena  $XX_W = [X_W, Y_W, Z_W, I]^T$ :

\_

$$s \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \mathbf{P} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$
(1.11)

Unde:

 $[x_s, y_s]$  coordonatele imagine ale punctului **p** scalate cu un factor de scalare  $s = z_s$ .

Deducerea coordonatelor imagine [u,v] din coordonatele scalate se calculează în felul următor:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x_s / z_s \\ y_s / z_s \end{bmatrix}$$
(1.12)

După cum se poate vedea, matricea de proiecție generează un sistem de ecuații liniar (1.11) care nu include distorsiunile lentilelor. În cazul în care se dorește și modelarea acestui tip de perturbație, practica uzuală este de a corecta distorsiunea prin corecția imaginii și aplicarea modelului de proiecție liniar (1.11).

#### 1.1.2. Calibrarea parametrilor extrinseci ai camerei FIR

Pentru experimentarea metodelor de calibrare si detectie in domeniul FIR (Far Infra Red) / LWIR (LongWaveleghth Ifra Red) s-a folosit o camera model "Pathfinder IR" produsa de FLIR [Fli2017], dedicata special pentru viziunea noctura in aplicatiii automotive, avand urmatoarele caracteristici:

- resolutia native a senzorului: 320 x 240 [pixeli]
- camp vizual (deschidere unghiulara): 36°H x 27°V
- banda spectrala: 8000 ... 14000 nm

Parametrii intrinseci ai camerei sunt furnizati de catre producator (nemaifiind necesara calibrarea acestora). Deoarece camera poate functiona in 2 rezolutii, vor exista 2 seturi de parametric (parametrii pentru rezolutia VGA obtinandu-se prin scalarea celor corespunzatori rezolutiei native):

- QVGA (nativ): fx=fy=410, u0=160, v0=120 [pixeli]
- VGA (scalat):fx=fy=820, u0=320, v0=240 [pixeli]

Calibrarea parametrilor extrinseci reprezinta procedura prin care se estimeaza vectorul de translatie (TFW si matricea de rotatie TFW) dintre sistemul de coordinate al camerei IR si sistemul de coordinate al lumii in care se raporteaza masuratorile (fig. 1).

In [Kro2007] se prezinta o metoda de calibrare bazata pe metoda [Bou2017]. Deoarece camera FIR este sensibila la amprenta termala a scenei, autorii au incalzit un sablon de calibrare (de tip tabla de sah cu o lampa cu halogen de intenistate mare pentru a obine o amprenta termala a sablonului vizibila in camera FIR (fig. 1.2). Iluminarea uniforma a sablonului a redus efectele de voalare associate difuziei termale, permitand extragerea colturilor sablonului cu precizie de sub-pixel (similar cu aplicarea procedurii de calibrare pe imaginea din spectrul vizibil. O abordare similara a fost folosita de Torabi et al. [Thor2013].



Fig. 1.2. Scenariul de calibrare bazat pe metoda [Bou2017] folosit de Krotosky et al. in [Kro2007].

In lucrarile mai sus mentionate autorii nu prezinta o evaluare cantitativa a erorilor de calibrare, ci masuri globala a errorii "registration" a metodelor implementate (acest subiect va fi reluat in capitolul 1.2), pentru scenarii indoor cu personae/pietoni aflati la distante maxime de ordinal m. Desi metoda de calibrare a parametrilor intrinseci furnizata de [Bou2015] este folosita ca referinta in literature de specialitate, metodologia de calibrare a parametrilor extrinseci folosind o singura vedere a unui sablon de calibrare planar plasat la distanta mica (2 ... 3 m) de camera nu este fiabila pentru aplicatii automotive unde este necesara perceptia corecta/precisa a mediului in scene cu adancimi de ordinul zecilor de metri [Ned2012].

Din acest motiv, o abordare similara cu ce aprezentata in [Mar06] ar oferi rezultate mult mai precise. In locul tintelor in forma de 'x' detectabile doar in spectrul vizibil s-ar putea folosi puncte de control cu amprenta termica (de exemblu un bec cu incandescenta). Un astfel de experiment s-a si realizat prin plasarea unui bec cu incandescenta in 10 pozitii diferiten cu coordinate 3D masurate/ cunoscute (fig. 1.3). Pozitia centrului imaginii becului in fiecare imagine este selectata manual si apoi extrasa cu precizie de 0.5 pixeli prin binarizarea regiunii adiacente si extragerea centrului de masa a acesteia. Desi metoda ofera o precizie de estimare a parametrilor comparabila cu cea din [Mar06], procedura de construire/masurare a scenei de calibrare este mare consumatoare de timp si trebuie repetata de fiecare data cand camera FLIR este montata (camera FLIR fiind montata printr-un sistem de prindere magnetic pe tavanul exterior al autovehiculului de test, aproximativ deasupra rig-ului stereo; din motive de securitate ea trebuie montata/demontata la fiecare utilizare).



Fig. 1.3. Scena de calibrare cu puncte de control reprezantand imaginea unui bec cu incandescenta(stanga) si ilustrarea erorilor de reproiectie a punctelor de control (dreapta).

Tabel 1.1. Valori numerice pentru punctele de control	corespinzatoare scenariului din figura 1.	.3 si
erorile de reproiectie ale acestora pe planul imagine in o	conformitate cu (1.15).	

	Puncte contr	rol FLIR <b>p</b> i	Coordonate 3D $\mathbf{P}_i$ (sensor stereo)			Erori de reproiectie (1.15)		
ID	u <sub>i</sub> [pixel]	v <sub>i</sub> [pixel]	X <sub>i</sub> [mm]	Y <sub>i</sub> [mm]	Zi [mm]	£х	εy	
1	158.8667	88.5333	0	-270	11977	1.41	0.80	
2	193.7273	77.9545	1111	-715	11063	0.55	1.18	
3	113.0400	80.4000	-1288	-715	9956	0.90	0.95	
4	206.9600	102.4000	1247	-275	8748	1.27	1.29	
5	94.6333	106.3667	-1504	-275	7850	1.18	0.55	
6	217.5000	71.5000	1247	-1174	6939	0.87	0.86	
7	109.0833	73.9722	-933	-1172	6239	0.08	1.46	
8	201.9756	128.5854	773	-275	5421	1.52	1.71	
9	117.5000	137.5000	-646	-275	4624	1.36	0.97	
10	181.9474	158.3684	350	-95	4020	1.09	0.13	
RMS error					1.10	1.08		

Din acest motiv s-a propus o metoda de calibrare semi-automata care sa utilizeze informatia 3D furnizata de senzorul de stero-viziune. Prin corelarea unor puncte de control intre imaginea din domeniul vizibil a unei camera (de ex. camera stanga) a senzorului de steroviziune si corespondentii lor din imaginea dreapta, se pot estima parametrii extrinseci ai camerei FIR prin minimzarea erori de reproiectie a coordonatelor 3D ale punctelor de control pe planul imagini FLIR.

Arhitectura senzorului de stereoviziune folosit este cea prezentata in [Ned2011]. Senzorul stereo foloseste 2 camere cu sensibilitate spectrala in domeniul vizibil si infra-rosu apropiat (400 – 850 nm) montate in spatele parbizului autovehiculului de test. Senzorul furnizeaza o harta 3D densa

a mediului folosind o implementare pe GPU a algoritmului SGM de stereocorelatie [Pan2011]. Metode de perceptie de nivel inalt [Ned2009] implementate peste datele senzoriale primare (imaginea 2D, harta 3D si datele senzoriale odometrice ale autvehiculului) sunt folosite pentru detectia si urmarirea si clasificarea obstacolelor.

Algoritmul de calibrare propus este descries mai jos:

- 1. Achizitia de perechi de imagini stereo sincronizate temporal cu imaginile FLIR (preferabil intr-un scenario static) care sa contina puncte de interes 2D vizibile in ambele imbele spectre vizuale (in exemplu cu un astfel de scenario este prezentat in fig. 1.4).
- 2. Selectia a unui numar minim de 6 puncte de interes (de prefereabil cu localizare precisa in ambele directi punctele de colt) din imaginea stereo stanga si care sa aiba si informatie 3D asociata:  $\mathbf{P}_i(X_i, Y_i, Z_i)$ .
- 3. Selectia punctelor lor corespondente in imaginea FLIR  $\mathbf{p}(u_i, v_i)$ .
- 4. Estimarea parametrilor extrinseci ai camerei FLIR ( $\mathbf{T}_{F}^{W}$  si  $\mathbf{R}_{F}^{W}$ ) prin minimizarea erorii de proiectie a punctelor 3D  $\mathbf{P}_{i}(X_{i}, Y_{i}, Z_{i})$  pe planul imagine FLIR vs. coordonatele selectate  $\mathbf{p}(u_{i}, v_{i})$  folosind metoda Gauss-Newton [Ned2012]



Fig. 1.4. Puncte de control folosite in procesul de calibrare

	Puncte contr	ol FLIR <b>p</b> i	Coordonate 3D $\mathbf{P}_i$ (stereo sensor)			
ID	u <sub>i</sub> [pixel]	v <sub>i</sub> [pixel]	X <sub>i</sub> [mm]	Y <sub>i</sub> [mm]	Zi [mm]	
1	153.25	86.92	-323	-926	7751	
2	153.99	128.94	-304	85	7775	
3	195.64	128.76	700	57	7733	
4	195.27	85.82	689	-924	7899	
5	93.16	81.03	-2165	-876	10258	
6	93.02	115.14	-2160	117	10070	
7	127.13	115.71	-1148	101	9941	
8	126.98	80.89	-1106	-911	9539	
9	237.83	87.84	1689	-905	7898	
10	236.49	129.85	1661	88	7732	

Tabel 1.2. Valori numerice pentru punctele de control corespinzatoare scenariului din figura 4 .

Pasii 2 si 3 se pot automatiza daca se alege un obiect planar de culoare neagra, forma patrata/dreptunghiulara si bordure alba, facilitand astfel deectia celor 4 colturi in amble imagini spectrale. Prin ajustarea parametrilor detectorului de colturi se poate restrictiona functia de raspuns a detectorului de colturi strict la cele 4 cloturi de interes. Corelatia intre colturile detectate in imaginea FIR si ce a din domeniul visibil se poate face pe baza pozitiei spatiale sau se poate aplica o metoda de corelatie bazata pe trasturi invariante [Agu2012].

Pentru evaluarea parametrilor extrinseci estimati, s-a folosit ca si metrica eroarea proiectiei 2D a unui set de puncte 3D generate cu senzorul stereo vs. corespondentul lor in imaginea FLIR (pentru un set de puncte independent de cel folosit in calibrare). Proiectia punctelor 3D:  $\mathbf{P}_i(X_i, Y_i, Z_i)$ 

pe planul imaginii FLIR in punctele  $\mathbf{p}_i[u_i, v_i]$  se face pe baza matricei de proiectie (vezi ec. (1.10) ... (1.12):

$$\mathbf{P}_{F} = \mathbf{A}_{F} \cdot \left[ \mathbf{R}_{W}^{F} \mid \mathbf{T}_{W}^{F} \right]$$
(1.13)

$$s \cdot \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \mathbf{P}_F \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}$$
(1.14)

Eroarea de proiectie pentru fiecare punct va fi:

$$\varepsilon = \begin{bmatrix} u_{GT} - u_{PR} \\ v_{GT} - v_{PR} \end{bmatrix}$$
(1.15)

unde:

- $[u_{GT}, v_{GT}]$  sunt cooordonatele de control 2D selectate in imaginea FIR (ground thruth GT)
- [*u<sub>PR</sub>*, *v<sub>PR</sub>*] sunt proiectiile imagine ale coordonatelor 3D ale punctelor de control corespondente furnizate de senzorul stereo.

In figura 1.5 este prezentat rezultatul calitativ al erorilor de proiectie sub forma vizuala (punctele verzi sunt punctele proiectate iar cele rosii sunt coordonatele GT selectate manual in imaginile FIR). In tabelul 1.2 sunt prezentate erorile numerice conform metricii (1.15) pentru 2 tipuri de scenarii:

- Scenariul 1 (static) cu imagini FIR achizitionate la rezolutia nativa QWGA folosind calibrarea obtinuta pe scena din fig. 4;
- Scenariul 2 (dynamic) cu imagini FIR achizitionate la rezolutia VGA in care s-a repetat procedura de calibrare pentru o noua pozitie a camerei FLIR



a. Erori de proiectie Scenariul 1(resolutie QVGA)



b. Erori de proiectie Scenariul 2 (resolutie VGA)



c. Erori de proiectie Scenariul 2 (resolutie VGA)

Fig. 1.5. Scenarii folosite pentru evaluarea parametrilor de calibrare pe baza erorii de proiectie a unor puncte 3D pe planul imaginiilor FIR.

Scenariul 2: QVGA			Scenariul 2: VGA			
ID (fig. 5.a)	£Х	εy	ID (fig 5.b)	£Х	εy	
1	-0.17	1.76	1	1.15	-0.61	
2	0.97	1.54	2	1.64	6.50	
3	0.81	1.22	3	0.14	1.77	
4	1.10	0.99	4	1.54	1.56	
5	1.09	1.56	5	-0.01	1.58	
6	-0.13	1.46	<b>RMS error</b>	1.13	3.18	
7	-0.02	0.94	ID (Fig 5.c)	£Х	εy	
8	1.76	1.02	1	2.09	-0.27	
9	1.81	0.69	2	-0.01	-2.30	
10	1.21	0.84	3	1.65	-1.26	
<b>RMS</b> error	1.09	1.25	<b>RMS</b> error	1.54	1.52	

Tabel 1.3. Valori numerice ale erorilor de proiectie.

Pentru scenariul static si imagini FIR la rezolutia VGA, eroarea RMS a proiectiilor a fost mai mica de 1.25 pixeli. Pentru scenariul dynamic la resolutia VGA erorile au fost mai mici de 2.5 pixeli (ceea ce pare explicabil datorita dublarii dimensiunii pixelului virtual si implicit a distantei focale in pixeli) precum si efectului de voalare a difuziei termale amplificata de miscare.

#### Concluzii

S-a propus o metoda de calibrare a parametrilor extrinseci ai camerei FIR care nu are nevoie de constructia unei scene de calibrare, ci doar de corelarea unor puncte de interes (de preferinta colturi) intre imaginea FLIR si imaginea stanga a senzorului stereo. Informatia de localizare spatiala a punctelor de control este furnizata de senzorul de steroviziune care se considara calibrat in prealabil si este folosit ca referinta. Metodele de calibrare dezvoltate au fost implementate si integrate intr-un prototip experimental dezvoltat in mediul MATLAB. Acuratetea de estimare a aprametrilor extrionseci s-a facut pe baza erorii de reproiectie (pe imaginea FLIR) a punctelor de interes/control plecand de la coordonatele lor 3D (furnizate de senzorul de stereoviziune). Erorile obtinute s-au dovedit comparabile cu cele in care s-a folosit o scenea de calibrare 3D masurata maual (vezi Tabel 1.3 vs. Tabel 1.1). Chiar si folosite in calibrare dar cu o precizie inferiaoara (Tabel 1.3). Totusi ca si metodologie de calibrare se recomanda folosirea unui scenario static si a unui obiect de calibrare planar cu puncte de interes /colturi vizibile/detectabile in amele imagini spectrale, plasat in diverse pozitii/adancimi fata de camera (preferabil sa acopere o arie de ordinal [m] sau zecilor de [m].

#### 1.1.3. Calibrare parametrilor extrinseci ai senzorului LiDAR

Mașinile autonome folosesc o varietate de senzori pentru a naviga în mediul înconjurător. Se dorește adăugarea informației de adâncime imaginilor 2D, care provine de la senzorii laser, precum și adăugarea informației de culoare punctelor 3D din norii de puncte, care provine de la camere. Fuziunea senzorilor formează fundația dezvoltării unor noi funcționalități complexe precum detecția pietonilor, recunoașterea semnelor din trafic și localizare și cartografiere. Pentru aceasta, este nevoie de cunoașterea pozițiilor senzorilor. Există două tipuri de parametri – intrinseci și extrinseci – care trebuie calibrați. Parametrii intrinseci ai unei camere sunt lungimea focală, centrul optic, coeficientul de înclinare și coeficienții de distorsiune, iar pentru LIDAR aceștia sunt unghiurile de înclinare ale razelor laser, unghiul azimut și reflectivitatea. Se pot utiliza setările din fabrică ale senzorilor, dar câteodată sunt necesare corecții. [Mir2012] și [Gle2010] oferă explicații detaliate ale structurii interne a unui scaner laser, în mod specific pentru tipul Velodyne HDL-64E, arătând poziționarea razelor laser și folosind o metodă a celor mai mici pătrate pentru estimarea inițială și rafinarea parametrilor interni.

Parametrii externi dau orientarea dintre senzor și obiectul folosit pentru a face măsurătorile. Locația camerei și a LIDAR-ului în lume pot fi determinate dacă se știu translația și rotația din sistemul de coordonate al senzorului în cel al lumii. Există trei metode pentru calibrarea camerei relativ senzorului laser, după [Hua2010]: calibrarea razelor vizibile – prin observarea punctelor reflectate ce necesită cameră infraroșu, calibrare bazată pe LIDAR 3D – utilizează colțuri și muchii ale unor obiecte specifice utilizate în procesul de calibrare pentru a obține relația între cele două sisteme de coordonate și calibrare bazată pe un plan 2D – observarea unui plan al unui obiect și rezolvarea constrângerilor de distanță între cei doi senzori. O metodă cunoscută este cea prezentată de [Hua2009], care constă în utilizarea unui model de tablă de șah observat în același timp de ambii senzori, din poziții și orientări diferite și estimarea coeficienților calibrării prin metoda celor mai mici pătrate.

Multe metode folosesc un obiect specific văzut de ambii senzori în același timp, pentru a găsi valoarea calibrării: [Ali2012] utilizează un cerc negru, rezolvând problema aliniamentului între normala planului elipsei detectate în imagine și normala planului detectată în punctele obținute cu ajutorul scanerului laser; în [Pan2010] se utilizează un zid și un model de tablă de șah pentru a rezolva o problemă de optimizare Levenberg-Marquardt; [Deb2013] folosește modele triangulare care sunt detectate atât în imagini 2D cât și în norul de puncte 3D, iar inter-calibrarea este obținută folosind algoritmul Levenberg-Marquardt; în [Fre2008] se utilizează un cerc negru cu interiorul gol, exploatând astfel informația de reflectivitate dată de senzorul laser; în [Vel2014], un model de tablă de șah și diferite modele planare având forme geometrice sunt detectate de cei doi senzori și corelate.

Dezavantajul celor mai multe metode este că au nevoie de un obiect specific, care nu este întotdeauna disponibil dacă se dorește recalibrarea. Astfel, cea mai evidentă alegere a unor trăsături care apar atât în imagini cât și în norii de puncte sunt muchiile. În articolul [Cas2016], muchiile sunt potrivite folosind o variație normală ponderată extrasă din imaginea de adâncime fuzionată. [Sim2016] și [Web2010] prezintă diferite metode de a detecta trăsături proeminente dintr-un nor de puncte 3D folosind gruparea bazată pe harta lui Gauss, care pot fi în continuare folosite pentru a alinia camera cu LIDAR-ul. O altă problemă a celor mai multe metode este că nu pot fi rulate în timp real. În domeniul vehiculelor autonome, o calibrare fixă nu este o soluție potrivită, deoarece senzorii pot deriva chiar și cu o mică eroare de la poziția inițială, din cauza vibrațiilor mașinii de exemplu. Astfel, inter-calibrarea senzorilor trebuie corectată. Recalibrarea este o soluție costisitoare, deci apare nevoia unui algoritm care oferă posibilitatea corecției în condiții normale de funcționare a mașinii. În [Rod2014], drumul – care este o suprafață plată – și un obstacol văzut de ambii senzori sunt folosiți pentru a calibra o cameră stereo și un scaner laser, utilizând algoritmul MSAC pentru a găsi planul drumului și rezolvând mai apoi problema alinierii datelor. În [Car2011], a metodă automată care folosește conceptul de SLAM (Simultaneous Localization and Mapping, Localizare și Cartografiere Simultană) este prezentată. În [Mik2010] se explică o metodă automată de inter-calibrare care are la bază conceptul epipolilor, aplicat camerelor mono. Articolul [Lev2013] prezintă o metodă care a reprezentat punctul de pornire al soluției implementate, prin suprapunerea muchiilor din imagine peste cele din norul de puncte.



Fig. 1.6. Arhitectura sistemului de inter-calibrare online

Figura 1.6 ilustrează pașii algoritmului online de inter-calibrare a camerei și LIDAR-ului. Se începe cu inițializarea parametrilor calibrării obținuți dintr-o metodă offline cum ar fi [Hua2009] sau [Ata2011]. Apoi, cât timp există cadre, se repetă o serie de pași. Mai întâi se detectează muchiile din imaginile 2D folosind algoritmul distance transform. Mai apoi, se corectează datele LIDAR pentru a obține o reprezentare a punctelor fără a fi influențate de viteza mașinii. Norul de puncte corectat este folosit pentru a se crea o imagine de adâncime, peste care se aplică o filtrare pentru extragerea muchiilor. Acestea vor fi suprapuse peste muchiile obținute din imagine și se va aplica algoritmul grid search pentru a se optimiza cei 6 parametrii de calibrare (translația pe axele x, y, z și cele 3 unghiuri de rotație). Dacă funcția de cost este îmbunătățită, parametrii vor fi actualizați și procesul va fi reluat în cadrul următor.

În metoda noastră, există două sisteme care lucrează împreună și anume cele două tipuri de calibrări, cei offline și cea online. Prima este obligatorie pentru a inițializa parametrii calibrării, atunci când senzorii sunt pentru prima dată poziționați pe mașină. Aceștia sunt găsiți folosind o metodă offline, utilizând un model de tablă de șah poziționat în fața celor doi senzori, simultan, în diferite poziții și orientări. Calibrarea internă a camerei este realizată folosind Camera Calibration Toolox [Bou2003]. Mai apoi se află normalele planului, folosind RANSAC și se aliniază rezolvând probleme celor mai mici pătrate pentru a afla parametrii, care sunt mai apoi rafinați utilizând Levenberg-Marquardt. Valorile inițiale sunt folosite ca intrare pentru metoda online, care va rula pe toată durata conducerii, corectând devierile senzorilor. În secțiunile următoare, se va pune accentul pe explicarea pașilor soluției.

Pentru detecția muchiilor din imagine, folosim algoritmul distance transform. Rezultatul este o imagine în tonuri de gri care arată similar cu imaginea dată ca intrare, doar că intensitățile punctelor sunt modificate pentru a arăta distanța către cel mai apropiat contur. Mai întâi se transformă imaginea de intrare într-o imagine în tonuri de gri în care valoarea fiecărui pixel este maximul dintre el și cei 8 vecini. Mai apoi se folosește metoda bazată pe masca Chamfer pentru a scana imaginea în două direcții. Aceasta va grăbi procesul de calcul. Pașii principali ai algoritmului sunt:

1. Se alege o mască de 3x3, care este mai apoi descompusă în două părți:



2. Se realizează o scanare dublă, de sus în jos și de la stânga la dreapta, mai apoi de jos în sus și de la dreapta la stânga, ca în figura 1.7, folosind masca anterior definită.



Fig. 1.7. Scanare dublă folosind masca descompusă

Se aplică următoarea actualizare a pixelilor:

$$DT(i,j) = \min_{(k,l)\in Mask} (DT(i+k,j+l) + weight(k,l))$$
(1.16)

Unde weight(k, l) este wHV daca direcția (k, l) este orizontală sau verticală relativ centrului măștii, sau wD daca direcția este diagonală. Relația dintre cele două valori trebuie să fie  $wD = wHV \cdot \sqrt{2}$ ; în implementarea noastră, folosim valorile wD=7 și wHV=5. Rezultatul poate fi văzut în figura 1.8.



Fig. 1.8. Rezultatul aplicării algoritmului distance transform

Pentru a detecta muchiile din norul de puncte 3D obținut cu ajutorul senzorului LIDAR, se aplică doi pași. Mai întâi se calculează imaginea de adâncime, folosind [Hol2015], care conține informații despre coordonatele punctelor și adâncimea acestora. Mai apoi se filtrează punctele bazat pe discontinuități. Un punct din LIDAR este notat cu *L*, iar vecinii din stânga, respectiv dreapta sa sunt notați cu  $L_{left}$ , respectiv  $L_{right}$ . Diferența dintre vecini este maximul diferenței valorilor lor, calculat astfel:

$$m = \max(L_{left} - L, L_{right} - L, 0)^{\gamma}$$
(1.17)

unde  $\gamma = 0.5$ . Mai apoi, se aplică o filtrare bazată pe adâncimea fiecărui punct:

$$E.range = \begin{cases} -\infty, if \ R < 8 \ and \log(R^{0.26}) > m \\ -\infty, if \ \log(R^{0.5}) > m \\ m, otherwise \end{cases}$$
(1.18)

unde E este noul nor de puncte, format din punctele care sunt situare pe muchii. Valorile 0.26 și 0.5 au fost obținute din măsurători și sunt specifice senzorilor utilizați în sistemul nostru. Această ultimă filtrare este necesară deoarece s-a observat că se obține o densitate mai mare de muchii la distanțe mai mari, dar informația din imagini nu este suficient de detaliată pentru a fi relevantă.

Dându-se calibrarea inițială dintre cameră și LIDAR, putem proiecta punctele de muchie 3D peste imaginea care conține muchiile obținute din cameră. Această proiecție va fi folosită pentru a evalua funcția de cost care exploatează faptul că valoarea calibrării va fi optimă atunci când există o suprapunere corectă a celor două seturi de date. Funcția de cost este suma produsului dintre intensitatea punctelor de muchie din imaginea 2D și valoarea adâncimii punctelor de pe muchie din norul de puncte 3D, luată din ultimele *w* cadre:

$$C = \sum_{j=f-w}^{f} \sum_{i=1}^{N} \sqrt{DT_i \cdot E_i \cdot range}$$
(1.19)

unde *f* denotă numărul curent al cadrului, *w* este numărul de cadre, *N* este numărul punctelor 3D proiectate pe imaginea 2D, *DT* reprezintă imaginea de muchii, iar E – norul de puncte aflate pe muchii, din care luăm doar informația de adâncime.

Pentru a determina valoarea cea mai bună a calibrării, cei 6 parametrii vor fi variați astfel: îi putem crește sau descrește cu o valoare sau pot fi lăsați neschimbați, astfel obținând 729 valori ale funcției de cost, din care se alege setul care o maximizează. Funcția noastră de cost va converge garantat doar dacă există o alterare minimală a celor 6 parametrii, depinzând în același timp de datele de la senzorul laser. Când se obține o valoare mai bună pentru funcția de cost, parametrii calibrării vor fi actualizați. În figura 1.9 se observă proiecția întregului nor de puncte peste imagine, iar în figura 1.10 se observă proiecția punctelor de muchie din norul de puncte peste imaginea obținută în urma aplicării distance transform.



Fig.1.9. Proiecția punctelor LIDAR peste imaginea în tonuri de gri



Fig.1.10. Muchiile din norul de puncte 3D suprapuse peste muchiile din imagine obținute în urma aplicării algoritmului distance transform

În sistemul nostru, parametrii calibrării sunt notați cu  $\theta_s^r = \{x_s^r, y_s^r, z_s^r, \alpha_s^r, \beta_s^r, \gamma_s^r\}$ , unde *s* reprezintă senzorul, iar *r* este sistemul de coordonate de referință. Primele 3 valori reprezintă vectorul de translație, iar ultimele 3 valori reprezintă unghiurile de rotație, care sunt folosite pentru aflarea matricii de rotație. Există mai multe matrice de calibrare. Una dintre ele este pentru calibrarea internă a camerei, notată astfel:

$$K_{C}^{i} = \begin{bmatrix} R_{C}^{i} & T_{C}^{i} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
(1.20)

După aplicarea metodei offline de calibrare între cameră și LIDAR, se obține o matrice care transformă punctele din coordonatele camerei în cele ale lumii:

$$K_{C}^{W} = \begin{bmatrix} R_{C}^{W} & T_{C}^{W} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
(1.21)

Încă o matrice este obținută, care reprezintă transformarea din coordonatele senzorului laser în cele ale camerei.

$$K_L^C = \begin{bmatrix} R_L^C & T_L^C \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
(1.22)

Se poate obține matricea de calibrare externă a LIDAR-ului în raport cu sistemul de coordonate al lumii aplicând următorul calcul:

$$K_L^W = K_C^W \cdot K_L^C \tag{1.23}$$

Sistemul nostru a fost dezvoltat în C++, folosind o serie de librării precum: PCL (Point Cloud Library), OpenCV, Boost, Qt și Eigen. Proprietățile hardware pe care s-au rulat testele sunt procesor Intel x64, cu frecvență de 4 GHz și 64 GB RAM memorie, iar sistemul de operare este Windows 10. Ca senzori s-au utilizat un Velodyne VLP-16 cu 16 straturi, orientate între -15 și +15 grade vertical, acoperind un câmp de viziune vertical de 30 de grade și orizontal de 360 de grade. Poate măsura obiecte pâna la distanța de 100 m, cu eroare de -/+3 cm. Camera utilizată este JAI BM 141GE, care oferă imagini monocrom. Sistemele de coordonate au axa z orientată în față, axa x – la dreapta și axa y – în jos.

O serie de experimente au fost realizate pentru a se testa viteza și acuratețea metodei. Cel mai important parametru care a fost variat este dimensiunea ferestrei de cadre utilizată, deoarece atunci când se folosea doar cadrul curent se obținea un timp de rulare de 16ms, pe când dacă se utiliza un număr de 9 cadre (cel curent și 8 anterioare), se obținea un timp de rulare de 90ms. Parametrii calibrării au fost alterați în mod sintetic, cu zgomot care varia între 1 și 2 cm pentru valorile de translație și 2 grade pentru rotații. Pentru evaluarea corectitudinii, s-a folosit Mean Squared Error (MSE), iar în figura 1.11 și figura 1.12 se poate observa cum aceasta variază luând în considerare diferite dimensiuni ale ferestrei de cadre. Eroare descrește o dată cu creșterea numărului de cadre, așa cum se observă în figura 1.13, dar fiind luat și timpul de rulare în considerare, concluzia a fost că valoarea optimă este de 4 cadre.



c. Translația pe axa z

Fig. 1.11. Variația parametrilor de translație pe cele 3 axe, în funcție de numărul de cadre



c. Unghiul de rotație  $\gamma$ 

Fig. 1.12. Variația parametrilor de translație pe cele 3 axe, în funcție de numărul de cadre



Fig. 1.13. Valoarea totală a erorii, variată în funcție de numărul de cadre

În figura 1.14 se pot observa calitativ proiecțiile punctelor 3D pe imaginea 2D înainte și după calibrare. Metoda rafinează cu succes parametrii calibrării, în timp real.





În tabelul 1.4, rezultatele obținute sunt comparate cu alte metode din literatură și se observă că soluția implementată obține un grad de corectitudine mai ridicat pentru 5 din 6 parametrii. Nu se obțin îmbunătățiri majore pentru translația pe axa y, deoarece senzorul LIDAR utilizat are doar 16 straturi, distanța dintre ele fiind în medie de 30 de cm, ceea ce dă o densitate scăzută pe muchiile verticale, comparativ cu informația din muchiile obținute din imagine, care este densă.

	Translație (m)			Rotație (grade)		
Articol	Х	у	Z	α	β	γ
[Nap2013]	0.0045	0.0052	0.0046	0.38	0.39	0.44
[Lev2013]	0.02	0.014	0.006	0.672	0.628	0.476
[Pan2010]	0.305	-0.005	-0.426	-0.15	0	0.27
Metoda proprie	0.002	0.015	-0.005	-0.016	0.002	0.01

Tabel 1.4. Rezultatele inter-calibrării online

#### Concluzii

Această metodă reprezintă o soluție pentru inter-calibrarea camerei cu senzorul LIDAR. Sistemul nostru poate detecta erori în cazul în care senzorii suferă modificări ale poziției în timpul conducerii mașinii, pe care le poate corecta, extrăgând astfel matricele de calibrare ale senzorilor relativ cu sistemul lumii. S-au obținut rezultate mai bune decât ale altor metode prezente în literatură, prin folosirea corecției datelor 3D, crearea unei imagini de adâncime, optimizarea în timp real al parametrilor calibrării și actualizarea acestora. Rezultatele experimentale arată că sistemul poate fi folosit în timp real, ceea ce face ca soluția să fie potrivită ca o componentă a unui sistem de asistare a conducerii autonome.

# **1.2. Studiul metodelor de de aliniere si reprezentarea 3D spatio-temporala si bazata pe aparente a datelor multi-sensoriale si multi-spectrale primare (SO2.1-1, SO2.2-1). Elaborare specificatii**

#### 1.2.1. Alinierea spatiala dintre imaginile din spectru vizibil si infra-rosu indepartat (FIR)

Problema alinierii dateror multisenzoriale consta in aducerea lor in acelasi sistem de coordonate. In cazul nostru problema se poate descompune in doua sub-probleme:

a. Asocierea dintre datele 3D furnizate se senzorul de stereoviziune si imaginea FIR

b. Asocierea intre pixelii imaginii stangi a senzorului de steroviziune si imaginea FIR.

#### 1.2.1.a. Asocierea dintre datele 3D furnizate se senzorul de stereoviziune si imaginea FIR

Avand parametrii extrinseci ai celor doi senzori calculati (relativ la sistemul unic de coordinate al lumii), pixelii  $\mathbf{p}_{\mathbf{v}}[\mathbf{x}_i, y_i]$  ai imaginii din spectrul vizibil (imaginea stanga a senzorului stereo) care au asociata si informatie 3D  $\mathbf{P}(X_i, Y_i, Z_i)$  se pot asocia cu puncte de pe imaginea FIR prin proiectia acestor puncte 3D pe planul imaginii FIR in  $\mathbf{p}_{\mathbf{F}}[u_i, v_i]$ , in conformitate cu (1.33) si (1.14).



a.

b.

с.

Fig.2.1. Ilustrarea procesului de proiectie al punctelor 3D asociate pietonilor detectati de senzorul stereo (fig. 6.c) pe imaginea stanga a senzorului (6.b) respectiv pe imaginea FIR (6.a) in conformitate cu (1.13) ,(1.14)



Fig. 2.2. a. Obiecte detectate si clasificate de senzorul stereo; b. Harta punctelor 3D reconstruite [Ned2011].

Dupa cum se poate vedea din figurile 2.1 si 2.2, atat datele de nivel coborat cat si cele de nivel inalt furnizate de senzorul de stereoviziune pot fi incomplete sau eronate:

- puncte 3D lipsa in zone fara textura sau saturati ale intensitatii
- obiecte nedetectate sau detectate/clasificate eronat.

Pentru a compensa aceste neajunsuri este nevoie de implementarea unor metode care sa permita asocierea intre pixelii imaginii stangi a senzorului de steroviziune si imaginea FIR indiferent de prezenta/absenta informatiei 3D furnizate de senzorul de stereoviziune.

#### 1.2.1.b. Asocierea intre pixelii imaginii stangi a senzorului de steroviziune si imaginea FIR

In conformitate cu [Har2013] nu se poate define o relatie de proiectie (omografie: H) intre planele imagine a 2 camere decat in anumite cazuri particulare (rotatie pura, puncte din scena coplanare sau la infinit), conditii care nu sunt respectate in majoritatea scenariilor de perceptie a mediului. In cazul general (exista translatie intre centrele optice ale celor doua camere), corespondenta se poate stabili exclusiv pe baza matricii fundamentale (F).

Conoscand parametrii intrinseci si extrinseci ai celor doua camere putem deduce matricea fundamentala F astfel [Ned2012]:

$$\mathbf{F} = \left(\mathbf{A}_{2}^{-1}\right)^{T} \mathbf{E} \mathbf{A}_{1}^{-1}$$
(2.1)

unde:

 $\mathbf{A}_1$ ,  $\mathbf{A}_2$  - matricele interne ale celor două camere

 $\mathbf{E} = \mathbf{R}_{12}\mathbf{S}$  - matricea esențială

 $\mathbf{R}_{12} = \mathbf{R}_{C2}^{W^T} \mathbf{R}_{C1}^W$  - matricea de rotație relativă din sistemul camerei 1 în sistemul camerei 2

 $\mathbf{T}_{12} = \begin{bmatrix} T_x & T_y & T_z \end{bmatrix}^T = \mathbf{R}_{C1}^w (\mathbf{T}_{C2}^w - \mathbf{T}_{C1}^w)$  - este vectorul de translație relativă din sistemul camerei din stânga în sistemul camerei din dreapta din a carui elemente se deduce matricea S:

$$\mathbf{S} = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$
(2.2)

Principala constrângere a geometriei epipoare folosită la reducerea spațiului de căutare în procesul de corelatia dintre 2 imagini afirmă că, pentru fiecare punct  $P_1(x_1, y_2)$  din imaginea 1, corespondentul său din imaginea 2  $P_2(x_2, y_2)$  se va găsi pe linia epipolară corespunzătoare  $e_2$  având următoarea ecuație (și reciproc):

$$a_2 * x + b_2 * y + c_2 = 0 \tag{2.3}$$



Fig. 8. Ilustrarea constrangerii epipolare folosita la reducerea spatiului de cautare in procesul de corelatie dintre 2 imagini.

Parametrii liniei epipolare e2 pot fi calculați cunoscând matricea fundamentală F:

$$\begin{vmatrix} a_2 \\ b_2 \\ c_2 \end{vmatrix} = \mathbf{F} * \mathbf{P}_1$$
(2.4)

Problema corespndentei/corelatiei intre pixelii a doua imagini nu este o problema triviala nici chiar in cazul in care cele doua camere au aceasi banda spectrala (vezi metoda de stereocorelatie in spectrul vizibil descrisa in [Pan2011]). Problema se complica si mai mult in cazul in care camerele au benzi spectrale diferite (corespondenta intre imaginile din spectrul vizibil si FIR) datorita relatiei neliniare dintre intensitatile pixelilor corespondenti: intensitatea pixelilor din imaginea FIR este proportionala cu temperatura obiectelor in timp ce intensitatea pixelilor din imaginea vdin spectrul vizibil este data de culoarea si reflectanta obiectelor din scena. Mai mult decat atat, imaginile FIR sunt mai voalate, au contrastul mai scazut, au mai putine detalii, textura, iar gradienti de intensitate au amplitudini mai mici (in special prdatorita procesului de difuzie termica). In continuare se va face o trecere in revista a abordarilor relevante din literatura de specialitate

#### Metode de corelatie a unor puncte de interes

In [Agu2012] autorii propun o metoda de potrivire in 3 pasi pe baza unor punctelor de interes: (1) detectarea punctelor de interes in amble imagini folosind detectorul propus de [Low2003]; (2) asocierea la fiecare punct de interes a unui descriptor invariant calculat pe baza trasaturilor EOH (Edge-Oriented-Histogram) calculate pe o vecinatati a imaginii de muchii (Metoda Canny); (3) Potrivirea punctelor de interes intre cele 2 imagini spectrale pe baza metricii de distanta euclidiena calculata pe vectorii de trasaturi (fara sa se mentioneze o strategie de cautare sau reducere a spatiului de cautare). Autorii raporteaza o precizie de 59% comparati cu 93% pentru folosirea desriptorilor SIFT ([Low2003]) respectiv 97% pentru descriptori SURF [Bay2006]; referitor la rata de recall au obtinut 74% fata de 6% (SIFT) si 45% (SURF), ceea ce indica o mai maica rata a potrivirilor fals-negative (recall) dar o mai prosta a fals pozitivelor (precision) comparativ cu SIFT/SURF. Dezavantajul principal este legat de corespondenta "sparse" a tarsaturilor. De asemena metoda este sensibila la stetarea parametrilor metodei de detectie a muchiilor (Canny) si la dimensiunea vecinatatii/regiunui pe care se calculeaza treasaturile EOH. Acuratetea potrivirilor ar putea imbunatatita prin utilizarea constrangerilor epipolare sau metoda RANSAC (parametrizata prin constrangerile epipolare).

#### Metode de corelatie densa a suprafetelor planare

In [Bar2012] este abordata problema corelatiei dense intre imagini multispectrale pe baza modelarii scenei prin suprafete planare. Metoda are trei pasi principali: (1) construirea unei harti de disparitate "sparse" folosind corelatia locala cu o functie de cost multi-modala adaptata (ca suma ponderata intre informatia mutuala a valorilor pixelilor si gradul de similaritate al vectorilor gradientilor); (2) generarea unui set de ipoteze planare care descriu suprafetele scenei prin segmentare split&merge; (3) informatia de la pasii anteriori este combinata folosind teoria campurilor Markov aleatoare, si rezolvata cu ajutorul algoritmului graph-cut. Imaginile multispectrale folosite de aceasta metoda necesita reactificare prealabila a imaginilor multispectrale.

#### Metode bazate pe corelatia fluxului optic

In [Onk2015] este prezentata o metoda de corelatie intre imaginile multi-spectrale pe baza fluxului optic dens (calculat pe fiecare imagine spectrala in parte) ca si o varianta alternativa la corelatie photometrica (cunoscuta a fi dificila intre imaginile multispectrale. Ca si metrica de corelatie se pot folosi SAD (Suma diferentelor absolute), SSD (a diferentelor patratice) sau NCC (Normalized cross corelation). Autorii nu prezinta o evaluare cantitativa a metodei dar ideea, mai ales in contextul efectelor de difuzie termala si de "motion blur" specifice seventelor dinamice.

#### Metode bazate pe informatia multuala

Elementul principal al orcarei metode de potrivire este selectia metricii de similaritate care sa potriveasca cat mai precis proprietatile fizice disparate ale imaginilor inregistrate cu senzori multimodali. Informatia mutuala furnizeaza o metrica pentru situatiile in care exista asocieri complexe intre intensitatile pixelilor obiectelor din fiecare imagine senzoriala datorita mecanismelor fizice disparate care dau nastere imaginilor multimodale. [Egn2000] a aratat ca informatia mutuala este o metrica viabila de similaritate pentru potrivirea imaginilor stereo multimodale in situatia in care dimensiunea ferestrelor mutuale de informatie este destul de mare ca sa populeze suficient histograma de probabilitate comuna a informatiei mutuale. [The2000] investigheaza proprietatile si aplicabilitatea informatiei mutuale pentru masuratori de corespondenta pe fereastra. Provocarile abordarii constau in obtinerea dimensiunilor potrivite pentru ferestre pentru a calcula informatia mutuala in scenele in care apar multi oameni si ocluzii. Coerenta modelarii datelor este asigurata de constructia unei piramide de imagini multi-rezolutie, un proces de interpolare, un instrument de optimizare si criteriul de similaritate insusi. Metoda propusa de [Kro2006] se bazeaza pe un vot cumulativ al valorilor disparitatilor din ferestre de scanare corespunzatoare. Metoda ofera o potrivire robusta pentru situatiile in care exista multe obiecte in scena cat si in situatiile cand exista obiecte partial acoperite (ocluzii). Totodata [Kro2006] prezinta si metoda de urmarire a pesoanelor in imagini video multimodale si un studiu experimental care ilustreaza utilizarea disparitatilor generate din potrivirea multimodala ca o trasatura pentru algoritmul de urmarire (tracking).

#### Metode bazate pe local-self simmilarity

Robustetea metodelor care utilizeaza informatia mutuala este restrictionata de alegerea dimensiunii ferestrelor in care se calculeaza aceasta informatie. [Tor2013] aplica auto-similaritatea locala (local self-similarity LSS) ca o metrica de similaritate multimodala densa si arata avantajele ei in comparatie cu informatia mutuala pentru potrivirea regiunilor de interes in care apar pietoni. Totodata lucrarea propune o potrivire bazata pe LSS a imaginilor termale cu imagini vizuale stereo si demonstreaza robustetea metodei in contextul scenelor cu multe persoane si acoperiri partiale. LSS capteaza parti relevante din imagine deoarece unitatea pentru a masura statisticile comune ale pixelilor este o parte mica a imaginii. Descriptorul de auto-similaritate locala capteaza aranjarea geometrica interna a auto-similaritatilor (cum ar fi muchiile) intr-o regiune a imagini. Descriptorul descrie co-ocurenta statistica a unor zone mici din imagine (de exemplu de 5x5 pixeli) intr-o regiune inconjuratoare mai mare (de exemplu de 40x40 pixeli). Pentru inceput se calculeaza o suprafata de corelatie data de suma diferentelor patratice (sum of the square differences SSD) dintre o regiune mica cu centrul la pixelul p si toate celelalte regiuni din zona inconjuratoare. SSD este normalizata cu maximul dintre varianta intensitatii din regiunea mica si zgomot. Pentru zgomot se foloseste o constanta care corespunde la variatii fotometrice acceptabile ale culorii sau luminozitatii. Apoi corelatia suprafetei este transformata intr-o reprezentare logaritmica polara impartita in 80 de intervale (20 de unghiuri si 4 intervale radiale). Descriptorul LSS este definit prin selectia valorii maxime in fiecare interval. Astfel descriptorul va avea 80 de valori. Un descriptor LSS este calculat la inceput pe o regiunea de interes din prima imagine si apoi este comparat cu alti descriptori LSS din a doua imagine folosind metrica de distanta L1.

#### Metode bazate pe trasaturi wavelet

Teoria trasaturilor wavelet a fost studiata si aplicata in diverse domenii cum ar fi reducerea zgomotului, clasificarea texturii, detectia fetelor sau a pietonilor. [Sap2016] si [Sap2016-1] aplica aceste trasaturi pentru a realiza potrivirea a doua imagini una pentru spectrul vizibil (notate cu IVS) si spectrul termal (notate cu IIR). Spectrul termal cuprinde atat infrarosul apropiat cat si infrarosul indepartat. Fiecare imagine are dimensiunea de mxn pixeli. Prin fuziunea celor doua imagini rezulta imaginea IF, de dimenziunea mxn pixeli. In schema de potrivire bazata pe trasaturi wavelet imaginile sunt sunt filtrate cu filtre trece sus si filtre trece jos pentru fiecare nivel de descompunere. Se obtin astfel reprezentari de detaliu care corespund la detalii verticale (VD), detalii orizontale si detalii diagonale. Pentru fiecare nivel de descompunere se calculeaza mai multi coeficienti la care se pot

aplica diferite scheme de fuziune cum ar fi: a) fuziune prin inlocuire in care informatia de la o imagine este complet inlocuita cu informatia de la cealalta imagine. b) fuziunea aditiva in care informatia de la o imagine se adauga la informatia celei de a doua imagini. c) fuziune ponderata in care o strategie de fuziune ponderata a imaginilor este aplicata. Transformata discreta wavelet inversa este aplicata apoi rezultatului.

#### Metode bazate pe retele neuronale

In ultimii ani retelele neuronale de convolutie (Convolutional Neural Networks) au obtinut rezultate excelente in diverse domenii de viziune artificiala. Aceste retele au fost utilizate si la potrivirea imaginilor multi-spectrale. [Agu2017] prezinta o arhitectura de retea neuronala de convolutie (Q-net) care invata descriptori de trastaturi locale care sunt folositi pentru a potrivi zone din imagini multi-spectrale. Reteaua de antrenare consta din patru copii a aceleasi retele CNN. Astfel ponderile retelelor sunt impartite. Retelele au ca intrare o perehe de zone care se potrivesc (matching patches). Zonele care se potrivesc constau din bucati de imagine care surprind aceeasi scena indiferent de spectrul de la care provin aceste zone. O pereche de zone care nu se potrivesc consta din bucati de imagine care cuprind scene diferite ale mediului in care au fost captate imaginile. La trecerea inainte reteaua calculeaza mai multe distante L2 intre iesirile fiecarei retele CNN si se obtine perechea de potriviri care are cea mai mare distanta L2 si perechea de zone care nu se potrivesc se selecteaza din bucatile care au cea mai mica distanta L2. Experimentele facute cu descriptorii obtinuti cu reteaua CNN propusa au aratat o imbunatatire de 3% fata de metodele existente.

#### Specifcatii

1. Proiectarea si implementarea unei metode de sincronizare temporala intre imagnile achizitionate de sistemul setero si camera FIR.

2. Experimentarea metodelor optime de reducere a spatiului de cautare in procesul de corelare intre pixelii imaginilor multispectrale pe baza constrangerilor epipolare si/sau rectificarea imaginilor (tinand cont de anumite restrictii legate de geometria relativa a camerelor).

3. Proiectarea si implementarea de metode de corelare mixte dintre pixelii imaginilor multispectrale prin proiectie (cand informatia 3D furnizata de senzorul de stereoviziune este disponibila) si/sau algoritmi de mathcing bazati pe trasaturi invariante locale sau globale.

# **1.2.2.** Alinierea dintre imaginile 2D si punctele 3D pentru construirea unei reprezentari spatio-temporale si bazata pe aparente a datelor sensoriale primare

#### Metode de aliniere intre imaginile 2D si puncte 3D furnizate de LiDAR

Fuziunea dintre LiDAR si camera propune unificarea informatiei acestora pentru o mai buna descriere a mediului inconjurator[Kam2007]. La baza acestei fuziuni sta proiectia punctelor 3D in imaginea camerei si atribuirea informatiei specifice imaginii pentru fiecare punct.

Pentru a face trecerea din sistemul de coordinate al LiDAR-ului in cel al camerei, punctele 3D date de senzorul LiDAR sunt proiectate direct in planul imagine cu ajutorul parametrilor extrinseci (matricea de translatie T si matricea de rotatie R ) si intrinseci (matricea K) ai camerei, luand in considerare si modelul distorsiunii lentilelor (ex. radiala, tangentiala) ce poate exista sau nu.

Fie un punct 3D dat de senzorul LiDAR  $M = [X, Y, Z]^t$  exprimat in sistemul de coordonate Euclidian. Proiectia acestuia in pixelul  $m_p = [u, v]^t$  al imaginii camerei este data de ecuatia:

$$\boldsymbol{m}_{\boldsymbol{p}} = \boldsymbol{A}[\boldsymbol{R}|\boldsymbol{T}]\boldsymbol{M} \tag{2.5}$$

unde: matricea A este de forma:

$$\begin{array}{cccccccc}
f_{x} & 0 & x_{0} \\
A = & s & f_{y} & y_{0} \\
c_{x} & c_{y} & 1
\end{array}$$
(2.6)

R, T sunt matricea, respectiv vectorul de translatie din sistemul de coordonate al LiDAR-ului in sistemul de coordonate al camerei (vezi 1.22)

In urma proiectiei, punctele 3D vor lua fie valoarea pixelilor fie a trasaturilor zonei in care cad in imagine.

Exista doua metode de a fuziona informatia de la Lidar cu camera: *metoda centralizata si metoda descentralizata*. In metoda centralizata fuziunea dintre senzori are loc la nivel de pixel sau trasatura[Bab2013],[Mah2013],[Wan2012],[Lai2012],[Sch2010]. Metoda descentralizata clasifica separate datele pentru fiecare sensor si mai apoi combina rezultatele[Dan2017], [Zha2014]. Exista de asemenea metode ce combina aceste doua strategii. [Gar2010],de exemplu, folosesc vectorii de trasaturi rezultati din informatia senzorilor pe care ii include intr-un clasificator de tip arbore de decizie combinat cu AdaBoost.

In cazul *metodei centralizate*, se observa ca numarul initial de puncte 3D va fi redus semnificativ in urma proiectiei norului de puncte pe imagine. Pentru imbunatatirea timpului computational, imaginea este de obicei presegmentata (in mai multe celule de tip grid [Wan2012], in segmente omogene[Bab2013],[Mah2013],[Dan2017]). Segmentele omogene sunt definite ca regiuni de pixeli a caror valoare a culorii este similara. Prin segmentarea imaginii in aceste regiuni obtinem superpixeli. Superpixelii rezultati din segmentarea imaginii camerei sunt asociati mai apoi cu o colectie de puncte 3D(numita patch)[Bab2013],[Mah2013] sau o colectie de trasaturi extrase din punctele 3D[Has2010][Lai2012].

In [Bab2013] autorii presupun ca exista fatade ale cladirilor pe ambele parti ale masinii cat si ca dimensiunea strazii este constanta si drumul nu este aglomerat. Se aproximeaza 2 curbe hiperbolice pe punctele 3D ce au fost deja aduse in sistemul de coordinate al imaginii adaugandu-se la ele o a treia dimensiune data de distanta de la puncte la sensorul LiDAR. Autorii propun pastrarea punctele ce sunt la o distanta mai mica de 50m, restul fiind eliminate. Doar punctele ce vor fi intre cele 2 curbe hiperbolice vor fi luate in final in considerare ca facand parte din imagine.

Autorii din [Mah2013] proiecteaza punctele 3D in imaginea camerei procesata din care s-au extras superpixelii cu ajutorul algoritmului de segmentare de tip SLIC(Simple Lidear Iterative Clustering)[ Achanta2010]. Pentru evitarea outlierilor, ei fixeaza o limita pentru minimul de puncte ce trebuie sa cada intr-un segment din imagine. Segmentele al caror numar de puncte 3D nu depaseste limita pusa sunt ignorate. Punctele 3D devin supervoxeli pe care autorii ii vor folosi pentru a calcula alte atribute utile ( eg. Normalele, centrul geometric, s.a.) pentru a segmenta sau clasifica informatia fuzionata.

In [Has2010] autorii folosesc o reprezentare de tip grid si calculeaza pentru fiecare celula trasaturi a inaltimii si a rugozitatii unei suprafete din punctele 3D si alte trasaturi de ajutor precum textura, culoare din imagine clasificand mai apoi regiunea din care s-au extras acestea. [Lai2012] utilizeaza aceeasi reprezentare. Ei se folosesc de deviatia standard pe inaltime si de intensitatea a punctelor ce cad in fiecare celula a gridului.

Trebuie sa tinem cont de faptul ca erori pot aparea in cadrul asocierii deoarece calibrarea este imperfecta si unele puncte pot fi colorate incorect din cauza aparitiei unor zgomote, a miscarii obiectelor din jur, s.a.. De asemenea, LiDARUL vede de multe ori si obiectele ce se afla in spatele camerei si care nu sunt vizibile in imagine. Zonele de interes sunt date de cele vizibile in imagine. O etapa importanta pentru fuziunea celor doi senzori va include excludere acestor outlieri.

O metoda folosita in literatura de specialitate pentru excluderea outlierilor implica asociarea regiunilor de grid ale imaginii cu patchuri de puncte 3D[Wan2012]. Imaginea camerei este astfel impartita in mai multe celule patrate de dimensiune fixa. Pentru fiecare celula se inregistreaza

punctele 3D ce cad in ea. Dupa ce se calculeaza pentru fiecare punct din celula distanta pana la senzorul LiDAR, punctele sunt ordonate in functie de valoarea distantei si sunt incluse intr-o histograma. Daca procentul numarului de puncte din nivelul curent al histogramei este mai mare decat o limita variabila, valoarea adancimii pentru acea celula este selectata. Aceasta metoda da un timp computational mare din cauza numarului mare de celule din grid care poate fi evitat prin reducerea complexitatii imaginii prin gasirea regiunilor omogene.

[Sch2010] propune exluderea outlierilor prin segmentarea cloudului de puncte si nu al imaginii ca in restul exemplelor. Calculeaza mai apoi proiectia segmentelor in imagine pastrand invelisul convex al acestora. Pentru fiecare pixel se va alege cel mai apropiat segment de care raza acestuia se loveste. Aceasta metoda este insa costisitoare. Metodele ce folosesc superpixeli pentru asociere se remarca inca o data.

O metoda buna pentru proiectul de fata ar fi combinarea metodei propuse in [Wan2012], intrucat aceasta nu include constangeri majore si da rezultatele asteptate cu cea propusa in [Mah2013] pentru imbunatatirea timpului de executie si acuratete a sistemului de fuziune. Proiectia punctelor 3D se va realiza asadar la nivel de superpixel folosind ecuatia (2.5) si va fi mai apoi corectata prin gasirea colectiei de puncte 3D cu adancimea minima si ponderea cea mai mare din cadrul fiecarui cluster. Aceasta va fi folosita in cazul in care dorim ca fuziunea sa aibe loc inainte de a detecta obiectele. Se va testa de asemenea si fuziunea senzorilor de dupa obtinerea unor scoruri de probabilitati a claselor pentru informatia venita de la camera si LiDAR. In final metoda cu acuratete si precizie cea mai mare cat si cu un timp de executie bun.

#### Segmentarea imaginii

Segmentarea semantica si detectia pietonilor reprezinta domenii cu un pontetial ridicat pentru aplicatii de timp real. Una din abordarile de baza in segmentarea semantica este data de [Sho2009] care folosesc trasaturi de tip texton. Acestea sunt combinate cu dictionare vizuale de trasaturi de textura si sunt utilizate pentru a genera canale de textoni. Pixelii individuali sunt clasificati utilizand metode de boosting aplicate pe sume calculate in regiuni dreptunghiulare ale diferitelor canale de textoni. Rezultatele clasificarii sunt integrate intr-un Conditional Random Field (CRF). Au fost propuse arhitecturi mai complexe de tip CRF care utilizeaza modele de ordin mai mare de tip  $P^n$  Potts [Koh2007] si modele robuste de tip  $P^n$  Potts [Koh2009], pixeli ierarhici [Rus2009], co-ocurente statistice [Lad2010], relatii intra-clasa spatiale[Gou2008]. Metodele non-parametrice de segmentare semantica reprezinta o alternativa fata de cele prezentate mai sus [Sin2013, Tig2013, Sha2015]. Aceste abordari extrag imagini similare din punct de vedere vizual din baze mari de date si folosesc tehnici de transfer al etichetei pentru a prezice clasele si pentru a schimba bazele de date de imagini cu un numar mare de clase semantice.

#### Fuziunea de nivel scazut 2D si 3D

[Cos2017] prezinta o abordare care foloseste perceptia multisenzoriala si exploateaza informatia de culoare, miscare si adancime. Aceasta abodare poate fi utilizata in acest proiect adaugand si informatia data de imagini grayscale. Abordarea foloseste o filtrare multimodala multirezolutie a semnalului de intensitate, a magnitudinii gradientului si a canalelor de orientare pentru a capta structura imaginii la scale multiple si la diferite orientari. O scehma de corectie este propusa pentru a obtine trasaturi invariante la scalare. Metoda incorporeaza contextul 2D si 3D prin generarea canalelor spatiale, geometrice si simetrice. Sunt obtinute rezultate foarte bune si metoda ruleaza cu 25 de cadre pe secunda.

#### Alinierea punctelor 3D din scanari LiDAR succesive si corectia erorilor de miscare

Problema alinierii a două scanări 3D într-un cadru de referință comun a fost cercetat de peste trei decenii. Arun și Kolab. [Aru1987] au prezentat o soluție a acestei probleme in care se presupune că este cunoscută corespondența punctelor din cele doua achizitii. Pentru un anumit set de puncte corespondente au propus o soluție bazata pe least squares a transformării corpului rigid

[R, t], pe baza descompunerii (SVD) a unei matrice  $3 \times 3$ . Horn [Horn1990] a propus o soluție iterativă a aceleiași probleme folosind quaternioni pentru a reprezenta rotația dintre cadrele de referință ale seturilor de puncte. Deși metodele descrise oferă o soluție la problema alinierii, ele presupun că este cunoscută corespondența dintre puncte. În cele mai multe cazuri practice, această corespondență între cele două scanări diferite nu este disponibilă direct. Deci, pentru a rezolva problema compensarii de miscare a celor doua noruri de puncte, trebuie să se stabilească mai întâi corespondența punctelor dintre cele două scanări. Au fost propuse mai multe metode de rezolvare a acestei probleme în ultimele două decenii, care pot fi clasificate pe larg în următoarele categorii: metode iterative, metode probabilistice și alte metode.

#### a. Metode iterative

Metodele iterative sunt una dintre cele mai comune metode de potrivire a scanarilor, și au cunoscut o îmbunătățire constantă de la introducerea lor de către Besl și McKay [Bes1992]. Besl și McKay a propus o metodă iterativă pentru minimizarea distanței euclidane între punctele corespunzătoare pentru a obține transformarea relativă dintre cele două scanări. Metoda lor este popular cunoscut sub numele iterativ closest point(ICP) și este utilizata pe scară largă în rezolvarea problemei de aliniere. Figura 1 prezintă cadrul algoritmului clasic ICP. Chen și Medioni [Che1992] au introdus varianta punct-la-plan a ICP datorită faptului că majoritatea măsurătorilor sunt esantionate dintr-o suprafață plană locală a mediului.



Figura 1. Corectie folosind algoritmul ICP

De la introducerea algoritmului ICP de către Chen și Medioni [Chen1992] și Besl și McKay [Bes1992], au fost propuse numeroase variații ale algoritmului pentru a crește viteza solutiei, îmbunătățirea acurateței și creșterea rezistenței la zgomot a datelor [Als2007], [Fit2003], [Gel2003], [Gre2003] [Par2003]. Un bun review ale tuturor acestor variații ale algoritmului clasic ICP este prezentat de Rusinkiewicz și Levoy [Rus2001].

O metoda moderna care considera modificarea imbunatatirea algoritmului ICP prin estimarea vitezei. De cele mai multe ori din cauza faptului ca algoritmii de acumulare considera ca toate punctele sunt achizitionate simultan se introduc erori care se acumuleaza in timp. Asadar algoritmii ICP clasici sunt predispusi sa acumuleze erori foarte mari in special in cazul miscarii rapide a vehicolului pe care sunt montati senzorii sau din cauza miscarii rapide a vehiculelor din jur. Cu toate acestea distorsiunea unei scanari in mascare poate fi compensata folosind o estimare a vitezei. Aditional, punctele care nu sunt bune sunt rejectate in iteratia de actualizare a vitezei, lucru care inseamna ca miscarea e mai robusta si mai precisa. O metoda care trateaza acest subiect este prezentata in [Hon2010].

#### b. Metode prbabilistice

Unul dintre principalele motive pentru popularitatea metodelor bazate pe ICP este simplitatea și viteza lor. Cu toate acestea, majoritatea algoritmilor determiniști discutati până acum nu iau în considerare faptul că în seturile de date din lumea reală, când scanează din două momente diferite, nu obținem niciodată corespondențe punctuale exacte. Mai mult decât atât, scanările sunt în general doar parțial suprapuse și este dificil să se stabilească corespondențe punctuale prin aplicarea unui prag pe distanța punct-la-punct. Recent, au fost propuse câteva tehnici probabilistice care modelează datele din lumea reală mai bine decât metodele deterministe. Granger și Pennec [Gra2002] au introdus o estimare generală a maximum likelyhood (ML), a transformării care aliniază cei doi nori de puncte zgomotoase, pe care ei au numit-o punctul cel mai apropiat de maximizare a așteptărilor, Expectation Maximization Iterative Closest Point (EM-ICP). Ei au arătat că, în cazul specific al zgomotului Gaussian, acesta corespunde algoritmului ICP cu distanța Mahalanobis.

#### c. Alte Metode

În această secțiune discutăm câteva metode care includ indicii vizuale din camere (senzori vizuali) în cadrul ICP. Încadrarea informațiilor vizuale în cadrul ICP a fost sugerată în mai multe variante. Johnson și Kang [Joh1999] și Godin et al. [God2001] a propus o abordare simplă de încorporare a informațiilor de culoare în cadrul ICP prin mărirea celor trei canale de culoare la coordonatele 3D ale norului punct. Deși această tehnică adaugă informațiile despre culoare în cadrul ICP, nu reușește să justifice amestecarea coordonatelor 3D ale unui punct și valorile RGB din canalul de culoare, deoarece acestea sunt două entități complet diferite. Recent, Akca [Akc2007] a propus o metodă nouă de utilizare a informațiilor despre intensitatea scanării. El a propus conceptul unei cvasi suprafețe, care este generată de scalarea normală la un anumit punct 3D prin culoarea sa și apoi potrivirea suprafeței geometrice și a cvasi-suprafeței într-un model de estimare combinat. Această abordare funcționează bine atunci când mediul este structurat și normală la un anumit punct este ușor de calculat.

#### Specifcatii

1. Sinconizarea si corectia datelor senzoriale primare

- Proiectarea si implementarea unei metode de sincronizare temporala intre senzorul de stereoviziune si LIDAR
- Proiectarea si implementarea unei solutii pentru corectia la miscare a punctelor 3D furnizate de LiDAR, la un timp referinta, la nivel de masuratoare
- Experimentarea unor algoritmi de fuziunea temporala a punctelor din mai multe cadre pentru densificarea norului de puncte

2. Proiectarea si implementarea unor modele de reprezentarea imaginilor 2D prin gruparea pixelilor in superpixeli si segmente/entitati relevante folosind agregarea datelor senzoriale primare multisenzoriale si multispectrale la nivel de pixel: intensitate/culoare, miscare si trasaturi derivate din acestea.

3. Proiectarea si dezvoltarea unui model de reprezentare spatio-temporal si bazat pe aparente prin asocierea datelor 3D furnizate de sistemul stereo si/sau LiDAR cu reprezentarea 2D la nivel de pixel si segmente/entitati relevante.

4. Testarea si validarea metodelor de aliniere si reprezentare spatio-temporale si bazate pe aparente a datelor senzoriale primare

# **1.3. Studiul metodelor de detecție si clasificare a obstacolelor si de reprezentare si fuziune a datelor senzoriale de nivel inalt (SO3.1-1, SO3.2-1, SO3.3-1). Elaborare specificatii**

Detecția și clasificarea obiectelor în sistemele avansate de asistare a conducerii reprezintă domenii vaste, intens explorate în lumea științifică dar și in industria producatorilor de masini. Aceste domenii exploreaza metode care sunt fie specifice pentru anumite tipuri de senzori sau combină informația de la mei multe tipuri de senzori. Cele mai populare abordări folosesc camere monoculare, sisteme de stereoviziune, senzori infrarosu (apropiat sau îndepărtat), senzori laser sau lidar, camere de tip bird eye cu un unghi de 360 de grade pentru explorarea informației. Majoritatea abordarilor au ca obiectiv principal detectia si clasificarea obiectelor dintre care pietonii prezinta un interes deosebit. Pentru a trece în revistă abordările existente în domeniu vom începe cu prezentarea unor articole care prezintă studii de sinteză în aceste domenii. Astfel [Gan2006] prezinta detectia pietonilor in contextul sistemelor de evitare a coliziunii. Lucrarea realizeaza o comparatie a diferitelor modalitati senzoriale cum ar fi camere pentru spectrul vizibil, camere termale senzitive la infrarosul indepartat si la infrarosul apropiat, RADAR si scaner LASER. Sunt analizate abordarile existente pentru detectia pietonilor in domeniul vizibil si in domeniul infrarosu si metodele de fuziune senzoriala de nivel inalt.

Un studiu extins al metodelor existente pentru detectia pietonilor este realizat de [Ger2010]. Lucrarea imparte procesul de detectie si clasificare in mai multe etape principale si anume: preprocesare, segmentarea fundalului, clasificarea obiectelor, verificare si ajustare, urmarirea objectelor. Metodele existente care parcurg toate sau o parte din etapele mentionate sunt prezentate si analizate in contextul senzorilor monoculari si de stereoviziune, a senzorilor pentru infrarosul indepartat si apropiat dar si din perspectiva fuziunii informatiei de la mai multi senzori. Metodele de preprocesare implica setari si experimente cu timpul de expunere, calibrari ale senzorilor. Segmentarea fundalului presupune extragerea regiunilor de interes din imagini. Abordarile existente pentru imaginile monoculare cuprind ferestre de scanare, algoritmi bazati pe retele neuronale, metode care folosesc simetria pietonilor. Pentru imaginile infrarosu abordarile existente folosesc proprietatea de simetrie, binarizare adaptiva, proiectia histogramei, retele de hipermutatii pentru clasificarea pixelilor. Regiunile de interes extrase sunt transmise modulului de clasificare care foloseste fie abordari bazate pe potrivirea siluetelor, abordari bazate pe aspect care implica extragerea unei multimi de trasaturi si antrenarea unui model de clasificare, sau abordari bazate pe retele neuronale (deep convolutional neural networks) [Tom2016]. Pasul de rafinare si ajustare a detectiilor implica metode de filtrare a instantelor fals pozitive dar si segmentarea precisa a pietonului fata de fundal. Metodele de urmarire (tracking) sunt revizuite detaliat si este subliniata importanta lor pentru eliminarea detectiilor false, pentru predictia traiectoriei pietonilor in timp si pentru analiza comportamentelor. [Ger2010] prezinta si metodele de referinta si bazele de imagini pentru testare si evaluare. In sectiunea urmatoare vom prezenta metodele existente pentru definirea regiunilor de interes in imagini mono, pentru sisteme stereo si pentru sisteme infrarosu.

#### 1.3.1. Definirea regiunilor de interes

In aceasta sectiune sunt trecute in revista principalele abordari de generare a regiunilor de interes pentru diferite sisteme senzoriale. Exista situatii cand pasul de generare a regiunii de interes si pasul de detectie a obiectelor sunt similare.

In contextul sistemelor de stereoviziune [Llo2012] trec in revista metodele existente pentru generarea regiunilor de interes in detectia obiectelor folosind informatie stereo. Ipoteza unui obiect in general este data de masuratori 3D cum ar fi harta de disparitati, v-disparitate, u-disparitate, si imagine de disparitate virtuala. Toate aceste metode se bazeaza pe ideea ca obiectele au o densitate de puncte 3D mai mare decat drumul. Abordari cu rezultate remarcabile folosesc metode de grupare

3D (subtractive clustering) [Llo2009], extinderea regiunilor pe o harta de disparitati [Ned2009], stixeli [Enz2012].

Stereoviziunea densa este folosita pentru a estima profilul drumului si pentru a rafina ariile de interes in raport cu posibilele locatii ale obiectelor din scena. [Kel2011] folosesc trasaturi de tip histograma orientarii gradientului calculate pe harta de adancime pentru a imbunatatii calitatea algoritmului de detectie a pietonilor. Abordarea lor este motivata de rezultatele din [Kel2009] care investigeaza beneficiile stereoviziunii dense pentru generarea regiunilor de interes a pietonilor.

O fuziune dintre segmentarea bazata pe miscare, harta de disparitati si detectori de pietoni bazat pe histrograma orientarii gradientului si clasificatori de tip vector suport este propusa in [Kel2011-2] pentru a construi un senzor activ de siguranta in contextul sistemelor autonome de asistare a conducerii.

In contextul senzorilor de viziune monoculara am identificat doua categorii de abordari care furnizeaza locatiile posibile pentru obiectele de interes din scenele de trafic, in special pietonii: a) abordari bazate pe context b) abordari care folosesc ferestre de scanare.

Abordarile bazate pe context privesc imaginea ca o combinatie de regiuni sau puncte de interes. Rezultatul acestor abordari il reprezinta o segmentare a regiunilor de interes fata de fundal. Algoritmii propusi pentru aceasta categorie folosesc detectia candidatilor pe baza trasaturilor vizuale cum ar fi puncte de interes invariante la scalare [Lei2008], simetrie [Ber2003], trasaturi de histograma a orientarii gradientului calculate pe diferite spatii de culoare [Rem2011]. O alta directie in generarea regiunilor de interes se bazeaza pe un vocabular de cuvinte vizuale construit pentru parti distincte si relevante ale obiectelor de interes [See2007], [Lei2008]. Localizarea pe baza informatiei semantice este propusa de [Che2012] care segmenteaza semantic imaginea in clase de interes cum ar fi drum, cer, cladiri, vehicule si pietoni. Alte abordari folosesc geometria scenei pentru a estima drumul iar ca regiuni de interes sunt pastrate doar zonele care sunt pe drum [Ben2012], [Enz2008], .

Analiza miscarii din scenele de trafic este o alta directie utilizata pentru a prezice regiunile de interes. De exemplu [Zha2013] folosesc flux optic pentru realizarea unei segmentari a miscarii si generarea ipotezelor de pietoni. [Bre2011] utilizeaza filtru de particule pentru detectia si urmarirea persoanelor.

Metodele bazate pe ferestre de scanare filtreaza numarul mare al ferestrelor de scanare care se pot genera pentru o imagine pe baza informatiei din aceste ferestre sau folosind constrangeri de localizare geometrica. Astfel regiunile de interes cuprind ferestre dreptunghiulare care au o dimensiune in pixeli cuprinsa intr-un anumit interval. Acest interval este calculat din parametrii intrinseci ai sistemului si din relatia dintre dimensiunea pietonului si distanta sa fata de camera [Sud2011], [Enz2008],[Jou2011]. O metoda care reduce semnificativ numarul de ferestre de scanare este propusa de [Che2012]. In primul rand se extrage informatia geometrica din scena si se segmenteaza imaginea in trei clase geometrice: orizontal, vertical si cer. Apoi imaginea este segmentata si impartita in superpixeli care apartin celor 3 clase diferite. Fiecare superpixel este descris de informatie de distanta care include culoare, locatie, perspectiva si textura. In pasul urmator imaginea este segmentata in functie de un interval de distanta care cuprinde zone intre 0-10m, 10-25m si mai mari de 25m. Pentru a calcula hartile de distanta se antreneaza trei clasificatori binari de tip AdaBoost care utilizeaza ca si trasaturi media de culoare, histograme pentru canalele RGB si HSV, textura, si locatia pixelilor.

Abordarea cu scanarea seriala folosind o fereastra de dimensiune fixa este depasita de o paradigma numita " multiple stage particle windows" introdusa de [Gua2012]. Se foloseste o cautare focalizata care incepe in paralel cu mai multe ferestre uniform distribuite in imagine si se concentreaza pe zonele in care exista o probabilitate mai mare de a gasi obiecte. Functia de probabilitate este estimata folosind esantionare Monte Carlo bazata pe raspunsul a patru clasificatori care folosesc matricea de covarianta, trasaturi Haar si AdaBoost, trasaturi HOG si SVM, si clasificator de tip canale de trasaturi integrale. Pentru a creste acuratetea si eficienta detectiilor se exploreaza

coerenta temporala a obiectelor folosint o retea Bayes. [Dol2014] introduc o abordare care foloseste o piramida de imagini multi-scala. Ei demonstreaza ca o clasa de trasaturi din imagine pot fi aproximate rapid fara pierderi in acuratetea detectiei prin extrapolarea valorilor trasaturilor intre scale apropiate ale dimensiunii imaginii. Se utilizeaza trasaturi de tip histograma de gradient, magnitudinea normalizata a gradientului, spatiul de culoare LUV.

Domeniul infrarosu este de mare interes deoarece surprinde caldura emisa de obiecte. Astfel utilizarea unui sensor infrarosu poate imbunatatii detectia si clasificarea obiectelor din scenele de trafic (in special a pietonilor) in conditii de noapte, in conditii meteo extreme cum ar fi ploaie, zapada, ceata. In acest proiect ne vom focaliza pe generarea regiunilor de interes pentru pietoni in imagini infrarosu deoarece aspectul lor termal poate fi recunoscut mai usor decat aspectul altor obiecte din trafic cum ar fi masini, cladiri, drum sau vegetatie. Pentru generarea regiunilor de interes (ROI) metodele existente combina prezenta petelor de caldura cu informatie de muchie si de textura [Fan2004]], [Mei2003]. O segmentare orizontala a imaginii infrarosu este prezentata in [Olm2008]. O alta abordare [Olm2009] combina binarizare, detectie de muchii si simetrie a gradientului pentru identificarea regiunii de interes. In [Kro2007] se combina simetria de muchii cu analiza profilului de intensitate si detectia drumului. O metoda de generare a zonelor candidat pentru pietoni foloseste localizarea capului pietonilor [Olm2012], [Li2009]. Printr-o operatie simpla de binarizare (transformarea imaginii din nuante de gri in alb si negru) se gasesc zonele unde ar putea fi capul pietonilor. Corelatia calculata pe trasaturi de nuante de gri este utilizata pentru a valida dreptunghiurile de incadrare a pietonilor. Totodata se foloseste si simetria muchiilor verticale calculata in dreptunghiurile de incadrare si zonele cu simetrie verticala puternica sunt pastrate pentru a extrage siluetele pietonilor. Regiunea de interes este extrasa folosind puncte discrete calculate pe imaginea de coerenta a fazei utilizand maximul si minimul momentelor de covarianta.

Pentru a genera o regiune de interes a obiectele din norul de puncte 3D provenit de la masuratorile LiDAR exista doua categorii de abordari care se bazeaza pe (1) structuri arborescente sau (2) structuri de tip grila. In acest context detectia obiectelor coincide cu generarea regiunilor de interes.

In detectia care foloseste structuri arborescente se folosesc structuri de date cum ar fi "octree" sau "range tree" [Ben2012]. Aceste metode sunt foarte bune pentru cautarea pe o gama de valori dar sunt ineficiente computational la initializare. Alte aborari recente utilizeaza metode de crestere a regiunii (region growing) pentru a detecta robust obiectele. [Azi2012] prezinta o grila de ocupare bazata pe octree care modeleaza mediul din apropierea vehiculului si detecteaza obiectele in miscare din inconsistentele intre scanari succesive. Metodele bazate pe structuri de tip grila se focalizeaza pe o procesarea 3D rapida pentru detectia obiectelor. In [Him2008] se foloseste o segmentare a norului de puncte 3D utilizand componente conexe pe o grila de ocupare 2D. In [Her1989] se descrie conceptul de harta de elevatie sau harti 2.5 care stocheaza in fiecare celula inaltimea obiectelor care sunt deasupra nivelului solului. [Rot1989] propun o grila 3D realizata din voxeli. Aceasta metoda necesita multe resurse de calcul deoarece voxelii definiti acopera tot spatiul chiar daca in realitate doar cateva masuratori sunt prezente intr-o regiune. De obicei pentru detectia obiectelor metodele bazate pe grila functioneaza impreuna cu metode de detectie a suprafetei drumului. In [Oli2015] metoda RANSAC este folosita pentru a estima planul drumului. Aceasta metoda este eficienta cand drumul este o suprafata plana. [Oni2011] propun o solutie care functioneaza si pentru drumuri curbe care au elevatie diferita. Un model cuadratic de suprafata este potrivit initial cu o regiune din fata vehiculului pentru a estima planul drumului.

#### Concluzii

Extragerea regiunilor de interes din scenele de trafic este o etapa extrem de importanta intrun sistem avansat de asistare a conducerii. Pasul de extragere a regiunii de interes reduce dimensiunea spatiului de cautare pentru detectia si clasificarea obiectelor si astfel garanteaza un timp mai redus de executie. In general obiectele de interes care apar in trafic sunt: masini, pietoni, cladiri, drum, vegetatie. Solutia propusa se va focaliza pe generarea unei regiuni de interes pentru obiectele din trafic folosind metode de stereo-viziune combinat cu lidar si informatie termala, si totodata generarea regiunilor in care exista probabilitate mare de a avea pietoni va fi tratata detaliata.

Pietonii reprezinta entitati din trafic care au un grad ridicat de variabilitate dat de forma si dimensiune, miscarea partilor componente ale corpului, diferitele atitudini: mers, alergat, stat, orientarea față de cameră, vestimentație și accesorii, vizibilitate parțială ale părților corpului, fundal complex, condiții meteo variabile, condiții de iluminare.

Generatorul de regiuni de interes va fi supus constrangerilor de rapiditate cu un timp cat mai scazut de executie, precizie data de pastrarea regiunilor cu probabilitate ridicata de a contine obiecte de interes indifierent de distanta, vizibilitate sau conditii meteorologice.

#### Specificatiile principale pentru generatorul de regiuni de interes:

- 1. Proiectarea si realizarea unor modele individuale de generare a regiunilor de interes specifice fiecarui tip de senzor in parte.
- 2. Ajustarea modelelor individuale pentru regiuni de interes care contin pietoni.
- 3. Testarea si validarea modelelor individuale.
- 4. Proiectarea si realizarea unui model de fuziune a rezultatelor modelelor individuale de generare a regiunilor de interes.
- 5. Testarea si evaluarea modelului de fuziune pentru generatorul de regiuni de interes raportat la imagini proprii adnotate in contextul acestui proiect.

#### 1.3.2. Construirea unei colectii adnotate de imagini si reprezentarea obiectelor relevante

Pentru a evalua acuratetea algoritmilor propusi exista in literatura diferite baze de date de referinta (benchmark) care contin imagini adnotate cu pietoni.

O colectie de referinta este KITTI Vision Benchmark [Gei2012] care contine adnotari 2D in imagini de intensitate, adnotari 3D pentru imagini stereo, adnotari pentru flux optic si odometrie.

O alta colectie de imagini – KAIST<sup>1</sup> – care cuprinde anotari pentru imagini infrarosu si imagini monoculare este prezentata in [Hwa2015]. Sistemul multisenzorial cu care au fost achizitionate secventele cuprinde o camera color, o camera termala si un divizor de fascicul luminos. Baza de imagini contine 95000 de perechi de imagini color-termale de dimensiune 640x480 de pixeli, la 20Hz. Toate perechile de imagini sunt adnotate manual si cuprind clase cum ar fi persoana, grup de persoane si biciclist. Exista in total.103128 adnotari diferite si 1182 pietoni unici. Adnotarile contin corespondente temporale intre aparitiile aceluias pieton in imagini succesive dintr-o secventa.

Colectivul de cercetare de la Daimler a pus la dispozitie cateva baze de imagini cu adnotari pentru imagini mono si stereo – ambele cu imagini cu nuante de gri. Acestea sunt descrise in [Kel2011-3] si [Che2012]. Adnotarile contin pietoni vizibili complet, grupuri de pietoni, pietoni partial acoperiti, biciclisti si motociclisti. Imaginile au o dimensiune de 640x480 de pixeli.

Colectia de imagini de la ETHZ este descrisa in [Ess2008]. Imaginile sunt color si au o dimensiune de 640x480 de pixeli.

Baza de imagini de la Caltech a fost introdusa de [Dol2009]. Contine imagini color achizitionate cu o camera mono. Are foarte multe imagini si contine adnotari pentru pietoni, pietoni partial vizibili si grupuri de pietoni.

<sup>&</sup>lt;sup>1</sup> <u>https://sites.google.com/site/pedestrianbenchmark/</u>

Baza de imagini TUD-Brussels a fost introdusa de [Wu2007]. Datele sunt impartite in 3 secvente: (1) "tudpedestrians" care contine 250 de imagini cu 311 pietoni adnotati vazuti din lateral, cu variatii semnificative de pozitie si imbracaminte (2) "TUD Crossing Sequence" care contine 201 imagini cu 1008 pietoni adnotati. Majoritatea sunt pietoni vazuti din lateral si multi sunt partial vizibili. (3) "TUD Campus Sequence" contine 71 de imagini cu 303 pietoni adnotati – toti vazuti din lateral.

Baza de date de la INRIA este descrisa in [Fel2010]. Contine doua multimi de imagini color cu dimensiune fixa (1) pozitive pentru pietoni si (2) negative – reprezinta imagini fara pietoni.

Baza de date de la MIT [Luo2014] a fost generata din imagini color extrase din secvente video achizitionate in Boston si Cambridge pe parcursul a mai multe anotimpuri utilizand diferite camere digitale. Postura pietonilor este limitata la vizualizare frontala si din spate. Fiecare imagine cu pietoni a fost scalata la o dimensiune fixa de 64x128 si a fost aliniata astfel incat corpul pietonului sa fie in centrul imaginii. Inaltimea pietonilor a fost aleasa astfel incat distanta dintre umeri si picioare sa fie aproximativ 80 de pixeli.

#### Specificatii constructie colectie de imagini adnotate

- 1. Realizarea unei colectii adnotate de imagini: imagini mono, stereo, infrarosu in care pietonii sunt adnotati. Acestora li se vor adauga si masuratori LiDAR.
- 2. Adnotarile cuprind tipul obiectului (pieton) si pozitia in fiecare tip de imagine.
- 3. Se vor adnota secvente luate in conditii variate de iluminare si vizibilitate.

#### 1.3.3. Detectia si clasificarea multi-senzoriala redundanta si fuziunea datelor la nivel de obiect

Dupa realizarea modului de generare a regiunilor de interes si construirea unei baze de date cu imagini adnotate pentru evaluarea intregului sistem se pot explora modalitati de detectie multisenzoriala a obiectelor din imagini si apoi de clasificare a pietonilor.

In contextul sistemelor autonome de asistare a conducerii detectia obiectelor se refera la precizarea pozitiei unde se afla un obiect in scena fara a stii ce anume reprezinta acest obiect.

Clasificarea atribuie o categorie semantica obiectelor detectate. Cele mai de interes entitati in sistemele autonome de asistare a conducerii sunt pietonii. Proiectul de fata isi propune sa realizeze o detectie multi-senzoriala a obiectelor din scenele de trafic si mai departe o clasificare a pietonilor fata de restul obiectelor.

#### Detectia obiectelor in sisteme multi-senzoriale

Exista numeroase abordari in literatura care combina diferiti senzori pentru a realiza detectia obiectelor. In particular pentru scenele de trafic se pot folosi si combina LiDar cu stereo-viziune [Ada2012],[Oh2016], LiDAR cu CCD [Gon2015], LiDAR cu radar [Nus2015] sau chiar LiDAR, radar si stereo-viziune [Cho2014]. Pentru a realiza o detectie cat mai precisa se combina informatia de la mai multi senzori.

Fuziunea de nivel inalt presupune existenta unor detectori de obiecte specifici pentru fiecare tip de senzor si realizarea unui mecanism de rationare care funzioneaza detectiile reduntante ale obiectelor. Astfel [Cha2016] propun un sistem care imbunatateste detectia si urmarirea obiectelor in miscare prin rationamente pe incertitudini. [Cho2014] extrag trasaturi independente folosind informatia de interes provenita de la fiecare senzor pentru obiectele din scena si combina aceasta informatie pentru clasificarea miscarii si urmarirea obiectelor in miscare.

[Pre2014] combina informatie de culoare si date de la LiDAR pentru a antrena un detector deformabil de pietoni care foloseste trasaturi HOG.

#### Clasificarea pietonilor in sisteme multi-senzoriale

Dupa detectia obiectelor in general se pot realiza modele de clasificare a pietonilor care permit diferentierea acestora fata de alte obiecte din scena. Clasificarea se realizeaza utilizand o multime de trasaturi extrase pe diferite parti ale pietonilor, pe intreg corpul sau combina parti multiple sau componente cu posturi diferite ale pietonilor.

Ca si trasaturi extrase in imaginile mono si stereo pentru diferite reprezentari ale pietonilor in literatura exista numeroase abordari care folosesc: trasaturi de forma [Hil2005,Fel2005], descriptori vizuali cum ar fi trasaturi Haar [Lad2012, Li2010, Wu2007], trasaturi de tip histograma orientarii gradientului [Fel2010,Wu2007, Gei2008], si variatii ale acestora [Pri2013, Fel2011], edgelets [Dol2012, Gav2007], trasaturi texturale [Ye2012], trasaturi spatio-temporale, combinatii de trasaturi cum ar fi HOG si LBP [Yu2008] calculate pe imagine de intensitate, pe flux optic si pe disparitate, canale de trasaturi [Lu02014,Ger2010] canale de cuvinte vizuale [Enz2008-1].

Avand o reprezentare de nivel inalt a pietonilor se realizeaza modele de clasificare. In literatura am identificat urmatoarele categorii de abordari: potrivirea sabloanelor care realizeaza potrivirea dintre o multime de sabloane specifice pietonilor si o regiune de interes folosind diferite metrici de potrivire cum ar fi transformata de distanta [Enz2010], potrivire ierahica a sabloanelor [Hil2005, Gon2014], clasificare bazata pe aspectul pietonilor – aceste metode folosesc algoritmi de invatare cum ar fi vectori suport (SVM) [Wu2007, Fel2010]; random forests [Gav2000, Xu2014], Hough forests [Hos2014, Lei2008-1], diferite variante de boosting [Li2010, Wu2007, Che2012], retele neuronale [Leh2009, Gei2008] dar si combinatii de clasificatori.

In domeniul infrarosu multe abordari folosesc modele care antreneaza clasificatori pe o multime de trasaturi relevante ale pietonilor in imaginea termala. Astfel de trasaturi pot fi edgelets [Cao2011], multi-block binary patterns(MLBP) [Sua2006], local binary patterns(LBP) si variatiile lor [Mal2010], histograma orientarii gradientului (HOG) [Lin2011], [Jun2010], combinatii de HOG cu trasaturi de contur [Bes2010]. Alte metode exploateza beneficiile descriptorilor invarianti la scalare cum ar fi puncte cheie discrete [Xia2010], SURF [Bes2010-1], sau dictionare de cuvinte vizuale construite pe descriptori SURF [Mir2012]. O evaluare a mai multor combinatii de trasaturi si clasificatori este realizata in [Bes2011] care includ trasaturi precum componente principale (PCA), LBP, HOG si HOPE.

Pentru detectia efectiva a pietonilor [Mei2003] utilizeaza un algoritm de corelatie a formelor bazat pe modele 3D de pietoni. Corelatii cu modele probabilistice deformabile sunt folosite si in [Olm2012], [Li2009]. O metoda de potrivire a sabloanelor este propusa de [Olm2008] care compara similaritate dintre vectori de trasaturi independente de forma calculate pe regiuni de interes cu un sablon generic de pieton. Vectorul de trasaturi cuprinde histograme de luminozitate, contrast si masuratori de inertie a imaginii. Corelatia dintre sabloane si clasificarea sabloanelor a fost explorata in [Mei2004] care combina (1)potrivire ierarhica de sabloane bazata pe transformata de distanta (2) clasificare cascadata folosind trasaturi Haar si (3) cateva retele de hiper-permutari care transforma imaginea originala in imaginea de probabilitate per pixel. Vectori suport sunt folositi de [Lin2011] care ii utilizeaza intr-un sistem compus din senzori stereo infrarosu. SVM este folosit si de [Jun2010], [Mir2012, Olm2012-1, Bes2010] si [Xia2010, Mah2005].

Abordarile bazate pe masuratorile date de LiDAR pentru detectia pietonilor combina senzori rapizi care sunt folositi pentru detectia regiunilor de interes si apoi senzori mai inceti care cauta pietonii in acele regiuni. In [Zhu2012] se foloseste un LiDAR cu cost mic (avand un unghi de vedere orizontal de 36 de grade si o rezolutie de 0.08, un unghi de vedere vertical de 7.125 cu o rezolutie de 1.425) pentru a detecta regiunile de interes. O camera color capteaza doar regiunile de interes si un clasificator de tip retea neuronala de convolutie (CNN) confirma daca regiunea de interes este sau nu

un pieton. [Pre2014] propun o abordare de detectie a pietonilor utilizand un LiDAR 3D. Se folosesc mai multe modele de clasificare pentru a detecta pietonii pana la o distanta de 30m. Metoda se bazeaza le liniaritatea si circularitatea datelor 2D pentru extragerea de trasaturi. In [Spi2010] se utilizeaza norul de puncte 3D care este impartit in diferiti nori de puncte 2D obtinute prin taierea norului de puncte 3D la diferite inaltimi. Metoda este foarte senzitiva in raport cu distanta de vizualizare.

In [Nav2010] se prezinta o metoda care extrage informatia de la un nor de puncte provenit de la un LiDAR cu rezolutie mare. Norul este impartiti in parti care corespund la partea de sus a pietonilor (piept, cap) si la picioare. Trasaturile sunt date de varianta punctelor 3D pentru fiecare parte. Abordarea nu mai este asa de precisa daca distanta de vizualizare creste.

[Kid2011] propun o metoda care realizeaza grupuri de puncte in norul 3D. Gruparea punctelor este data de inaltimea lor in raport cu solul si de imprastirea pe care o au. Se realizeaza astfel o generare a regiunilor de interes. Regiunile de interes sunt apoi clasificate si abordarea are o rata buna de detectie pentru pietoni care sunt la o distanta pana la 50m.

#### Fuziunea de nivel inalt a obiectelor detectate cu sisteme multi-senzoriale

Fuziunea senzoriala de date reprezinta procesul de a combina informatii senzoriale incomplete, imperfecte sau complementare intr-o maniera in care intelegerea unui fenomen din lumea reala poate fi mai bine perceput si modelat. Rezultatul fuziunii mai multor senzori ofera, de regula, un output mai bun decat rezultatul individual al oricarui senzor. Desi exista multe modele de fuziune senzoriala, create pentru a trata o gama variata de probleme, in contextul automotive folosind si informatiile din [Wol2014] identificam componentele pipeline-ului general al unui sistem de fuziune. Acest pieline cuprinde etape precum achizitie de date, alinierea si corectarea informatiilor intr-un sistem de referinta comun, filtrare stochastica, analiza logica, integrarea temporala si spatiala, exploatarea redundantelor. De mentionat este faptul ca dezvoltarea sistemelor moderne de fuziune senzoriala si dezvoltarea ramurii stintifice corespunzatoare au fost stimulate de creerea sistemelor de calcul cu frecvente ridicate, procesoarelor performante si programabile dar si de initiativa lui Ronald Regan din martie 1983 numita "Strategic Defense Initiative (SDI)". Dupa ce domeniul a atins un nivel de maturitate, un consiliu consultativ pentru departamentul de aparare, numit Joint Directors of Laboratories (JDL), au intreprins prima incercare de sistematizare stintifica a noii tehnologii si ramurilor inrudite [Leg2008]. Inclusiv in zilele noastre comunitatea stintifica care se ocupa de fuziunea senzoriala vorbeste de modelul JDL de Fuziune informationala.

Cercetarea in domeniul fuziunii senzoriale se efectueaza considerand un context interdisciplinar. In acest domeniu se aplica rezultate obtinute in domenii precum stinte ale naturii care abordeaza proprietatiile fizice ale senzorilor, stiinte ingineresti (in special ingineria senzorilor), metrologie, automatica, teoria controlului, matematica si informatica.

In acest document amintim si contributiile pionierilor in domeniu avand in vedere ca rezultatele cercetarilor contemporane se bazeaza pe pe acestea. Carl Friedrich Gauss a avut nenumarate contributii in stinta si tehnologie. Gauss a reusit sa estimeze parametrii orbitei planetei Ceres din masuratori zgomotoase folosind o metoda, least-squares, de compensare a errorilor definita recursiv. Aceasta metodologie [Sor1970] poate fi interpretata ca un caz restrans a filtrarii Kalman, un algorithm de baza pentru fuziunea senzoriala. O alta notiune de o importanta fundamentala este aceea de probabilitate Bayesiana. Acest concept interpreteaza conceptul de probabilitate ca o masura de cunoastere a unei stari [Jay2003], nu ca si o frecventa relativa similar cazului statisticii clasice. Functia de likelyhood reprezinta atat informatia imperfecta si incompleta oferita de senzor cat si informatia de context a performantei senzorului. Probabilitate a priori reprezinta gradul de incredere inainte ca noile masuratori senzoriale sa fie disponibile. O importanta deosebita in domeniul fuziunii senzoriale o are munca lui Abraham Wald[Wal1950] un matematician Austro-Ungar care creeaza ramura din matematica numita analiza secventiala, cu importanta in managementul urmaririi obiecteleor si testarea consistentelor. In cartea sa Statistical Decision Function, recunoaste rolu

fundamental al metodelor Bayesiene si isi numeste metodele de decizie Strategii Bayes. Un nume cu rezonannta foarte mare, care este legat de dezvoltarea sistemelor de fuziune senzoriala este Rudolf E. Kalman. Filtrul Kalman [Zar2000] este un exemplu de algoritm capabil sa infere starea unor variabile, care reprezinta proprietati ale unor obiecte ce pot evolua intr-un mod incert in timp. Fundamentele metodelor de urmarire a mai multor obiecte avand o origine incerta(multiple obiecte apropiate carora trebuie sa li se gasesca corespondenta intre cadre succesive) au fost introduse de Robert W. Sittler, care a remarcat problema [Mil1964]. In [Rei1979] Donald B. Reid a publicat o metoda care rezolva problema asocierilor multiple pentru urmarirea mai multor obiecte, rezolvand problema lui Sittler. Van Keuk, Sam S. Blackman si Yaakov Bar-Shalom au fost printre primii care au transformat metoda lui Reid in algoritmi practici [Sha2011]. Henk A. P. Blom au introdus procesele hibride stocastice in fuziunea de date [Blo2006] sub numele de "Interacting multiple models" pentru modelarea proceselor dinamice. Acesta a aplicat fuziune Bavesiana unui sistem de trafic aerian in conditii severe de functionalitate. Lary Stone este un pionier in folosirea metodelor bayesiene de urmarire si de fuziune de date pe senzori ultrasonici cu propagare in medii complexe [Sto1999]. Neil-Gordon a fost printre primii care au aplicat tehnici Monte Carlo pe probleme de urmarire neliniare, cunoscute si sub numele de "filtre de particule", care au dus la o dezvoltare a domeniului[Ris2004]. Roy Streit a introdus tehnici pentru maximizarea expectantei folosind tehnici Poisson, pentru a rezolva multipe probleme de asociere de date in contextul urmaririi si fuziunii de obiecte [Str2010]. In literatura s-au publicat numeroase carti de initiere in domeniul fuziunii senzoriale folosind diverse metode si senzori [Mit2007], [Str2001]. In cartea "Advanced Signal Processing: Theory and Implementation for Sonar, Radar, and Non-invasive Medical Diagnostics systems" [Mit2007] se ilustreaza modalitati pentru fuzionarea mai multor senzori. In [Koc2013] se prezinta o introducere in fuziunea senzoriala, ca si o ramura a stintelor ingineresti. In partea I a cartii se prezinta fundamentele teoretice necesare intelegerii aplicatiilor prezentate in a doua parte a cartii. Cartea exprima si viziunea autorului despre subiect si se subliniaza si contributiile personale in diferite stadii de dezvoltare din pipeline-ul de fuziune senzoriala. In manuscriptul numit "Sensor Data Fusion in Automotive Applications" [Pan2009] autorii descriu un model JDL revizuit pentru aplicatii automotive (Fig. 3.1.) continand nivele precum: semnal, obiecte, situatie si aplicatie.



Fig. 3.1. Modelul JDL revizuit pentru domeniul automotiv [Pan2009].

De asemenea, modelele arhitecturale folosite in fuziunea senzoriala sunt descrise atat in [Pan2009] cat si in [Koc2013]. Aceste modele arhitecturale pot fi impartite in 3 mari categorii: centralizate, distribuite si hibride. Arhitecturile centralizate sunt teoretic cele mai simple si au cea mai buna performanta, cand toti senzorii sunt aliniati (cand toti senzorii masoara aceleasi cantitati fizice). In aceasta arhitectura masuratorile brute (asa cum vin de la senzori) sunt colectate intr-un nivel central de prelucrare. Algoritmul de fuziune, va avea astfel toate masuratorile, de la toti senzorii, in momentul in care se doreste inceperea procesului de fuziune. Dezavantajul acestei solutii consta in

faptul ca algoritmul de fuziune este mult mai complex (comparativ cu cazul arhitecturilor distribuite), intrucat trebuie sa interpreteze si sa prelucreze date brute la frecvente mai mari. In figura 3.2 este prezentata arhitectura generala a procesului de fuziune centralizat. Prin Sensor1, Sensor2, etc. ne referim la senzorii la care dorim sa le aplicam algoritmul de fuziune, prin Spatial and Temporal Alignment ne referim la pasii necesari aducerii in acelasi sistem de referinta si corectiilor de miscare aplicate outputurilor senzorilor, gating se refera la procesul de limitare al al cautarilor detectiilor, asocierea reprezinta procesul de corespondenta intre masuratori (pentru a sti ce masuratoare de la un anume senzor corespunde carei predictii), componenta central track management se ocupa cu managementul masuratorilor, adica spre exemplu folosirea matricilor de covarinata corespunazoare senzorului de la care provine masuratoarea, ultima componenta filtering and prediction se refera la estimarea noilor proprietati ca urmare a fuziunii.



Fig. 3.2. Diagrama unui sistem centralizat de fuziune [Pan2009].

Arhitectura distribuita de fuziune este ilustrata in figura 3.3. Avantajul principal al acestei arhitecturi descentralizate este lipsa sensibilitatii la alinierea corecta a senzorilor. De asemenea, aceasta arhitectura are un caracter scalabil, modular, eliminand congestiile de date, si este robusta la defectarea senzorilor. In cazul fuziunii distribuite, datele pre-procesate sunt intrarile in procesorul central. Pentru fiecare senzor procesarea la nivelul semnalului poate fi efectuata in domeniul frecvential, temporal sau la nivel de pixel (in cazul fuziunii cu camere) si intrarea finala la procesorul central va fi entitatea cu toate atributele ei, cu un anumit grad de incredere in cazul fuziuniilor ulterioare in nivelul central de fuziune. Asumptia in cazul acestei arhitecturi, este ca fiecare senzor este independent de ceilalti. Redundanta informatiei este principalul dezavantaj al acestei arhitecturi.



Fig. 3.3. Arhitectura unui sistem de fuziune distribuit [Pan2009].

In cazul arhitecturilor hibride (Fig. 3.4), arhitectura centraliza este complementata de diferiti algoritmi de procesare pentru fiecare sensor. Aceste arhitecturi permit trimiterea informatiilor unui system de fuziune la nivel de masuratoare de rezerva. Arhitectura hibrida permite pastrarea avantajelor arhitecturilor centralizate si permite aditional fuzionarea informatiilor venite de la fiecare sensor individual. Dezavantajul acestui tip de arhitectura este data de complexitatea crescuta a procesului, cerintele mari la nivelul de transfer de date si corelarea intre rezultatele urmaririlor locale si centrale.



Fig. 3.4. Arhitectura hibrida [Pan2009].

Observam ca in toate arhitecturile de fuziune apar algoritmi de urmarire. In literatura exista doua metodologii de efectuare a urmaririi. O prima metoda de estimare a starii tinta din date observate este procesarea directa a semnalului brut si cautarea unor patternuri de urmarire. Aceasta metoda poarta numele de Track Before Detect(TBD). Aceasta solutie se aplica cazului in care avem un "signal to noise ratio" (SNR) scazut,adica in cazul in care semnalul provenit de la o tinta este slab si indescifrabil fata de zgomotul de mediu [Lin2008]. Metodele TBD, folosesc intregul set de masuratori din rezolutia senzorului si integreaza posibile tinte in cadre multiple in timp [Lin1964]. Ca urmare metodele TBD sunt foarte costisitoare din punct de vedere computational. O alta abordare ar fi sa aplicam prelucram semnalul brut venit de la senzori si sa extragem masuratorile dorite (denumite si detectii in literatura) [Bei1985]. Bazandu-ne pe detectiile extrase, noi valori pot fi urmarite sau detectii deja existente din masuratori precedente pot fi asociate noilor masuratori si starile lor modificate dupa noile valori[Bib2004]. In acest scenariu datorita faptului ca mai intai se realizeaza detectii iar apoi se efectueaza procesul de urmarire, algoritmii se numesc Detect-Before-Track(DBT).

#### Specificatii:

- 1. Proiectarea si realizarea unor modele individuale de detectie a obiectelor.
- 2. Proiectarea si realizarea unor modele individuale de clasificare a pietonilor aplicate pe obiectele detectate.
- 3. Testarea si evaluarea modelelor de detectie individuale si a modelelor de clasificare.
- 4. Proiectarea si realizarea unui model de fuziune a detectiilor si clasificarilor individuale de la fiecare sistem senzorial in parte folosind algoritmi de fuziune de nivel inalt.

- Proiectarea si implementarea unui mecanism de integrare intr-un sistem de referinta comun pentru obiectele detectate de fiecare senzori.
- Realizarea unui model de miscare ne-linear specific miscarii vehiculului pentru corectia miscarii obiectelor.
- Proiectarea si implementarea unui model de asociere pentru gasirea corespondentelor intre datele de nivel inalt oferite de senzori.
- Proiectarea si implementarea unei solutii de urmarire a obiectelor si utilizarea acesteia pentru fuziunea senzoriala de nivel inalt.
- 5. Testarea si evaluarea modelului de fuziune de nivel inalt.

# **Referinte bibliografice**

[Ach2010] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2010). Slic superpixels (No. EPFL-REPORT-149300).

[Agu2012] C. Aguilera et al., Multispectral Image Feature Points, *Sensors* vol. 12, p.12661-12672, 2012.

[Agu2017] Aguilera CA, Sappa AD, Aguilera C, Toledo R. Cross-Spectral Local Descriptors via Quadruplet Network. Passaro VMN, ed. *Sensors (Basel, Switzerland)*. 2017;17(4):873. doi:10.3390/s17040873.

[Akc2007] Akca, D. (2007), Matching of 3D surfaces and their intensities, *Journal of Photogrammetry and Remote Sensing*, 62(2), 112–121.

[Ali2012] H. Alismail, L. D. Baker, and B. Browning, "Automatic Calibration of a Range Sensor and Camera System," in 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, 2012, pp. 286-292.

[Als2007] Alshawa, M. (2007), ICL: Iterative closest line a novel point cloud registration algorithm based on linear features, *Theory and Application of Laser Scanning*, pp. 1–6.

[Aru1987] Arun, K.S, T. S. Huang, and S. D. Blostein (1987), Least Squares fitting of two point sets, IEEE Transaction on Pattern Analysis and Machine Intelligence, 9(5), 698-700

[Ata2011] G. Atanacio-Jiménez, J.-J. González-Barbosa, J. B. Hurtado-Ramos, F. J. Ornelas-Rodríguez, H. Jiménez-Hernández, T. García-Ramirez, et al., "LIDAR Velodyne HDL-64E Calibration Using Pattern Planes," International Journal of Advanced Robotic Systems, vol. 8, p. 59, 2011.

[Azi2012] Azim, A., Aycard, O.: Detection, classification and tracking of moving objects in a 3D environment. In: Intelligent Vehicles Symposium. pp. 802–807 (2012)

[Bab2013] Babahajiani, Pouria et al.,. Semantic parsing of street scene images using 3d lidar point cloud. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2013. p. 714-721.

[Bar2012] Barrera, F., et al. Planar-Based Multispectral Stereo. Proc. Quantitative InfraRed Thermography, 2012.

[Bay2006] Bay, H.; Tuytelaars, T.; Gool, L.V. SURF: Speeded Up Robust Features. In Proceedings of European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 404–417.

[Bei1985]Beirlant, J., and M.C.A. V. Zuijlen (1985), The empirical distribution function and strong laws for functions of order statistics of uniform spacings, Journal of Multivariate Analysis, 16(3), 300-317

[Ben2012] Benedek, C., Moln´ar, D., Szir´anyi, T.: A Dynamic MRF Model for Foreground Detection on Range Data Sequences of Rotating Multi-Beam Lidar. In: International Workshop on Depth Image Analysis, LNCS. Tsukuba City, Japan (2012)

[Ben2012] R. Benenson, M. Mathias, R. Timofte and L. Van Gool, "Pedestrian detection at 100 frames per second," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 2903-2910.

[Ber2003] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi, "Shape based pedestrian detection and localization," in Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE, vol. 1, 2003, pp. 328-333 vol.1

[Bes1992] Besl, P. J., and N. D. McKay (1992), A method for registration of 3D shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(2), 239–256

[Bes2010] B. Besbes, A. Rogozan, and A. Bensrhair, "Pedestrian recognition based on hierarchical codebook of surf features in visible and infrared images," in Intelligent Vehicles Symposium (IV), 2010 IEEE, June 2010, pp. 156-161.

[Bes2010-1]B. Besbes, A. Apatean, A. Rogozan, and A. Bensrhair, "Combining surf-based local and global features for road obstacle recognition in far infrared images," in Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on, Sept 2010, pp. 1869-1874.

[Bes2011]B. Besbes, S. Ammar, Y. Kessentini, A. Rogozan, and A. Bensrhair, "Evidential combination of svm road obstacle classi\_ers in visible and far infrared images," in Intelligent Vehicles Symposium (IV), 2011 IEEE, June 2011, pp. 1074-1079

[Bib2004] Biber, P., S. Fleck, and W. Strasser (2004), A probabilistic framework for robust and accurate matching of point clouds, in Proceedings of 26th Pattern Recognition Symposium (DAGM-04), pp. 480-487, Tubingen, Germany

[Blo2006] H.A.P. Blom, J. Jygereos (eds. ) Stochastic Hybrid Systems: Theory and Safety Critical Aplications (Springer, Berlin, 2006)

[Bou2003] J.-Y. Bouguet, "Camera Calibration Toolbox for Matlab," 2003.

[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)

[Bre2011] M. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Online multiperson tracking-by-detection from a single, uncalibrated camera," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 33, no. 9, pp. 1820-1833, Sept 2011

[Breiman2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

[Cao2011]Y. Cao, S. Pranata, and H. Nishimura, "Local binary pattern features for pedestrian detection at night/dark environment." in ICIP, B. Macq and P. Schelkens, Eds. IEEE, 2011, pp. 2053-2056.

[Car2011] G. Carrera, A. Angeli, and A. J. Davison, "SLAM-based automatic extrinsic calibration of a multi-camera rig," in 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 2652-2659.

[Cas2016] J. Castorena, U. S. Kamilov, and P. T. Boufounos, "Autocalibration of lidar and optical cameras via edge alignment," in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 2862-2866.

[Che1992] Chen, Y., and G. Medioni (1992), Object modelling by registration of multiple range images, *Image and Vision Computing*, *10*(3), 145–155.

[Che2002] Chetverikov, D., D. Svirko, D. Stepanov, and P. Krsek (2002), The trimmed iterative closest point algorithm, in *Proceedings of the International Conference on Pattern Recognition*, vol. 3, pp. 545–548, Quebec City, Canada.

[Che2012] D. Cheda, D. Ponsa, and A. Lopez, "Pedestrian candidates generation using monocular cues," in Intelligent Vehicles Symposium (IV), 2012 IEEE, June 2012, pp. 7-12.

[Cos2017] Arthur Daniel Costea, Robert Varga, Sergiu Nedevschi, "Fast Boosting Based Detection Using Scale Invariant Multimodal Multiresolution Filtered Features", vol. 00, no., pp. 993-1002, 2017, doi:10.1109/CVPR.2017.112

[Dan 2017] Dang, K., Hoy, M., Dauwels, J., & Yuan, J. (2017, July). Real-time hierarchical fusion system for semantic segmentation in offroad scenes. In *Information Fusion (Fusion), 2017 20th International Conference on* (pp. 1-6). IEEE.

[Deb2013] S. Debattisti, L. Mazzei, and M. Panciroli, "Automated extrinsic laser and camera intercalibration using triangular targets," in 2013 IEEE Intelligent Vehicles Symposium (IV), 2013, pp. 696-701.

[Dol2009] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark,"in CVPR, June 2009.

[Dol2012] P. Dollar, R. Appel, and W. Kienzle, "Crosstalk cascades for frame-rate pedestrian detection," in ECCV, 2012.

[Dol2014] P. Dollar, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 36, no. 8, pp. 1532-1545, Aug 2014.

[Egn2000] G. Egnal, Mutual information as a stereo correspondence measure, Tech. Rep. MS-CIS-00-20, University of Pennsylvania, 2000.

[Enz2008] M. Enzweiler, P. Kanter, and D. Gavrila, "Monocular pedestrian recognition using motion parallax," in Intelligent Vehicles Symposium, 2008 IEEE, June 2008, pp. 792-797

[Enz2008-1] M. Enzweiler and D. Gavrila, "A mixed generative-discriminative framework for pedestrian classiffication," in Computer Vision and Pattern Recognition, 2008. CVPR 2008.IEEE Conference on, June 2008, pp. 1-8.

[Enz2010] M. Enzweiler, A. Eigenstetter, B. Schiele, and D. M. Gavrila, "Multi-cue pedestrian classiffication with partial occlusion handling," in The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA,13-18 June 2010. IEEE, 2010, pp. 990-997.

[Enz2012] M. Enzweiler, M. Hummel, D. Pfei er, and U. Franke, "Efficient stixel-based object recognition," in Intelligent Vehicles Symposium (IV), 2012 IEEE, June 2012, pp. 1066-1071.

[Ess2008]A. Ess, B. Leibe, K. Schindler, , and L. van Gool, "A mobile vision system for robust multiperson tracking," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08). IEEE Press, June 2008.

[Fan2004] Y. Fang, K. Yamada, Y. Ninomiya, B. Horn, and I. Masaki, "A shape-independent method for pedestrian detection with far-infrared-images," IEEE Transactions On Vehicular Technology, vol. 53, no. 5, p. 1679, 2004.

[Fel2005]P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," Int. J. Comput. Vision, vol. 61, no. 1, pp. 55-79, Jan. 2005.

[Fel2010] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627-1645, 2010.

[Fel2011] P. Felzenszwalb, "Object detection grammars," in Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, Nov 2011, pp. 691-691.

[Fit2003] Fitzgibbon, A. W. (2003), Robust registration of 2D and 3D point sets, *Image and Vision Computing*, *21*(14), 1145–1153.

[Fli2017] FLIR, Pathfinder IR - User manual, (citat 2017): http://www.flir.com/uploadedFiles/UserManual\_PathFindIR.pdf

[Fre2008] S. A. R. F, V. Fremont, and P. Bonnifait, "Extrinsic calibration between a multi-layer lidar and a camera," in 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2008, pp. 214-219.

[Gan2006] T. Gandhi and M. Trivedi, "Pedestrian collision avoidance systems: a survey of computer vision based recent studies," in Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE, Sept 2006, pp. 976-981.

[Gar2010] García, F., Olmeda, D., Armingol, J. M., & de la Escalera, A. (2010, June). Hybrid fusion scheme for pedestrian detection based on laser scanner and far infrared camera. In *Intelligent Vehicles Symposium (IV), 2010 IEEE* (pp. 1134-1139). IEEE.

[Gav2000] D. Gavrila, "Pedestrian detection from a moving vehicle," in Computer Vision ECCV 2000, 6th European Conference on Computer Vision, Dublin, Ireland, June 26 - July 1, 2000, Proceedings, Part II, 2000, pp. 37-49

[Gav2007]D. Gavrila and S. Munder, 'Multi-cue pedestrian detection and tracking from a moving vehicle," International Journal of Computer Vision, vol. 73, no. 1, pp. 41-59, 2007.

[Gei2008] P. Geismann and G. Schneider, "A two-staged approach to vision-based pedestrian recognition using haar and hog features," in Intelligent Vehicles Symposium, 2008 IEEE, June 2008, pp. 554-559.

[Gel2003]Gelfand, N., L. Ikemoto, S. Rusinkiewicz, and M. Levoy (2003), Geometrically stable sampling for the ICP algorithm, in *Proceedings of Fourth International Conference on 3D Digital Imaging and Modelling*, pp. 260–267, Banff, Alberta, Canada.

[Ger2010] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 7, pp. 1239-1258, 2010.

[Gle2010] C. Glennie and D. D. Lichti, "Static Calibration and Analysis of the Velodyne HDL-64E S2 for High Accuracy Mobile Scanning," Remote Sensing, vol. 2, p. 1610, 2010.

[God2001]Godin, G., D. Laurendeau, and R. Bergevin (2001), A method for the registration of attributed range images, in *Proceeding of International Conference on 3D Digital* 

[Gon2014] A. Gonzalez, S. Ramos, D. Vazquez, A. M. Lopez, and J. Amores, "Spatiotemporal stacked sequential learning for pedestrian detection," CoRR, vol. abs/1407.3686, 2014.

[Gou2008] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller. Multi-class segmentation with relative location prior. IJCV, 2008

[Gra2002]Granger, S., and X. Pennec (2002), Multi-scale EM-ICP: A fast and robust approach for surface registration, in *Proceedings of the European Conference on Computer Vision*, pp. 418–432, London, UK.

[Gre2003] Greenspan, M., and M. Yurick (2003), Approximate k-d tree search for efficient ICP, in *Proceedings of Fourth International Conference on 3D Digital Imaging and Modeling*, pp. 442–448, Banff, Alberta, Canada.

[Gua2012] G. Gualdi, A. Prati, and R. Cucchiara, "Multistage particle windows for fast and accurate object detection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 8, pp. 1589-1604, Aug 2012.

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

[Has2011] M. Häselich, M. Arends, D. Lang, D. Paulus, Terrain classification with markov random fields on fused camera and 3d laser range data, in: Proceedings of the Fifth European Conference on Mobile Robotics (ECMR), 2011, pp. 153–158.

[Her1989] M. Herbert, C. Caillas, E. Krotkov, I. S. Kweon, T. Kanade, Terrain mapping for a roving planetary explorer, in: Robotics and Automation, 1989. Proceedings. 1989 IEEE International Conference on, IEEE, 1989, pp. 997–1002.

[Hil2005]A. Hillel, D. Weinshall, and T. Hertz, Efficient learning of relational object class models," in Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, vol. 2, Oct 2005, pp. 1762-1769 Vol. 2.

[Him2008] Himmelsbach, M., Muller, A., Luttel, T., Wunsche, H.J.: LIDAR-based 3D Object Perception. In: Proceedings of 1st International Workshop on Cognition for Technical Systems. M<sup>-</sup>unchen (Oct 2008)

[Hol2015] D. Holz, A. E. Ichim, F. Tombari, R. B. Rusu, and S. Behnke, "Registration with the Point Cloud Library: A Modular Framework for Aligning in 3-D," IEEE Robotics & Automation Magazine, vol. 22, pp. 110-124, 2015.

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

[Hon2010]S. Hong, H. Ko and J. Kim, "VICP: Velocity updating iterative closest point algorithm," *2010 IEEE International Conference on Robotics and Automation*, Anchorage, AK, 2010, pp. 1893-1898.

[Hor1990] Horn, B.K.P.(1990), Relative orientation, International Journal of Computer Vision, 4(1), 59-78

[Hos2014]J. Hosang, R. Benenson, and B. Schiele, "How good are detection proposals, really?" in BMVC, 2014.

[Hu2014] Huang, W., & Gong, X. (2014). Fusion based holistic road scene understanding. *arXiv* preprint arXiv:1406.7525.

[Hua2009] L. Huang and M. Barth, "A novel multi-planar LIDAR and computer vision calibration procedure using 2D patterns for automated navigation," in 2009 IEEE Intelligent Vehicles Symposium, 2009, pp. 117-122.

[Hua2010] L. Huang, "Lidar, camera and inertial sensors based navigation techniques for advanced intelligent transportation systems," University of California, Riverside, 2010.

[Jay2003] E. T. Jaynes, Probability Theory: The Logic of Science( Cambridge University Press, Cambridge, 2003)

[Joh1999]Johnson, A., and S. B. Kang (1999), Registration and integration of textured 3D data, *Image and Vision Computing*, *17*, 135–147.

[Jou2011] J. H. Joung, M. S. Ryoo, S. Choi, W. Yu, and H. Chae, \Background-aware pedestrian/ vehicle detection system for driving environments," in Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on, Oct 2011, pp. 1331-1336.

[Jun2010] K. Jungling and M. Arens, \Pedestrian tracking in infrared from moving vehicles," in Intelligent Vehicles Symposium (IV), 2010 IEEE, June 2010, pp. 470-477.

[Kam2017] Kampchen N., Feature-Level Fusion of Laser Scanner and Video Data for Advanced Driver Assistance Systems, PhD DIssertation, Fakultat fur Ingenieurwissenschaften und Informatik der Universitat Ulm, 2007.

[Kel2009] C. Keller, D. Llorca, and D. Gavrila, "Dense stereo-based roi generation for pedestrian detection," in Pattern Recognition, ser. Lecture Notes in Computer Science, J. Denzler, G. Notni, and H. Se, Eds. Springer Berlin Heidelberg, 2009, vol. 5748, pp. 81-90.

[Kel2011] C. Keller, M. Enzweiler, M. Rohrbach, D. Fernandez Llorca, C. Schnorr, and D. Gavrila, "The benefits of dense stereo for pedestrian detection," Intelligent Transportation Systems, IEEE Transactions on, vol. 12, no. 4, pp. 1096-1106, Dec 2011.

[Kel2011-2] C. Keller, T. Dang, H. Fritz, A. Joos, C. Rabe, and D. Gavrila, "Active pedestrian safety by automatic braking and evasive steering," Intelligent Transportation Systems, IEEE Transactions on, vol. 12, no. 4, pp. 1292-1304, Dec 2011

[Kel2011-3] C. Keller, M. Enzweiler, and D. Gavrila, "A new benchmark for stereo-based pedestrian detection," in Intelligent Vehicles Symposium (IV), 2011 IEEE, June 2011, pp. 691-696.

[Kid2011] Kidono, K.; Miyasaka, T.; Watanabe, A.; Naito, T.; Miura, J. Pedestrian recognition using high-definition LIDAR. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, 5–9 June 2011; pp. 405–410.

[Koc2013] Koch, Wolfgang, Tracking and Sensor Data Fusion, Methodological Framework and Selected Applications, 2013

[Koh2007] P. Kohli, M. P. Kumar, and P. H. Torr. P3 & beyond: Solving energies with higher order cliques. In CVPR, 2007.

[Koh2009] P. Kohli, P. H. Torr, et al. Robust higher order potentials for enforcing label consistency. IJCV, 2009.

[Kro2006] S. Krotosky et al., "Mutual information based registration of multimodal stereo videos for person tracking", *Computer Vision and Image Understanding*, vol.106, no. 2, 270-287, 2007.

[Kro2007] S. Krotosky and M. Trivedi, "On color-, infrared-, and multimodal-stereo approaches to pedestrian detection," Intelligent Transportation Systems, IEEE Transactions on, vol. 8, no. 4, pp. 619-629, Dec 2007.

[Lad2010] L. Ladicky, C. Russell, P. Kohli, and P. H. Torr. Graph cut based inference with cooccurrence statistics. In ECCV, 2010.

[Lad2012] L. Ladicky, P. H. S. Torr, and A. Zisserman, "Latent syms for human detection with a locally afine deformation field," in British Machine Vision Conference, 2012.

[Lai2012] S. Laible, Y.N. Khan, K. Bohlmann, A. Zell, 3d lidar- and camera-based terrain classification under different lighting conditions, in: AMS, 2012, pp. 21–29.

[Leg2008] M.E. Leggins, D.L. Hall, J Llians (eds.), Handbook of Multisensor Data Fusion – Theory and Practice, 2nd edn.(CRC Press, Boca Raton, 2008)

[Leh2009]A. Lehmann, B. Leibe, and L. Van Gool, "Feature-centric efficient subwindow search," in Computer Vision, 2009 IEEE 12th International Conference on, Sept 2009, pp.940-947.

[Lei2008] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," International Journal of Computer Vision, vol. 77, no. 1-3, pp. 259-289, 2008

[Lei2008-1] B. Leibe, K. Schindler, N. Cornelis, and L. Van Gool, "Coupled object detection and tracking from static cameras and moving vehicles," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 30, no. 10, pp. 1683-1698, Oct 2008.

[Lev2013] J. Levinson and S. Thrun, "Automatic Online Calibration of Cameras and Lasers," International Symposium on Experimental Robotics (ISER), 2013.

[Li2009] J. Li and Y. Wang, "Pedestrian tracking in infrared image sequences using wavelet entropy features," in Computational Intelligence and Industrial Applications, 2009.PACIIA 2009. Asia-Pacific Conference on, vol. 1, Nov 2009, pp. 288-291.

[Li2010] Z. Lin and L. Davis, "Shape-based human detection and segmentation via hierarchical parttemplate matching," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 4, pp. 604-618, April 2010.

[Lin1964] Lindley, D. V. (1964), The bayesian analysis of contingency tables. The Annals of Mathematical Statistics, 35, 1622-1643

[Lin2008] Lin, Y, G Medioni, Mutual information computation and maximization using GPU, in Computer Vision and Pattern Recognition Workshop, pp 1-6 Anchorage, Alaska, USA, 2008

[Lin2011]Y.-C. Lin, Y.-M. Chan, L.-C. Chuang, L.-C. Fu, S.-S. Huang, P.-Y. Hsiao, and M.- F. Luo, "Near-infrared based nighttime pedestrian detection by combining multiple features," in Intelligent Transportation Systems (ITSC), 2011 14th International IEEE-Conference on, Oct 2011, pp. 1549-1554.

[Llo2009] D. Llorca, M. Sotelo, I. Parra, J. Naranjo, M. Gavilan, and S. Alvarez, "An experimental study on pitch compensation in pedestrian-protection systems for collision avoidance and mitigation," Intelligent Transportation Systems, IEEE Transactions on, vol. 10, no. 3, pp. 469-474, 2009.

[Llo2012] D. Llorca, M. Sotelo, A. Helln, A. Orellana, M. Gaviln, I. Daza, and A. Lorente, "Stereo regions-of-interest selection for pedestrian protection: A survey," Transportation Research Part C: Emerging Technologies, vol. 25, pp. 226 -237, 2012

[Low2003] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, Vol. 60 Issue 2, November 2004, Pages 91-110

[Luo2014]P. Luo, Y. Tian, X. Wang, and X. Tang, "Switchable deep network for pedestrian detection," in Proceedings of the 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'14),

[Mad2012] W. Maddern, A. Harrison, and P. Newman, "Lost in translation (and rotation): Rapid extrinsic calibration for 2D and 3D LIDARs," in 2012 IEEE International Conference on Robotics and Automation, 2012, pp. 3096-3102.

[Mah2005] M. Mahlisch, M. Oberlander, O. Lohlein, D. Gavrila, and W. Ritter, "A multiple detector approach to low-resolution fir pedestrian recognition," in Intelligent Vehicles Symposium, 2005. Proceedings. IEEE, June 2005, pp. 325-330.

[Mah2013] Mahmoudabadi, Hamid et al., Superpixel clustering and planar fit segmentation of 3d lidar point clouds. In: *Computing for Geospatial Research and Application (COM. Geo), 2013 Fourth International Conference on*. IEEE, 2013. p. 1-7.

[Mal2010]R. O'Malley, E. Jones, and M. Glavin, "Detection of pedestrians in far-infrared automotive night vision using region-growing and clothing distortion compensation," Infrared Physics & Technology, vol. 53, no. 6, pp. 439 - 449, 2010.

[Mar2006] T. Mariţa, 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*, June 13-15, 2006, Tokyo, Japan, pp. 356-363.

[Mar2007] T. Mariţa, F. Oniga, S. Nedevschi, T. Graf, Calibration Accuracy Assessment Methods for Stereovision Sensors Used in Vehicles, in *Proceedings of IEEE 3-rd International Conference on Intelligent Computer Communication and Processing* (ICCP2007), 6-8 Sept. 2007, Cluj-Napoca, Romania, pp. 111-118.

[Mei2003] U. Meis, W. Ritter, and H. Neumann, "Detection and classi\_cation of obstacles in night vision traffic scenes based on infrared imagery," in Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE, vol. 2, Oct 2003, pp. 1140-1144 vol.2.

[Mei2004] U. Meis, M. Oberlander, and W. Ritter, "Reinforcing the reliability of pedestrian detection in far-infrared sensing," in Intelligent Vehicles Symposium, 2004 IEEE, June 2004, pp. 779-783.

[Mik2010] M. Miksch, B. Yang, and K. Zimmermann, "Automatic extrinsic camera self-calibration based on homography and epipolar geometry," in 2010 IEEE Intelligent Vehicles Symposium, 2010, pp. 832-839.

[Mil1964] An optimal data association problem in surveillance theory. IEEE Trans. Mil. Electron. MIL-8, 2 (1964)

[Mir2012] F. M. Mirzaei, D. G. Kottas, and S. I. Roumeliotis, "3D LIDAR-camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization," Int. J. Rob. Res., vol. 31, pp. 452-467, 2012.

[Mir2012]A. Miron, B. Besbes, A. Rogozan, S. Ainouz, and A. Bensrhair, "Intensity self similarity features for pedestrian detection in far-infrared images," in Intelligent Vehicles Symposium (IV), 2012 IEEE, June 2012, pp. 1120-1125.

[Mit2007] H.B. Mitchell, Multi-Sensor Data Fusion(Springer, Berlin, 2007)

[Nap2013] A. Napier, P. Corke, and P. Newman, "Cross-calibration of push-broom 2D LIDARs and cameras in natural scenes," in 2013 IEEE International Conference on Robotics and Automation, 2013, pp. 3679-3684.

[Nav2010] Navarro-Serment, L.E.; Mertz, C.; Hebert, M. Pedestrian Detection and Tracking Using Three-dimensional LADAR Data. Int. J. Robot. Res. 2010, 29, 1516–1528.

[Ned2009] S. Nedevschi, S. Bota, and C. Tomiuc, "Stereo-based pedestrian detection for collisiona voidance applications," Intelligent Transportation Systems, IEEE Transactions on, vol. 10, no. 3, pp. 380-391, 2009.

[Ned2009] S. Nedevschi, T. Marita, R. Danescu, F. Oniga, S. Bota, On-board Stereo Sensor for Intersection Driving Assistance. Architecture and Specification, *Proceedings of the 5-th IEEE International Conference on Intelligent Computer Communication and Processing* (ICCP2009), Cluj-Napoca, Romania, August 27-29, 2009, pp. 409-416.

[Ned2011] S. Nedevschi, T. Marita, R. Danescu, F. Oniga, S. Bota, I. Haller, C.D. Pantilie, M. Drulea, C. Golban, On-board 6D Visual Sensor for Intersection Driving Assistance, chapter in *Advanced Microsystems for Automotive Applications 2010 : Smart Systems For Green Cars And Safe Mobility*, editors G. Meyer, J. Valldorf, published *by Springer*, 10 Febr. 2011, pp. 253-264.

[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

[Oli2015] M. Oliveira, V. Santos, A. Sappa, P.Dias, Scene representations for autonomous driving: an approach based on polygonal primitives, in: 2nd Iberian Robotics Conference, 2015.

[Olm2008] D. Olmeda, C. Hilario, A. de la Escal era, and J. Armingol, "Pedestrian detection and tracking based on far infrared visual information," in Advanced Concepts for Intelligent Vision Systems, ser. Lecture Notes in Computer Science, J. Blanc-Talon, S. Bourennane, W. Philips, D. Popescu, and P. Scheunders, Eds. Springer Berlin Heidelberg, 2008, vol. 5259, pp. 958-969.

[Olm2009] D. Olmeda, A. De la Escalera, and J. Armingol, "Detection and tracking of pedestrians in infrared images," in Signals, Circuits and Systems (SCS), 2009 3rd International Conference on, Nov 2009, pp. 1-6.

[Olm2012] D. Olmeda, J. Armingol, and A. de la Escalera, "Discrete features for rapid pedestrian detection in infrared images," in Intelligent Robots and Systems (IROS), 2012 IEEE RSJ International Conference on, Oct 2012, pp. 3067-3072.

[Olm2012-1] D. Olmeda, A. de la Escalera, and J. Armingol, "Contrast invariant features for human detection in far infrared images," in Intelligent Vehicles Symposium (IV), 2012 IEEE, June 2012, pp. 117-122.

[Oni2011] Florin Oniga, Sergiu Nedevschi, Processing Dense Stereo Data Using Elevation Maps:Road Surface, Traffic Isle and Obstacle detection, IEEE Transactions on Intelligent Transportation Systems, vol 12 No 4, December 2011, pp 1331-1342

[Onk2015] Onkarappa, N, et al. Cross-spectral stereo correspondence using dense flow fields. In: Computer Vision Theory and Applications (VISAPP), 2014 International Conference on. IEEE, 2014. p. 613-617.

[Ouy2012] W. Ouyang and X. Wang, "A discriminative deep model for pedestrian detection with occlusion handling," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, June 2012, pp. 3258-3265.

[Pan2009] Panagiotis Lytrivis, George Thomaidis Angelos Amditis, Sensor Data Fusion in Automotive Applications,2009

[Pan2010] G. Pandey, J. McBride, S. Savarese, and R. Eustice, "Extrinsic Calibration of a 3D Laser Scanner and an Omnidirectional Camera," presented at the 7th IFAC Symposium on Intelligent Autonomous Vehicles, IAV 2010 - Proceedings, 2010.

[Pan2011] C. D. Pantilie, I. Haller, M. Drulea, and S. Nedevschi, "Real-Time Image Rectification and Stereo Reconstruction System on the GPU", in *2011 10th International Symposium on Parallel and Distributed Computing*, 2011, p. 79–85.

[Par2003] Park, S.-Y., and M. Subbarao (2003), An accurate and fast point-to-plane registration technique, *Pattern Recognition Letters*, 24(16), 2967–2976.

[Pre2014] C. Premebida, J. Carreira, J. Batista and U. Nunes, "Pedestrian detection combining RGB and dense LIDAR data," *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Chicago, IL, 2014, pp. 4112-4117.

[Pri2013] A. Prioletti, A. Mogelmose, P. Grisleri, M. Trivedi, A. Broggi, and T. Moeslund, "Partbased pedestrian detection and feature-based tracking for driver assistance: Real-time, robust algorithms, and evaluation," Intelligent Transportation Systems, IEEE Transactions on, vol. 14, no. 3, pp. 1346-1359, Sept 2013.

[Rei1979] D.B. Reid, An algorithm for tracking multiple targets. IEEE Trans. Autom. Control AC-24, 6 1979.

[Rem2011] K. Rematas and B. Leibe, "Effcient object detection and segmentation with a cascaded hough forest," in Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, Nov 2011, pp. 966-973.

[Ris2004] B.Ristic, S. Arulampalam, N. Gordon, Beyond the Kalman Filter: Particle Filters for Tracking Applications(Artech House Radar Library, Norwood, 2004)

[Rod2014] C. H. Rodríguez-Garavito, A. Ponz, F. García, D. Martín, A. d. l. Escalera, and J. M. Armingol, "Automatic laser and camera extrinsic calibration for data fusion using road plane," in 17th International Conference on Information Fusion (FUSION), 2014, pp. 1-6.

[Rot1989] Roth-Tabak, R. Jain, Building an environment model using depth information, Computer 22 (6) (1989) 85–90

[Rus2001]Rusinkiewicz, S., and M. Levoy (2001), Efficient variants of the ICP algorithm, in *Proceedings of Third International Conference on 3-D Digital Imaging and Modeling*, pp. 145–152, Quebec City, Canada.

[Rus2009] C. Russell, P. Kohli, P. H. Torr, et al. Associative hierarchical crfs for object class image segmentation. In CVPR, 2009.

[Sap2016] Angel D. Sappa, Cristhian A. Aguilera, Juan A. Carvajal Ayala, Miguel Oliveira, Dennis Romero, Boris X. Vintimilla, Ricardo Toledo, Monocular visual odometry: A cross-spectral image fusion based approach, In Robotics and Autonomous Systems, Volume 85, 2016, Pages 26-36, ISSN 0921-8890

[Sap2016-1] Sappa AD, Carvajal JA, Aguilera CA, Oliveira M, Romero D, Vintimilla BX. Wavelet-Based Visible and Infrared Image Fusion: A Comparative Study. Spagnolo V, Indjin D, eds. *Sensors* (*Basel, Switzerland*). 2016;16(6):861. doi:10.3390/s16060861.

[Sch2010] S. Schneider, M. Himmelsbach, T. Luettel, H.-J. Wnsche, Fusing vision and lidar – synchronization, correction and occlusion reasoning, in: Intelligent Vehicles, Symposium, 2010, pp. 388–393.

[See2007] E. Seemann, M. Fritz, and B. Schiele, "Towards robust pedestrian detection in crowded image sequences," in Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, June 2007, pp. 1-8.

[Sha2011] Y. Bar-Shalom, P.K. Willett, X. Titan, Tracking and Data Fusion. A handbook of Algorithms (YBS Publishing, Storrs, 2011)

[Sha2015] A. Sharma, O. Tuzel, and D. W. Jacobs. Deep hierarchical parsing for semantic segmentation. In CVPR, 2015

[Sho2009] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. IJCV, 2009

[Sim2016] S. Sim, J. Sock, and K. Kwak, "Indirect Correspondence-Based Robust Extrinsic Calibration of LiDAR and Camera," Sensors, vol. 16, p. 933, 2016.

[Sin2013] G. Singh and J. Kosecka. Nonparametric scene parsing with adaptive feature relevance and semantic context. In CVPR, 2013.

[Sor1970] H.W. Sorenson, Least-squares estimation: from Gauss to Kalman. IEEE Spectr. 7, 63-68 (1970)

[Spi2010] Spinello, L.; Arras, K.; Triebel, R.; Siegwart, R. A Layered Approach to People Detection in 3D Range Data. In Proceedings of the AAAI Conference on Artificial Intelligence, Atlanta, GA, USA, 11–15 July 2010; pp. 1625–1630.

[Sto1999] L.D. Stone, T.L. Corwin, C.A. Barlow, Bayesian Multiple Target Tracking (Artech House, Norwood, 1999)

[Str2001] S. Stergiopoulos (ed. ), Advanced Signal Processing: Theory and Implementation for Sonar, Radar, and Non-Invasive Medical Diagnostic Systems (CRC Press, Boca Raton, 2001)

[Str2010] R.L. Streit, Poisson Point Processes. Imaging, Tracking, and Sensing(Springer, New York 2010).

[Sua2006]F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi, "Pedestrian detection using infrared images and histograms of oriented gradients," in Intelligent Vehicles Symposium, 2006 IEEE, 2006, pp. 206-212.

[Sud2011] P. Sudowe and B. Leibe, "Efficient use of geometric constraints for sliding-window object detection in video," in Computer Vision Systems, ser. Lecture Notes in Computer Science, J. Crowley, B. Draper, and M. Thonnat, Eds. Springer Berlin Heidelberg, 2011, vol. 6962, pp. 11-20.

[The2000] P. Thevenaz, M. Unser, Optimization of mutual information for multiresolution image registration, IEEE Trans. Image Process. 9 (12) (2000) 2083–2089.

[Tig2013] J. Tighe and S. Lazebnik. Finding things: Image parsing with regions and per-exemplar detectors. In CVPR, 2013

[Tom2016] D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, S. Tubaro, Deep Convolutional Neural Networks for pedestrian detection, In Signal Processing: Image Communication, Volume 47, 2016, Pages 482-489, ISSN 0923-5965,

[Tor2013] A. Torabi, G.A. Bilodeau. "Local self-similarity-based registration of human ROIs in pairs of stereo thermal-visible videos", *Pattern Recognition*, vol. 46, no. 2, p.578-589, 2013.

[Tru19998] Trucco E., Verri A, Introductory techniques for 3D Computer Vision, Prentice Hall, 1998.

[Vel2014] M. Velas, M. Spanel, Z. Materna, and A. Herout, "Calibration of RGB camera With Velodyne LiDAR," WSCG 2014 Communication Papers Proceedings, pp. 135-144, 2014.

[Wal1950] A. Wald, Statistical Decision Functions (Wiley, New York, 1950)

[Wan2012] Wang, R., Bach, J., Macfarlane, J., & Ferrie, F. P. (2012, January). A new upsampling method for mobile LiDAR data. In *Applications of Computer Vision (WACV), 2012 IEEE Workshop on* (pp. 17-24). IEEE.

[Web2010] C. Weber, S. Hahmann, and H. Hagen, "Sharp Feature Detection in Point Clouds," presented at the Proceedings of the 2010 Shape Modeling International Conference, 2010.

[Wol2014] Wolfgang Koch, Tracking and Sensor Data Fusion, Methodological Framework and Selected Applications, 2014

[Wu2007]B. Wu and R. Nevatia, "Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors," International Journal of Computer Vision, vol. 75, no. 2, pp. 247-266, 2007

[Xia2010] D. Xia, H. Sun, and Z. Shen, "Real-time infrared pedestrian detection based on multiblock lbp," in Computer Application and System Modeling (ICCASM), 2010 International Conference on, vol. 12, Oct 2010, pp. 139-142.

[Xu2014] J. Xu, S. Ramos, D. Vazquez, and A. Lopez, "Domain adaptation of deformable partbased models," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. PP, no. 99,

[Ye2012] Q. Ye, J. Liang, and J. Jiao, "Pedestrian detection in video images via error correcting output code classification of manifold subclasses," Intelligent Transportation Systems, IEEE Transactions on, vol. 13, no. 1, pp. 193-202, March 2012.

[Yu2008] L. Yu, W. Yao, H. Liu, and F. Liu, "A monocular vision based pedestrian detection system for intelligent vehicles," in Intelligent Vehicles Symposium, 2008 IEEE, June 2008, pp. 524-529.

[Zar2000] Paul Zarchan; Howard Musoff (2000). Fundamentals of Kalman Filtering: A Practical Approach. American Institute of Aeronautics and Astronautics, Incorporated. ISBN 978-1-56347-455-2

[Zha2013] S. Zhang, C. Bauckhage, D. Klein, and A. Cremers, "Moving pedestrian detection based on motion segmentation," in Robot Vision (WORV), 2013 IEEE Workshop on, Jan 2013, pp. 102-107

[Zha2014] G. Zhao X. Xiao J. Yuan G. W. Ng "Fusion of 3d-lidar and camera data for scene parsing" Journal of Visual Communication and Image Representation, vol. 25 no. 1 pp. 165-183 2014.*Imaging and Modeling*, pp. 179–186, Qubec City, Canada.

[Zhu2012] Zhu, Q.; Chen, L.; Li, Q.; Li, M.; Nüchter, A.; Wang, J. 3D LIDAR point cloud based intersection recognition for autonomous driving. In Proceedings of the 2012 IEEE Intelligent Vehicles Symposium (IV), Madrid, Spain, 3–7 June 2012; pp. 456–461

#### Lucrari publicate de colectivul proiectului (cu Acknowledgement la proiect)

Bianca-Cerasela-Zelia Blaga, Sergiu Nedevschi, "<u>Online Cross-Calibration of Camera and LIDAR</u>", 2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, Sept. 7-9, 2017, p. 295 – 301, ISBN: 978-1-5386-3368-7, DOI:<u>10.1109/ICCP.2017.8117020</u>

Mircea Paul Muresan, Sergiu Nedevschi, Ion Giosan, "<u>Real-Time Object Detection Using a Sparse</u> <u>4-Layer LIDAR</u>", 2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, Sept. 7-9, 2017, p. 317 – 322, ISBN: 978-1-5386-3368-7, DOI: <u>10.1109/ICCP.2017.8117023</u>