MINISTERUL EDUCAȚIEI NAȚIONALE

UNIVERSITATEA TEHNICĂ
DIN CLUJ-NAPOCA

FACULTATEA DE AUTOMATICĂ ŞI CALCULATOARE
DEPARTAMENTUL CALCULATOARE

Ing. Răzvan ITU

TEZĂ DE DOCTORAT

REZUMAT

Sistem de percepție monoculară folosind viziune și inteligență artificială

CONDUCĂTOR ŞTIINŢIFIC:
Prof.dr.ing. Radu DĂNESCU

Str. G. Barițiu 26-28, 400027 Cluj-Napoca, România
Tel. +40-264-401220, fax +40-264-599893
www.cs.utcluj.ro

Introducere, context și motivație

Dezvoltarea sistemelor moderne de asistență a conducătorilor auto a luat amploare în ultimii ani datorită avansului tehnologic. Aceste sisteme au fost dezvoltate încă din anii 1960, când erau rudimentare, iar sisteme bazate pe percepția scenei prin viziune au fost realizate încă din anii 1980 și mai ales în anii 1990. Principalul interes al acestor sisteme este de a spori siguranța participanților din traficul rutier prin reducerea numărului de accidente. Principalele sisteme active și pasive de siguranță a vehiculelor sunt: centura de siguranță, sisteme de menținere a traiectoriei sau a benzii de circulație, sisteme de avertizare în caz de coliziuni frontale. Aceste sisteme au la bază diferite tipuri de senzori folosiți pentru măsurarea și percepția scenei, cum ar fi: camerele video, laser sau lidar. Percepția având sisteme compuse din una sau mai multe camere reprezintă un avantaj datorită costului redus și datorită informației densă disponibilă. Dezavantajul este dat de necesitatea unei vizibilități bune și neobstrucționate a scenei observate, dar camerele video pot fi afectate și de condiții meteo nefavorabile. Totuși, principalul avantaj este că adăugarea unui sistem de asistență/percepție bazat pe o cameră video se poate realiza într-un mod mult mai ușor față de alte tipuri de senzori, astfel încât vehiculele existente și cele mai vechi pot fi îmbunătățite. Una din cele mai ușoare metode de a adăuga sisteme de percepție unui vehicul este prin utilizarea unor platforme mobile dotate cu mai mult senzori, cum ar fi dispozitive mobile.

Dispozitivele mobile inteligente sunt omniprezente și vin echipate cu una sau mai multe camere video de diferite rezoluții, dar și cu putere de procesare din ce în ce mai mare. De asemenea, dispozitivele mobile moderne dispun de o varietate de senzori adiționali, cum ar fi: senzori de poziționare prin satelit (GPS, GLONASS), accelerometru, giroscop, senzor geomagnetic, compas, ceea ce înseamnă că ele pot fi utilizate pentru sisteme de asistență a conducătorilor auto sau ca și sisteme de analiză și monitorizare a traficului rutier. Șoferii pot plasa dispozitivele mobile în parbriz, astfel încât camera principala să fie orientată spre drum. Totuși analiza sau percepția a mediului înconjurător va necesita o calibrare a camerei cu scopul de a compara trăsăturile din scenă 3D a lumii și trăsăturile din imaginea camerei video. Acest pas de calibrare este de multe ori neglijat sau omis de către un utilizator obișnuit.

Obiective

În cadrul acestei teze, principalul obiectiv constă in proiectarea și implementarea unui sistem de percepție a mediului de trafic rutier bazat pe viziune monoculară și senzori de măsurare a poziției și inerției care să fie capabil să ruleze în diferite condiții de iluminare, în scenarii de autostradă și mai ales în traficul urban.

Pentru îndeplinirea obiectivului general, am identificat o serie de obiective secundare:

1. crearea unui sistem de achiziție și procesare a datelor senzoriale
2. detecția obstacolelor:
    - identificarea prezenței obstacolelor în spațiul imaginii folosind segmentare semantică cu rețele neuronale convoluționale
    - extragere informații 3D din imaginea monoculară prin analiza imaginii cu efectul de perspectivă eliminat
3. urmărirea și estimarea pozițiilor obstacolelor folosind modele 3D:
    - urmărire folosind un filtru de particule cu hartă de ocupare

- urmărire la nivel de cuboid

4. calibrarea automată a camerei
    - calibrare automată inițială
    - ajustarea dinamică a parametrilor camerei

Un sistem modern de asistență a șoferului necesită o percepție cât mai bună a scenei. Am ales folosirea unei singure camere video montată în vehicul în spatele parbrizului. Avantajul unui dispozitiv mobil (telefon sau tabletă) pentru achiziția de imagini este dat de faptul că oferă o rezoluție bună a imaginilor, dar și o gamă largă de senzori adiționali, totul integrat într-un dispozitiv de dimensiuni reduse care poate fi poziționat și manevrat ușor. Sistemul de achiziție constă în accesarea imaginilor din camera dispozitivului mobil și a senzorilor disponibili (GPS, accelerometru, giroscop, etc.) și procesarea lor direct pe mobil sau salvarea și trimiterea datelor pe un sistem mai performant pentru procesare mai rapidă. Pentru a putea folosi camere de la producători diferiți montate în vehicule diferite este nevoie de un sistem de calibrare cât mai eficient și cât mai puțin elaborat. Inițial am realizat un sistem de calibrare manual care necesită intervenția minimală a utilizatorului, iar apoi am prezentat abordări pentru calibrare automată unde nu este necesar nici un demers al utilizatorului. Astfel, un sistem capabil să se auto-calibreze este mult mai portabil și poate fi instalat mai ușor și în vehicule vechi, care de obicei nu au nici un sistem de asistență a șoferului.

Segmentarea semantică a scenei din trafic reprezintă interpretarea imaginilor din camera monoculară și înțelegerea acestora. Pentru acest obiectiv am ales folosirea unor tehnici de procesare bazate pe inteligența artificială, mai specific folosirea unei rețele neuronale convoluționale capabilă să ofere informații despre zona de drum (“driveable road area”) din imagini. Această abordare este robustă și invariantă la scala imaginilor și oferă rezultate foarte bune, fiind necesar doar un pas de redimensionare a imaginilor de intrare (la dimensiunea cu care a fost antrenată rețeaua inițial). Am ales detecția drumului, a spațiului liber pe care se poate circula în siguranță, deoarece astfel se poate face presupunerea că tot ce nu este drum ar putea fi un obstacol.

Zonele care nu fac parte din drum vor fi analizate și urmărite în timp folosind modele de percepție 3D, bazate pe filtre de particule. Este construită o hartă de ocupare a scenei bazată pe un filtru de particule, unde fiecare obstacol din scenă este reprezentat de un set de particule. Acesta este un model dinamic de reprezentare, deoarece are fiecare particulă are o componentă de viteză, una de poziție în harta de ocupare. Aceste informații sunt folosite pentru predicția și actualizarea hărții, când sunt adăugate sau eliminate anumite particule ce reprezintă posibile obstacole.

Prin integrarea tuturor componentelor într-un singur proiect, am reușit implementarea unui sistem de percepție monocular. Sistemul dezvoltat și prezentat în această lucrare are la baza idei și algoritmi care au fost publicați în jurnale, capitole de carte și la conferințe de specialitate.

Structura tezei

În cadrul acestei teze de doctorat am implementat un sistem de percepție și urmărire a obstacolelor din traficul rutier folosind o cameră monoculară. Teza este structurată în patru

capitole: primul conține o introducere în domeniu și necesitatea unor sisteme de asistență a șoferilor și obiectivele acestei teze, al doilea capitol prezintă metode de detecție și urmărire a obstacolelor și al treilea capitol descrie tehnicile de calibrare a camerei. Ultimul capitol al tezei se referă la concluzii și dezvoltări ulterioare, urmat de bibliografie și anexe.

În primul capitol este descrisă o statistică a accidentelor din ultimii ani, de unde rezultă o nevoie clară de a avea vehicule cât mai inteligente pe șosele. Am prezentat o scurtă istorie a vehiculelor inteligente: primele din anii 1960 care foloseau tehnici rudimentare pentru a-și menține banda, continuând cu cele de la sfârșitul anilor 1980 și anii 1990 care foloseau deja viziune artificială. În anii 2000 au fost salturi majore în acest domeniu facilitate în principal de competiția celor de la DARPA, iar după anii 2010 au fost deja lansate vehicule cu grad mare de automatizare pentru șoferi. Standardele internaționale descriu în total 5 grade de automatizare, iar senzorii folosiți pentru a obține datele din trafic sunt de asemenea prezentate în primul capitol. Obiectivele propuse, structura tezei și partea de mulțumiri sunt prezentate în secțiunile 1.2-1.4.

Capitolul 2 descrie metodele de detecție și de urmărire a obstacolelor. Am prezentat abordări pentru urmărirea obstacolelor folosind probabilistică, în care sunt menținute mai multe ipoteze despre poziția sau starea curentă a obstacolelor sau a vehiculelor, iar fiecare ipoteză are o probabilitate diferită. Am prezentat filtrul Bayes și apoi alte abordări bazate pe acesta: filtrul Kalman, filtrul Kalman extins și filtrul de particule (util pentru sisteme care nu se pot modela Gaussian). Secțiunile 2.5 și 2.6 reprezintă o introducere în inteligență artificială și rețele neuronale artificiale, care au la bază un model inspirat din neuronul biologic. Am prezentat și exemplificat operația de convoluție, iar în secțiunea 2.7 am descris diferența între clasificare și regresie și apoi în 2.8 am prezentat diferite tipuri de funcții de activare folosite pentru cele două tipuri de învățare, dar și pentru segmentare semantică. Secțiunea 2.9 conține un sumar al algoritmului de propagare a ponderilor și tehnica “gradient descent”, iar în secțiunea 2.10 sunt funcțiile de cost pentru rețele neuronale. Partea de segmentare semantică folosind rețele neuronale este descrisă în 2.11, în timp ce în secțiunea 2.12 am introdus tehnici pentru percepția monoculară. Prima parte (2.12.1) reprezintă un studiu al soluțiilor actuale bazate pe dispozitive mobile inteligente. În 2.12.2 am descris tehnicile folosite pentru estimarea adâncimii în imagini capturate folosind o singură cameră și apoi am prezentat o tehnică ce am folosit-o în alte lucrări pentru a detecta obstacole în imagini monoculare în care efectul de perspectivă este eliminat. Contribuțiile principale sunt descrise începând cu capitolul 2.13, unde am prezentat mai întâi bazele de date cu imagini din trafic (2.13.1) pe care le-am folosit pentru dezvoltarea un sistem complex de percepție a traficului rutier. Software-ul propriu pentru achiziția de imagini noi este descris în secțiunea 2.13.2. Următorul subcapitol (2.13.3) descrie pe larg o rețea neuronală pe care am folosit-o pentru segmentarea imaginilor din traficul rutier care să extragă doar zona de drum dintr-o imagine. Rezultatele obținute sunt foarte bune și pe baza de date proprie (care nu a fost folosită în timpul antrenării), iar pe seturile de validare ale bazelor de date existente rezultatele sunt bune chiar dacă metrica de evaluare este mai redusă decât a altor abordări similare. Analizând imaginile în care rezultatele segmentării nu au fost bune am constat că există anumite erori în adnotările din bazele de date cunoscute, iar rețeaua dezvoltată de mine a oferit uneori predicții mai bune. Au fost și situații și opuse, în care predicția a fost eronată, iar toate acestea sunt prezentate și ilustrate în teză. În subcapitolul 2.13.4 am descris algoritmul de urmărire bazat pe filtre de particule și o hartă de ocupare dinamică. Această abordare nu este una originală și a mai fost publicată, dar modul în care am extins-o și cum am construit partea de măsurare este unică.

Harta de măsurare este folosită pentru a genera particule noi, care sunt grupate în celule din care se pot extrage cuboide care reprezintă obstacolele detectate și urmărite în scenă. Această hartă de măsurare este de fapt o hartă de ocupare binară, construită din imaginea de drum segmentată de rețeaua neuronală convoluțională. Având măsurătorile în fiecare cadru, particulele pot fi create, deplasate sau distruse, astfel se pot urmări obstacolele. Incertitudinea măsurătorii este determinată de distanța până la obstacole care se este determinată în imaginea IPM. În ultimul pas, particulele sunt grupate în funcție de similaritățile lor (proximitate, vectorii de viteză) și apoi se extrag cuboide din ele. O altă contribuție față de lucrările publicate în trecut o reprezintă rafinarea hărții de ocupare, prin procesarea imaginii segmentate: am realizat un algoritm care să unească vârfurile din histogramă care aparțin unui obstacol. Efectul asupra creării particulelor este bine ilustrat și exemplificat în subcapitolul 2.13.5 (a). O altă contribuție constă în rafinarea parametrilor camerei prin generarea unor imagini IPM având unghiuri de înclinare diferite și compararea hărților de ocupare cu cea actuală din filtrul de particule. În secțiunea 2.13.6 am descris o modalitate de a estima orientarea locală a obiectelor și dimensiunile lor folosind o rețea neuronală convoluțională. Implementarea aceasta nu este unică, dar contribuția constă în integrarea cu sistemul de percepție bazat pe filtrul de particule. Astfel, orientarea și dimensiunea cuboidelor extrase din filtrul de particule sunt ajustate folosind CNN-ul. Alte detalii de implementare a tehnologiilor folosite și arhitectura sistemului sunt prezentate în secțiunea 2.13.7. Capitolul 2 se încheie cu concluzii și o listă de publicații rezultate.

În capitolul 3 am descris tehnici de calibrare a camerei. În primul subcapitol (2.1) am prezentat modelul camerei și apoi parametri intrinseci (2.2) și cei extrinseci (2.3). Principalul obiectiv în urma calibrării este de a cunoaște parametrii camerei pentru a calcula matricea de proiecție, care poate fi folosită pentru a genera imagini cu efectul de perspectivă eliminat, o vedere periferică (de sus) a scenei de drum. În secțiunea 3.4 am realizat un studiu actual al tehnicilor de calibrare a camerelor, iar în 3.5 am prezentat modalități de calculare automată a distanței focale. Secțiunea 3.6 conține algoritmii de calculare a imaginilor IPM, iar în 3.7 am prezentat o interfață intuitivă pentru asistarea calibrării manuale a parametrilor extrinseci. O altă contribuție este dată de implementarea unui algoritm de auto-calibrare a parametrilor vectorului de translație, prin analizarea “offline” unei secvențe cadru cu cadru. Înălțimea camerei față de sol este ajustată astfel încât lățimea benzii de circulație detectată în imaginea IPM să corespundă cu valoarea în pixeli corespunzătoare unei lățimi standard de 3.5 metri. Tot aici am prezentat pe scurt modul de determinare a unei benzi de circulație. În secțiunea 3.9 am descris tehnici pentru determinarea punctului de fugă din imaginile capturate în traficul rutier. Am prezentat abordări existente și apoi o tehnică proprie, bazată pe procesarea de imagini în care sunt analizate orientările și magnitudinile punctelor de muchii care aparțin marcajelor rutiere. Folosind o schemă de votare proprie am realizat o hartă de voturi din care este extras punctul de fugă final. Rezultatele au fost suficient de bune încât să creez o bază de date de imagini și puncte fuga, care apoi am folosit-o pentru a antrena o rețea neuronală convoluțională care să ofere predicții ale coordonatelei x și y din imagine a punctului de fugă (secțiunea 3.10). Metoda a fost publicată la o conferință internațională. Am folosit mai multe tehnici de augmentare a bazei de date și apoi am extins această abordare prin folosirea unei rețele antrenată pentru predicția coordonatei x, și aceeași rețea antrenată pentru coordonata y a punctului de fugă, iar rezultatele față de varianta inițială sunt mai bune (secțiunea 3.11). Totodată, am implementat și testat această soluție și pe dispozitive mobile Android (3.11.4). În secțiunea 3.12 am descris o soluție de auto-calibrare a unei camere. Din analiza traiectoriei unui vehicul am determinat înălțimea camerei față de sol și unghiul de înclinare folosind un

filtru Kalman extins, dar și punctul de fugă care este folosit pentru corecția imaginii IPM. Întreaga soluție a fost publicată la o conferință internațională și implementată atât pe sisteme desktop, cât și pe dispozitive mobile Android. Capitolul 3 se încheie cu partea de concluzii și contribuții originale și cu o listă de lucrări publicate.

Contribuții

Principalele contribuții originale descrise în capitolul doi sunt: crearea unui sistem propriu de achiziție de date folosind senzorii disponibili pe un dispozitiv mobil inteligent (camera foto, senzori de poziție prin satelit, accelerometru, giroscop, etc.), crearea unor baze de date și modalitatea lor de a vizualiza aceste date pe sisteme desktop și implementarea unui sistem de percepție și urmărire a obstacolelor în traficul rutier bazat pe un filtru de particule și rețele neuronale convoluționale.

Pentru achiziția de imagini este folosită cameră foto principală a dispozitivului (în general cea din spate), deoarece aceste camere au un senzor mai bun, iar sistemul de achiziții va stoca și informații adiționale din timpul deplasării vehiculului: date despre accelerația și viteza de deplasare, poziția (exprimată în coordonate de latitudine și longitudine), orientarea față de nordul magnetic (din magnetometru) și orientarea locală a dispozitivului (obținută din giroscop). Toate aceste informații sunt stocate în memoria internă și sincronizate folosind un marcaj de timp (“timestamp”).

Am realizat un sistem de percepție bazat pe o singură cameră, combinând rețele neuronale convoluționale pentru a identifica obiecte din scenă și apoi folosind filtre de particule pentru a le urmări în timp. Primul pas constă în implementarea unei rețele artificiale pentru segmentarea semantică, având la bază arhitectura rețelei U-NET [Ronneberger2015]. Rețeaua neuronală convoluțională de tip U-Net modificată are următoarea structură: 5 straturi de codificare, un strat central și 5 straturi de decodificare. În urma antrenării și evaluării rețelei se poate remarca și performanța bună a predicției în scenarii de trafic complexe unde sunt multe vehicule în scenă (figura 1). Această delimitare precisă a ce este carosabil și ce nu este foarte utilă pentru crearea unui sistem de percepție monocular unde nu avem informații despre adâncime sau alte informații din scenă.

Figura 1. Exemplu de segmentare semantică folosind rețeaua artificială.

Imaginile segmentate sunt integrate într-un framework probabilistic, bazat pe o hartă dinamică a scenei din jurul autovehiculului. Din detecția zonei libere de drum, putem considera că tot ce nu este drum reprezintă un posibil obstacol și poate fi interpretat și urmărit în timp. Astfel, această soluție este utilă pentru a detecta orice tip de obstacol, nefiind limitată doar la pietoni sau vehicule. Pentru implementarea acestei funcționalități am ales folosirea unei hărți de ocupare bazată pe un filtru de particule. Harta de ocupare binară este generată din imaginea segmentată de CNN în care este eliminat efectul de perspectivă (IPM), iar apoi este analizată folosind un algoritm care scanează zona din fața vehiculului folosind raze cu originea în punctul în care e montată camera foto. Deoarece obiectele în imaginile IPM vor fi dispuse radial, se poate folosi această proprietate.

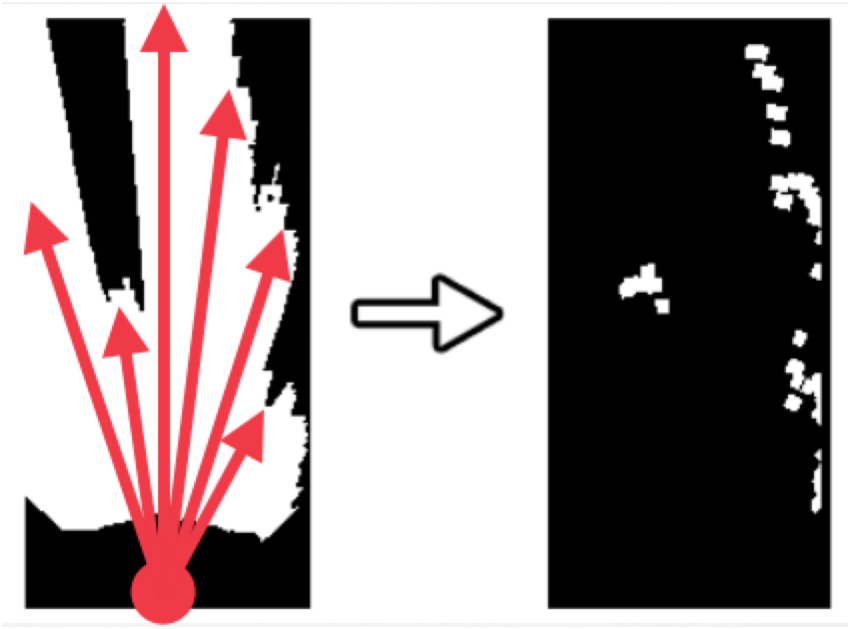

Figura 2. Harta binară de ocupare (dreapta), obținută din imaginea de drum segmentată (stânga). Sunt ilustrate și o parte din razele de-a lungul cărora se caută tranziții de intensitate.

Filtrul de particule utilizează o hartă de ocupare dinamică, care nu e o implementare originală (a fost publicată inițial în [Dănescu2011]), dar a fost extinsă (harta de ocupare este de dimensiune dublă pentru a permite urmărirea și a obstacolelor care părăsesc zona vizibilă a scenei), adaptată și îmbunătățită pentru imagini monoculare: am rafinat detecția de obstacole prin analizarea histogramelor radiale, dar și prin generarea mai multor imagini IPM pentru rafinarea parametrilor camerei. Primul pas al algoritmului de urmărire este cel de predicție și este aplicat pe fiecare particulă. Poziția particulelor din set este modificată în funcție de viteza lor (ele se mișcă în concordanță cu vectorul propriu de viteză) și în funcție de parametri de mișcare ai vehiculului propriu (viteza de deplasare și rata de rotație - “yaw rate”) citiți din senzorii vehiculului sau ai dispozitivului mobil: GPS, accelerometru și giroscop. Al doilea pas al algoritmului este procesul de actualizare, mai specific procesarea informațiilor măsurate și se bazează pe harta binară de ocupare a celulelor creată prin procesarea imaginii IPM. Informațiile din măsurare sunt folosite pentru a pondera particulele și apoi pentru a le reeșantiona într-un singur pas. Prin ponderare și reeșantionare, particulele dintr-o celulă se

pot multiplica sau distruge. Ultimul pas este de a calcula viteza și apoi de a estima probabilitatea unei celule de a fi ocupată de un obstacol.

Deoarece în celulele individuale ale hărții de ocupare se află particule, ele pot fi grupate în funcție de asemănările lor, astfel încât se pot extrage cuboide din ele. Algoritmul de grupare ia în calcul proximitatea celulelor, vectorii lor de viteză și deplasarea lor pentru a extrage regiuni conectate care vor reprezenta un obiect. În final, din aceste zone conectate este extrasă o formă dreptunghiulară orientată, iar din aceasta va fi construit cuboidul 3D (înălțimea cuboidului este setata fixă având 1,5 metri pentru toate obiectele).

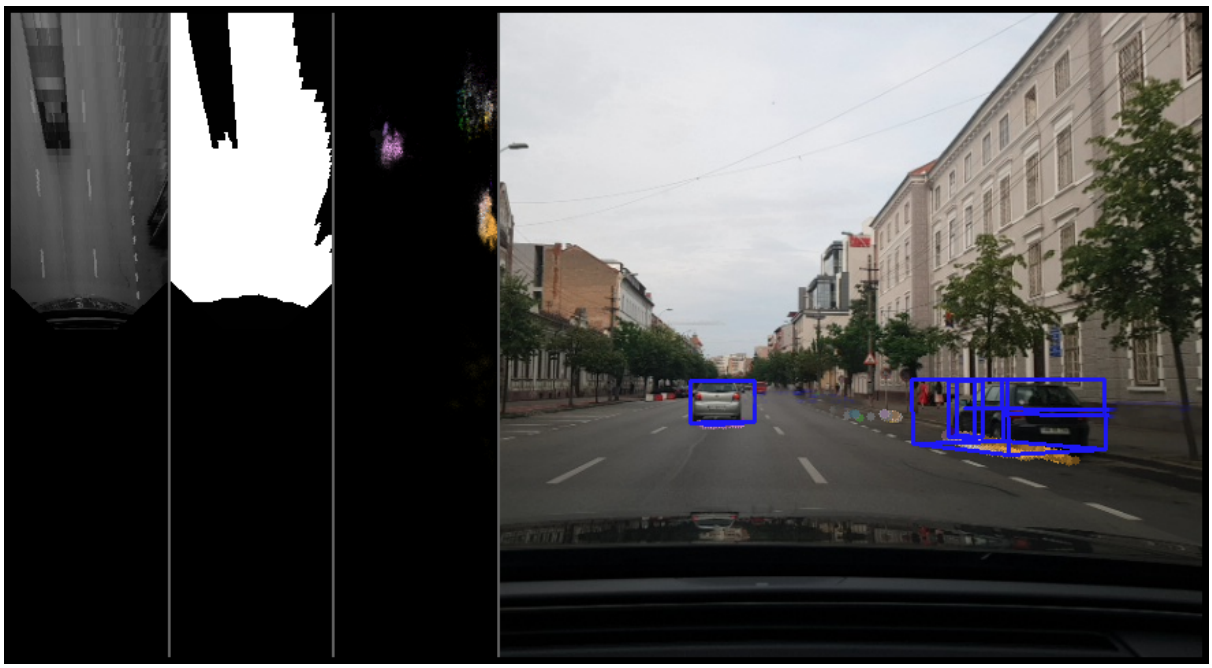

Figura 3. Framework-ul probabilistic de percepție a scenei.

Figura 3 ilustrează etapele de procesare din cadrul framework-ului probabilistic. În imaginea din stânga se pot observa etapele procesării și crearea hărții de ocupare dinamică: prima imagine reprezintă scena cu efectul de perspectivă eliminat (IPM), în a doua este ilustrată imaginea segmentată IPM (obținută din rețeaua neuronală convoluțională), a 3-a imagine ilustrează particulele și culoarea lor este codificată în funcție de vectorii de viteză (la fel ca în [Dănescu2011]: nuanța culorii reprezintă direcția de mers, saturația reprezintă magnitudinea vitezei, iar intensitatea codifică probabilitatea de ocupare). În imaginea din dreapta este prezentat rezultatul procesării și grupării particulelor și a celulelor în obstacole. Particulele sunt de asemenea proiectate și ilustrate în imaginea de perspectivă a scenei.

O contribuție și îmbunătățire a framework-ului față de lucrarea publicată în [Danescu2016], este pre-procesarea imaginii de măsurare. Pentru a îmbunătăți detecția de vehicule din zona segmentată (“free-space”) am folosit o tehnică având la bază unele idei din lucrarea [Bertozzi1998].

Obstacolele ating asfaltul doar cu roțile, iar acest lucru este amplificat în imaginile IPM, de unde rezultă o percepție falsă a distanței până la obstacole (în special la cele îndepărtate). Am rezolvat această problemă prin analiza histogramei radiale a fiecărui vehicul și prin unirea vârfurilor, adică umplerea zonei goale dintre roțile unui vehicul. Vârfurile din histogramă sunt

identificate prin calcularea maximelor locale și apoi sunt procesate pentru a identifica perechi de linii care descriu marginile unui obiect. Având aceste perechi de linii care vor corespunde unui obiect, următor pas a fost umplerea zonei dintre vârfuri. Această operație are rolul de a îmbunătăți semnificativ procesul de măsurare din filtrul de particule și al hărții de ocupare, deoarece este importantă determinarea zonei de contact și a distanței de la vehiculul propriu la restul de vehicule din scenă observată. Efectul aplicării acestor operații este ilustrat în figura următoare (figura 4), unde se poate observa o îmbunătățire semnificativă a modului în care sunt generate particulele în harta de ocupare (ele vor forma un obiect asemănător cu cel real).

Figura 4. Îmbunătățirea hărții dinamice de ocupare. Perechi de imagini cu particulele generate în funcție de imaginea de măsurătoare (segmentată) utilizată.

O altă contribuție constă în rafinarea parametrilor camerei. Parametri extrinseci referitori la unghiul de înclinare și cel de girație se pot corecta și rafina în timpul procesării. Pentru fiecare imagine dintr-o secvență se generează un număr total de *N* imagini cu efectul de perspectivă eliminat (IPM), fiecare având un unghi de înclinare diferit. În mod similar se procedează și pentru a genera *N* imagini IPM având *N* unghiuri de girație diferite. Acest pas se poate aplica după ce se face o auto-calibrare a camerei. În sistemul de percepție monocular am generat *N=10* imagini IPM folosind variații de 0.1 grade în jurul valorii unghiului de înclinare calculat folosind filtrul Kalman extins (secțiunea 3.12.2). În cele 10 imagini IPM generate sunt de fapt detectate obstacolele din scenă folosind algoritmul descris în secțiunea 3.6. Aceste imagini binare sunt comparate cu harta de ocupare dinamică a filtrului de particule (generată din imaginea segmentată din CNN) și se calculează procentului de pixeli care se suprapun. Imaginea cu un procent mai mare de un prag fix reprezintă o potrivire mai bună a datelor asupra scenei urmărite, ceea ce înseamnă că se poate ajusta unghiul de înclinare pentru a mări precizia și robustețea sistemului. În mod analog am procedat și pentru unghiul de girație. Harta de ocupare din filtrul de particule este folosită pentru validare și ajustare a parametrilor deoarece va estima cât mai apropiat deplasarea vehiculelor în scenă și va fi mai puțin influențată de trepidații și mișcări ale camerei în timpul condusului.

De altfel, am implementat o versiune proprie a unei rețele existente [Mousavian2017] pentru a oferi predicții asupra dimensiunii obiectelor detectate în scenă, dar și asupra orientării lor. Aceste informații sunt fuzionate cu cele din filtrul de particule pentru a îmbunătăți robustețea și precizia sistemului și reprezintă o contribuție importantă pentru sistemul de percepție monocular. Pentru determinarea dimensiunii si orientării vehiculelor am ales să

utilizez o rețea similară cu [Mousavian2017], cu mici modificări ale funcției de cost prezentată în articol. Rețeaua neuronală convoluțională are primele 5 straturi la fel cu cele ale rețelei VGG16 [Simonyan2014]. Am folosit tehnica de învățare prin transfer și am inițializat ponderile primelor 5 straturi cu cele are rețelei VGG16 antrenată pentru clasificarea obiectelor din imagini. Ieșirea ultimei convoluții este apoi folosită pentru regresia dimensiunilor, orientării și a unei probabilități a orientării. Estimarea dimensiunilor 3D și a orientării unui obiect este realizată prin extinderea unei rețele de detecție a obiectelor care oferă dreptunghiuri de încadrare în interiorul cărora se află obiectele detectate. Principala constrângere este dată de faptul că obiectul tridimensional din scenă se află în interiorul dreptunghiului 2D din imagine, astfel că proiecția 3D a acestuia se va afla în interiorul aceluiași dreptunghi (chenar) de încadrare obținut din detectorul de obiecte. Antrenarea unei rețele (SSD MobileNet [Wei2016], [Howard2017]) pentru detecția de obiecte este prezentată pe larg în secțiunea 3.12.1. Figura 5 prezintă rezultatele detecției de vehicule folosind rețeaua SSD MobileNet, iar în figura 6 este afișată predicția orientării și a dimensiunii pentru fiecare vehicul.

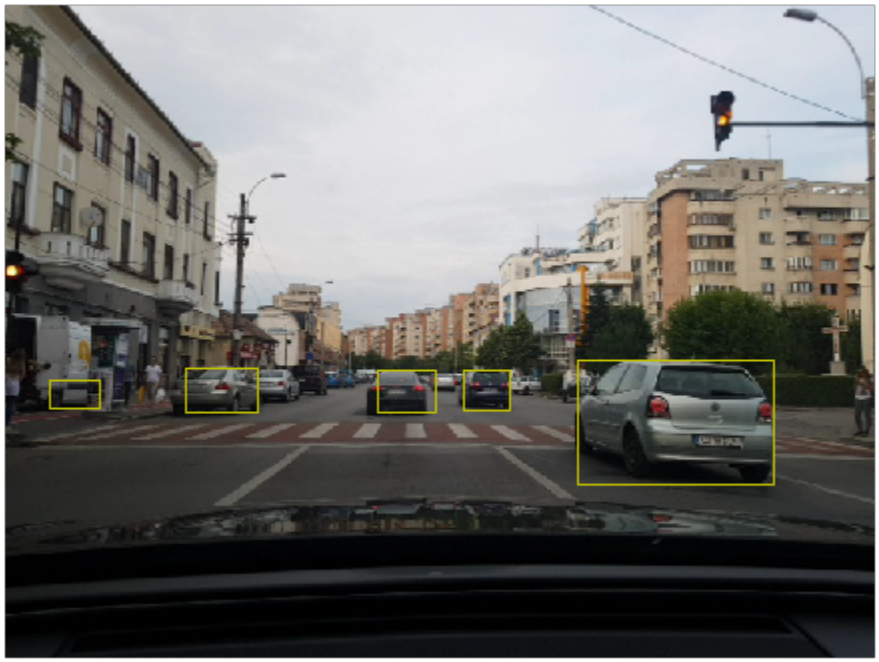

Figura 5. Imaginea de intrare având obstacole detectate și marcate prin chenare de încadrare obținute din rețeaua SSD MobileNet.

Figura 6. Rezultatul predicției rețelei de regresie a orientării și dimensiunii vehiculelor.

Dimensiunea și orientarea locală a unui vehicul este integrată în filtrul de particule, mai specific în algoritmul de grupare al particulelor care extrage cuboide din particulele din celule asemănătoare. Corespondența dintre vehiculele detectate de CNN și cele urmărite de filtrul de particule se face prin calcularea coeficientului Sorensen Dice (IoU). Am calculat IoU între latura din față sau cea din spate a cuboidului și chenarele de încadrare din rețeaua neuronală convoluțională SSD MobileNet. Situațiile cu un scor IoU mai mare de 0.5 sunt considerate valide și se modifică dimensiunile cuboidului și orientarea cu cele din rețeaua artificială din acest capitol. Astfel se obține fuziunea datelor dintre rețele convoluționale neuronale și urmărirea prin filtrul de particule.

Rezultatul fuziunii datelor este ilustrat în figura 7 unde este afișată o vedere de sus a scenei în care este afișat rezultatul extracției de cuboide direct din filtrul de particule în paralel cu varianta ajustată și îmbunătățită.

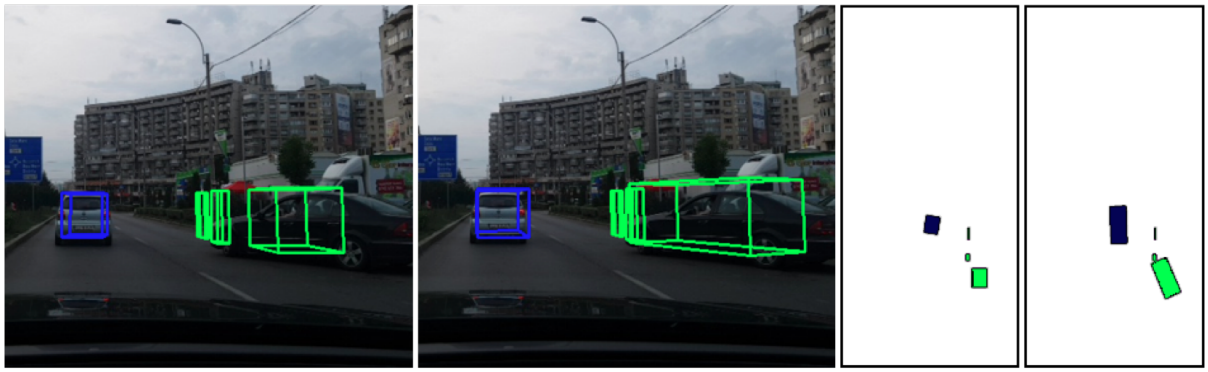

Figura 7. Fuziunea datelor folosind rețeaua CNN pentru orientarea obiectelor și a dimensiunilor. Prima imagine reprezintă cuboidele extrase din filtrul de particule, în centru sunt ilustrate cuboidele cu orientarea și lungimea lor ajustată din CNN. Ultima imagine reprezintă o vedere de sus a scenei cu aceleași obiecte (normale din filtrul de particule vs. ajustate din CNN).

Scopul sistemului de percepție este de a utiliza atât viziune artificială prin procesare de imagini, cât și folosirea rețelelor convoluționale neuronale și integrarea și fuziunea acestor module de procesare.

Contribuțiile originale rezultate în urma soluțiilor propuse în capitolul trei sunt: determinarea automată a distanțelor focale pe dispozitive mobile, crearea unui sistem de calibrare manual și intuitiv pe dispozitive mobile, unde utilizatorul poate ajusta parametri extrinseci ai camerei. O altă contribuție este crearea unui sistem de calibrare “offline” prin analiza imaginilor din trafic și determinarea lățimii unei benzi de circulație. Banda de circulație este folosită pentru a avea o corespondență între scena 3D a lumii și spațiul 2D al imaginii. Prin această metodă se determină înălțimea camerei față de sol (parte din vectorul de translație al camerei).

Una din contribuțiile principale se referă la elaborarea unui algoritm nou de calculare a punctului de fugă din imaginile din traficul rutier. Aceste imagini prezintă un efect de perspectivă în care liniile benzilor de circulație se vor intersecta în punctul de fugă. Algoritmul exploatează punctele din imagini care au o anumită magnitudine și orientare și le folosește pentru a construi o hartă de voturi care este analizată și validată folosind ferestre glisante. Punctul din fereastra cu cele mai multe voturi va reprezenta un punct de fugă. Figura 8 ilustrează în stânga imaginea cu harta de voturi și cu ferestrele de validare folosite, iar în dreapta este afișat rezultatul final al algoritmului cu punctul de fugă calculat.

Figura 8. Stânga: ferestrele de votare alese pentru determinarea unui maxim. Dreapta: rezultatul final al algoritmului cu punctul de fugă calculat.

Având această metodă am construit o bază de date de imagini și puncte de fugă, pe care am folosit-o pentru a antrena o rețea neuronală convoluțională cu scopul de a avea predicții din rețea asupra locației unui punct de fugă. Rețeaua are ca intrare imaginea scenei redimensionată, iar ieșirea va fi formată din cele 2 coordonate ale punctului de fugă. Structura rețelei este următoarea: 5 niveluri convoluționale cu număr diferit de filtre și având kernel de dimensiune variabilă, iar ultimele niveluri sunt reprezentate de 3 niveluri complet conectate (“fully-connected”). Ultimul strat complet conectat cu cele 2 valori ale neuronilor reprezintă valorile coordonatelor x și y ale punctului de fugă prezis. Un exemplu de predicție a punctului de fugă este ilustrat în figura următoare:

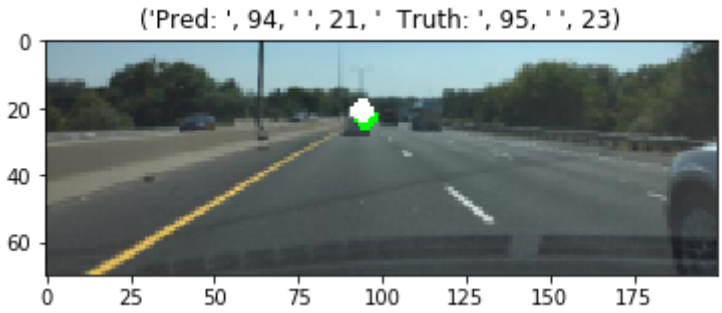

Figura 9. Exemplu în care punctul de fugă calculat de CNN (cercul alb) este mai bun decât cel obținut prin algoritmi tradiționali (cercul verde).

Rezultatele acestei abordări au fost foarte bune, iar apoi întreaga metodă a fost extinsă pentru a utiliza mai multe imagini de antrenare și două rețele neuronale convoluționale diferite: una pentru a prezice poziția pe coordonata orizontală, iar cealaltă pentru a prezice poziția punctului de fugă pe coordonata verticală. Am prezentat o comparație între metode, iar folosirea a două rețele a fost mai eficientă și a adus îmbunătățiri semnificative. Această soluție a fost implementată și pe dispozitive mobile, iar figura 10 prezintă modul de lucru pentru testare și dezvoltare al algoritmului într-un mediu controlat, unde dispozitivul mobil este amplasat într-un punct fix și orientat spre un ecran unde rulează o secvență de test.

Determinarea punctului de fugă din imagini de trafic rutier folosind rețele neuronale convoluționale reprezintă o contribuție originală.

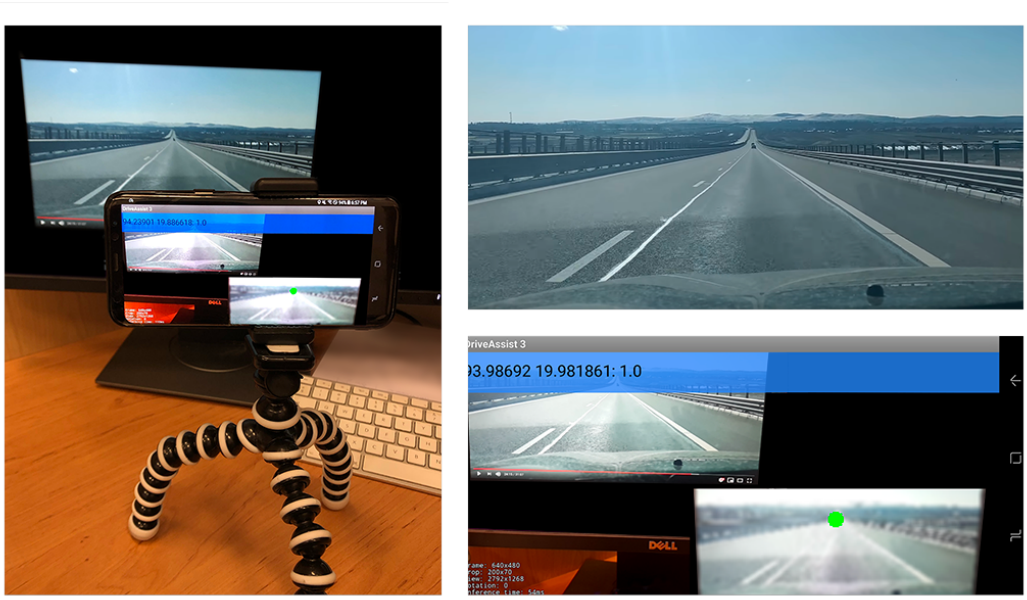

Figura 10. Calculare punct de fugă direct pe dispozitivul mobil. În dreapta sus este imaginea sursă, iar în dreapta jos este o captură de ecran de pe dispozitivul mobil.

Ultima secțiune constă în implementarea unei soluții de auto-calibrare folosind traiectoria vehiculelor. Detecția vehiculelor am realizat-o prin folosirea unei rețele convoluționale neuronale, mai specific rețeaua SSD și MobileNet care oferă un raport bun între precizia detecțiilor și viteza de execuție (timpul de predicție pe un cadru). Rezultatul rețelei va fi o listă de dreptunghiuri de încadrare corespunzătoare fiecărui vehicul din scenă. Aceste dreptunghiuri sunt urmărite de la un cadru la altul, iar pentru calibrare este necesar doar un anumit număr de vehicule detectate. Din coordonatele dreptunghiului de încadrare sunt extrase lățimile obiectelor din scenă și coordonata lor pe axa verticală (axa y a imaginii). Următorul pas constă în determinarea relației liniare dintre lățimi și coordonata verticală. Inițial am calculat relația liniară folosind RANSAC, iar apoi prin utilizarea algoritmului Hough modificat. Algoritmul de estimare a înălțimii camerei și a unghiului de înclinare folosește un filtru Kalman extins, iar vectorul de măsurare Z este format din lățimea w_1 și w_2 a vehiculelor la coordonata orizontală (axa x) $v_1 = 310$ și $v_2 = 360$, determinate din relația liniară calculată în pasul precedent. Am ales 2 puncte în spațiul 3D care să reprezinte un vehicul teoretic cu o lățime fixă (standard) de 1750 mm aflat la o distanță de $Z_1 = 7000$ mm și încă un set de puncte 3D pentru un vehicul aflat la o distanță $Z_2 = 20000$ mm. Aceste 4 puncte care descriu 2 vehicule la 2 distanțe diferite sunt proiectate în spațiul 2D al imaginii, iar perechile de coordonate laterale vor forma două linii care se vor intersecta într-un punct de fugă. În următorul pas se vor calcula coordonatele punctelor la indexul $v_1 = 310$ și $v_2 = 360$. Practic, algoritmul EKF va încerca să facă o potrivire între aceste 4 puncte teoretice ale celor 2 vehicule și cele 4 puncte ale vehiculelor din vectorul de măsurare. Algoritmul va ajusta înălțimea camerei și unghiul de înclinare până când rezultatul este stabil și se ajunge la o potrivire cât mai bună (în general sunt de ajuns 10 iterații ale filtrului EKF).

Algoritmul bazat pe rezultatele detecției de vehicule de la o rețea convoluțională neuronală este capabil să estimeze parametrii extrinseci ai unei camere monoculare în raport

cu un cadru de referință bazat pe un vehicul standard. Sistemul funcționează precis și robust când secvențele sunt lungi și în diverse situații de trafic, când sunt suficiente date disponibile. Deoarece detecția vehiculelor influențează estimarea parametrilor, în viitor propun să îmbunătățesc acest aspect.

Prin integrarea tuturor soluțiilor, am reușit implementarea unui sistem de percepție monoculară, iar fluxul de procesare este ilustrat în figura 11.

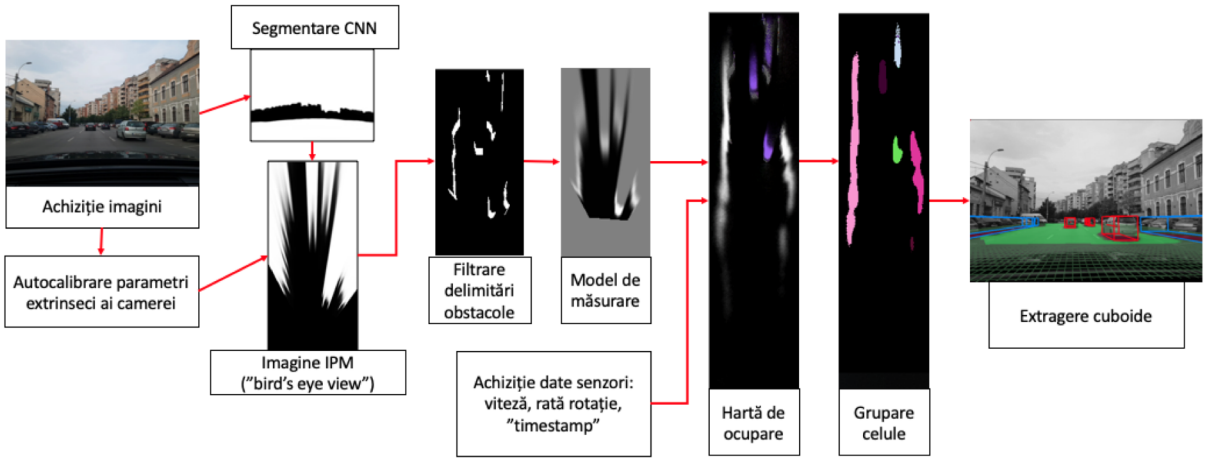

Figura 11. Fluxul de procesare al sistemului de percepție monoculară.

Primul pas constă în achiziția imaginilor și auto-calibrarea parametrilor extrinseci ai camerei, care sunt folosiți pentru a genera o imagine cu efectul de perspectivă eliminat din imaginea segmentată de rețeaua neuronală. Următorul pas constă în detecția delimitării obstacolelor, iar din această imagine se generează un model de măsurare care este folosit pentru a crea particule într-o hartă de ocupare. În ultimul pas celulele asemănătoare sunt grupate, iar din ele se extrag cuboide care sunt afișate în imaginea inițială a scenei de drum.

În concluzie, teza introduce contribuții în domeniul roboților și a vehiculelor autonome în care percepția este realizată folosind o singură cameră. Algoritmii de percepție au la baza algoritmi și tehnici de procesare de imagini, dar și algoritmi de inteligență artificială prin folosirea rețelelor neuronale convoluționale. Teza prezintă o soluție completă de percepție a scenei de trafic rutier, începând de la achiziția imaginilor și până la procesarea și afișarea rezultatelor. Sistemul monocular este capabil să se calibreze singur după un anumit timp, iar datele procesate pot fi folosite și integrate în sisteme inteligente de asistate a șoferului, cum ar fi sisteme de prevenție a accidentelor care pot acționa în locul șoferului său pot oferi doar avertizări acustice sau sonore. Soluțiile prezentate în teză au fost implementate atât pe sisteme mobile (smartphone-uri sau tablete), cât și pe PC-uri (desktop sau laptop). Prin metodele descrise în această lucrare se poate aduce un beneficiu major prin îmbunătățirea siguranței în traficul rutier, având avantajul că astfel de sisteme se pot integra și pe vehicule existente la un cost redus (prin folosirea unei singure camere video pentru analiza scenei). În concluzie, pe viitor îmi doresc să implementez împreună cu domnul profesor Dănescu un sistem complet care să fie ușor de integrat în orice vehicul.

Publicații

Următoarele publicații au rezultat din metodele propuse în cadrul acestui teze de doctorat:

1. **Razvan Itu**, Radu Danescu, "An Efficient Obstacle Awareness Application for Android Mobile Devices", International Conference on Intelligent Computer Communication and Processing, pp. 157-163, 2014.

2. Radu Danescu, **Razvan Itu**, Andra Petrovai, "Sensing the Driving Environment with Smart Mobile Devices", IEEE International Conference on Intelligent Computer Communication and Processing, pp. 271-278, 2015.

3. Radu Danescu, **Razvan Itu**, Andra Petrovai, "Generic Dynamic Environment Perception Using Smart Mobile Devices", Sensors, Vol. 16, No. 10, 2016, Art. No. 1721.

4. Radu Danescu, Andra Petrovai, **Razvan Itu**, Sergiu Nedevschi, "Generic Obstacle Detection for Mobile Devices Using a Dynamic Intermediate Representation", Advances in Intelligent Systems and Computing, vol. 427, 2016, pp. 629-639.

5. **Razvan Itu**, Diana Borza, Radu Danescu, "Automatic extrinsic camera parameters calibration using Convolutional Neural Networks", 2017 IEEE 13th International Conference on Intelligent Computer Communication and Processing (ICCP 2017), pp. 273-278.

6. **Razvan Itu**, Radu Danescu, "Machine Learning Based Automatic Extrinsic Calibration of an Onboard Monocular Camera for Driving Assistance Applications on Smart Mobile Devices", 2018 International Forum on Advanced Microsystems for Automotive Applications, pp. 16-28.

7. Radu Danescu, **Razvan Itu**, ”Camera Calibration for CNN based Generic Obstacle Detection”, 2019 Portuguese Conference on Artificial Intelligence (EPIA), acceptat iunie 2019.

Citări

Google Scholar:
30 citări
H-index 4

ISI web of science:
12 citări
H-index 2

Bibliografie

[Bertozzi1998] - M. Bertozzi, A. Broggi, “GOLD: a Parallel Real-Time Stereo Vision System for Generic Obstacle and Lane Detection”, IEEE Transactions on Image Processing, vol. 7, no. 1, pp. 62-81, 1998.

[Danescu2011] - R. Danescu, F. Oniga, S. Nedevschi, “Modeling and Tracking the Driving Environment with a Particle-Based Occupancy Grid”, IEEE Transactions on Intelligent Transportation Systems, volume 12, no. 4, pp. 1331-1342, 2011.

[Danescu2016] - R. Danescu, R. Itu, A. Petrovai, "Generic Dynamic Environment Perception Using Smart Mobile Devices", Sensors, vol. 16, no. 10, art. no. 1721, 2016.

[Howard2017] - A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications”, arXiv:1704.04861, 2017.

[Ronneberger2015] - O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, Vol. 9351, pp. 234-241, 2015.

[Simonyan2014] - K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, arXiv:1409.1556, 2014.

[Wei2016] - L. Wei, et. al., “SSD: Single Shot MultiBox Detector”, European Conference on Computer Vision, pp. 21-37, 2016.