



Universitatea Politehnica din București
Facultatea de Automatică și Calculatoare
Departamentul de Calculatoare



Teză de Doctorat - Rezumat

O Abordare Predictivă pentru Urmărirea Persoanelor în Aplicații Autonome

Autor:

Alexandra Ștefania Ghiță

Conducător de Doctorat:

Prof. Dr. Ing. Adina Magda Florea

Comisia de Doctorat

Președinte	Prof.Dr.Ing. Florin Pop	Universitatea Politehnica din București
Conducător Doctorat	Prof.Dr.Ing. Adina Magda Florea	Universitatea Politehnica din București
Membru	Prof.Dr.Ing. Rodica Potolea	Universitatea Tehnică din Cluj-Napoca
Membru	Prof.Dr.Ing. Costin Bădică	Universitatea din Craiova
Membru	Prof.Dr.Ing. Ștefan Trășan-Matu	Universitatea Politehnica din București

BUCUREȘTI

2022

Abstract

În prezent, importanța dispozitivelor cu operare autonomă crește odată cu creșterea numărului de aplicații care pot beneficia de pe urma lor. Aplicațiile autonome, cum ar fi platformele robotizate sau mașinile autonome, funcționează fără intervenția umană în medii în care oamenii se mișcă sau își desfășoară activitățile zilnice.

O componentă critică pentru dispozitivele autonome este capacitatea de a urmări și reidentifica aceleași persoane într-o secvență de imagini, într-un interval scurt de timp, pentru a genera comportamente sigure și robuste. În această teză, introducem un sistem de urmărire și reidentificare a persoanelor, în timp real, bazat pe o metodă de predicție a traiectoriei. Combinăm rezultatul unei metode de predicție a traiectoriei cu o tehnică simplă de urmărire pentru a crea un sistem stabil și precis. Abordăm problema predicției traiectoriei prin introducerea unui sistem care încorporează informații semantice din mediu și influență socială de la ceilalți participanți în mișcarea fiecărui individ, pentru a prezice traiectoriile cele mai probabile. Evaluăm sistemele introduse luând în considerare mai multe studii de caz, și anume robotică socială, conducere autonomă și drone autonome. Pentru contextul roboticii sociale prezentăm un sistem modular, creat pentru scenarii de îngrijire și asistență. Sistemul pune la dispoziție utilizatorilor servicii robotice pentru vedere computațională, navigare, interacțiunea în limbaj natural și dialog, precum și recunoașterea activităților și structurarea generală a comportamentului robotului. Pentru toate cele trei studii de caz, efectuăm mai multe experimente pentru a valida abordarea noastră luând în considerare atât componenta de predicție a traiectoriei, cât și sistemul de reidentificare a persoanei. Ne concentrăm atât pe evaluarea calitativă, cât și pe cea cantitativă, prin integrarea unor seturi de date existente și pe baza unor imagini colectate de noi.

Cuprins

Abstract

1	Introducere	1
1.1	Motivație	1
1.2	Obiective	2
1.3	Schița Tezei	3
2	Definirea Problemei	5
3	Lucrări din Domeniu	7
4	Seturi de Date Relevante	9
5	Aplicația Robotică AMIRO pentru Roboți Sociali	11
5.1	Prezentare Generală și Arhitectură	11
5.2	Experimente	13
6	Sistemul de Predicție a Traectoriei	14
6.1	Prezentare Generală și Arhitectură	14
6.2	Evaluare și Validare	16
7	Sistemul de Urmărire și Reidentificare în Timp Real a Persoanelor	17
7.1	Prezentare Generală și Arhitectură	17
7.2	Evaluare și Validare	19
8	Validarea Sistemelor în Alte Contexte	20
8.1	Validarea în Contextul Mașinilor Autonome	20
8.2	Validarea în Contextul Aplicațiilor cu Vedere de Sus	21
9	Concluzii	23
9.1	Contribuții	24
9.2	Perspectivă	26
	Referințe	28

Introducere

Inteligența artificială se dezvoltă și se îmbunătățește constant, atingând noi praguri în toate domeniile. Spectrul de aplicații la care poate fi aplicată se extinde, având un impact semnificativ asupra oamenilor și asupra calității vieții acestora. Progresul realizat în cercetare permite dezvoltarea unor proiecte mai complexe, permițând dispozitivelor de operare autonome, care pot ajuta sau înlocui activitățile făcute de oameni, să fie mai atractive și mai robuste. Această teză abordează mai multe probleme în contextul aplicațiilor autonome, luând în considerare cerințele lor de procesare în timp real.

1.1. Motivație

Dispozitivele autonome reprezintă o tendință recentă în domeniul cercetării având în vedere multitudinea de locuri în care pot fi aplicate. Vehiculele autonome și roboții sociali sunt cele mai răspândite aplicații care sunt în prezent studiate și sunt îmbunătățite constant, având în vedere progresele în inteligența artificială. Oamenii tind să depindă din ce în ce mai mult de dispozitivele care le pot face viața mai ușoară, reducându-și o parte din sarcinile zilnice. Roboții sociali, și în special roboții pentru asistență, au o varietate de scopuri, deoarece pot fi folosiți ca asistenți pentru persoanele în vârstă, ghiduri turistice în muzee, informații despre punctele de control în magazine, colectori de date în spitale etc. Vehiculele autonome ajută prin reducerea stresului și a anxietății pe care oamenii le simt atunci când conduc. Fiecare dispozitiv care poate crește calitatea vieții persoanelor este în prezent îmbunătățit ca să genereze sisteme mai stabile și mai robuste. Dezvoltarea cercetării, alături de aplicabilitatea largă a unor astfel de sisteme, a influențat decizia de a implementa un proiect care să poată genera un sistem adecvat pentru dispozitivele autonome.

Provocările care apar în contextul dispozitivelor autonome provin dintr-o

varietate de factori: evenimente neprevăzute, defecțiuni ale senzorilor, informații limitate, înțelegerea utilizatorului etc. Unul dintre cei mai dificili factori pentru un dispozitiv autonom este chiar omul. Oamenii au o natură foarte complexă și imprevizibilă, așa că un astfel de dispozitiv trebuie să fie capabil să facă față schimbărilor bruște, pentru a putea reacționa în consecință atunci când apare o situație neașteptată. În contextul roboților de asistență socială, pe lângă natura comportamentală complexă, oamenii au multe variații în ceea ce privește aspectul fizic (trăsături faciale, înălțime, îmbrăcăminte etc.) sau voce (ton, accent, viteză), făcând interacțiunea mult mai dificilă. În contextul conducerii autonome, aspectul vizual se schimbă semnificativ datorită mișcării rapide a vehiculului. Capacitatea de a depăși aceste provocări pentru a obține un sistem autonom sigur a fost unul dintre principalele motive în alegerea acestui proiect.

Având în vedere utilizarea generală a unui sistem autonom, o componentă pentru detectarea și urmărirea persoanelor este esențială. Sistemele autonome în general sunt concepute pentru a ajuta oamenii și pentru a reduce o parte din responsabilitățile acestora. Ca urmare, sistemele trebuie să cartografieze mediul în care operează și să înțeleagă comportamentele agenților din apropiere, pentru a-și executa sarcinile în siguranță și cu erori minime. Componenta de urmărire a persoanelor este relevantă din mai multe motive: poate reidentifica aceeași persoană într-o secvență de imagini, poate analiza comportamentul pentru a extrage tipare, poate prezice mișcarea pentru a preveni eventualele complicații. Importanța și varietatea funcțiilor pe care le implică o astfel de componentă, au determinat proiectarea unui sistem de urmărire a persoanelor în timp real care să se potrivească cerințelor unui dispozitiv autonom.

1.2. Obiective

Teza se concentrează pe implementarea sistemelor potrivite pentru aplicații autonome, în special pentru roboții de asistență socială. Scopul nostru este să integrăm un sistem propriu de urmărire a persoanelor într-o aplicație pe care am proiectat-o pentru dispozitive robotice. Aplicația robotică integrează mai multe capacități pentru a obține un sistem competent. Obiectivul său este să fie o platformă generală, în care pot fi integrați o varietate de roboți sociali în scenarii de asistență robotică. Oferă capacități în ceea ce privește înțelegerea vizuală, navigarea și interacțiunile vocale, care sunt combinate pentru a defini comportamente complexe. Componenta de urmărire a persoanelor îmbunătățește

funcționalitatea unei astfel de platforme. Obiectivul nostru este de a proiecta un sistem care nu se limitează doar la aplicații de robotică socială, ci poate fi aplicat pentru orice dispozitiv autonom, inclusiv pentru mașini și drone autonome. De asemenea am proiectat o componentă de predicție a traiectoriei care estimează traiectoriile viitoare ale oamenilor, pe care am integrat-o în componenta de urmărire a persoanelor, pentru a crea un sistem precis și robust.

În consecință, principalele obiective ale proiectului sunt:

1. Dezvoltarea unei aplicații robotice care integrează capabilități generale pentru o platformă robotică, pentru a crea comportamente complexe care pot fi aplicate în scenarii de robotică socială.
2. Proiectarea și implementarea unei metode de predicție în timp real a traiectoriei oamenilor care integrează mișcarea persoanei, influența socială și informațiile din mediu.
3. Proiectarea și implementarea unui sistem de urmărire și reidentificare a persoanelor în timp real bazat pe metoda de predicție a traiectoriei oamenilor.
4. Validarea sistemelor propuse în multiple contexte autonome: robotică socială, mașină cu conducere autonomă, drone.

1.3. Schița Tezei

Teza începe prin a prezenta problema pe care o abordează proiectul și detaliază lucrările relevante asociate cu problema, împreună cu seturile de date aferente. Continuă prin prezentarea aplicației robotice introduse pentru crearea de comportamente complexe pentru interacțiunile cu utilizatorii. Se explică apoi cele două sisteme introduse pentru analizarea comportamentelor oamenilor, și anume predicția traiectoriei oamenilor și urmărirea persoanelor, validate în contextul roboticii sociale. În cele din urmă, teza prezintă validarea sistemelor în alte contexte autonome: mașini autonome și drone.

Schița tezei prezentată capitol cu capitol este descrisă mai jos.

- Capitolul 2 definește problemele abordate de această teză din punctul de vedere al relevanței și cercetării. Capitolul definește problemele roboticii sociale, predicția traiectoriei oamenilor și urmărirea oamenilor, alături de provocările de cercetare asociate.

- Capitolul 3 raportează o analiză extinsă a lucrărilor existente pentru problemele abordate în această teză. Capitolul analizează diferite sisteme existente, prin prezentarea avantajelor și dezavantajelor corespunzătoare în contextul aplicațiilor autonome.
- Capitolul 4 prezintă mai multe seturi de date existente pentru aplicații autonome. Analizează seturile de date din punctul de vedere al calității, rezoluției și unghiului de vedere. De asemenea, prezintă datele pe care le-am colectat pentru a ne evalua sistemul în scenarii particulare.
- Capitolul 5 introduce o aplicație robotică concepută pentru dezvoltarea ușoară a scenariilor robotice de asistență. Aplicația integrează date de la mai mulți senzori pentru a genera capabilități specifice pentru un robot social. Capabilitățile pot fi combinate în comportamente complexe dedicate scenariilor de asistență socială.
- Capitolul 6 introduce o nouă metodă de predicție a traiectoriei oamenilor. Am proiectat o arhitectură care combină informații despre mișcarea oamenilor cu influența socială venită de la ceilalți participanți și informațiile de mediu extrase din scenă. Sistemul este validat într-un context robotic social, folosind un set de date existent și datele pe care le-am colectat.
- Capitolul 7 propune o arhitectură pentru reidentificarea și urmărirea oamenilor în timp real, bazată pe metoda de predicție a traiectoriei oamenilor. Sistemul reidentifică oamenii pe baza mișcării prezise, permițând reidentificarea după ocluzii sau după mișcarea camerei. Sistemul este validat luând în considerare aplicația robotică socială, folosind un set de date existent și datele pe care le-am colectat.
- Capitolul 8 validează sistemele pentru predicția traiectoriei oamenilor și urmărirea oamenilor în alte contexte autonome, și anume mașini și drone autonome. Evaluăm metodele utilizând seturile de date conexe existente și datele pe care le-am colectat, pentru a include mai multe variații în scenariile analizate.
- Capitolul 9 încheie lucrarea, prin conturarea principalelor contribuții ale tezei și a perspectivelor în ceea ce privește dezvoltarea viitoare.

Capitolul 2

Definirea Problemei

Obiectivul acestui capitol este de a defini problemele abordate de teză și de a le pune în contextul aplicațiilor autonome. Scopul tezei este de a dezvolta un sistem de urmărire a persoanelor bazat pe o abordare de predicție a traiectoriei și de a-l valida în contexte autonome. Pentru contextul roboticii sociale, teza introduce o aplicație robotică concepută pentru a crea comportamente robotice potrivite pentru scenarii de asistență socială. Având în vedere această structură, acest capitol definește problema roboticii sociale, predicția traiectoriei și urmărirea persoanelor, alături de provocările de cercetare asociate pe care acestea le implică.

Robotica de asistență socială [1] se referă la roboții care sunt menși să ajute oamenii într-o manieră care se concentrează pe interacțiunile sociale (de exemplu, vorbire, îndrumare, notificare, observare și amuzament). Deși interacțiunea fizică (de exemplu, transportul de obiecte) poate fi făcută doar de anumite tipuri de roboți, aceasta nu este necesară conform definiției menționate.

Unul dintre cele mai importante domenii pentru roboții de asistență socială (numiți și roboți *însoțitori*) este acela al sprijinirii populației în vârstă, în special a persoanelor care locuiesc singure sau în instituții de îngrijire, precum și a celor care sunt afectați de probleme medicale care necesită o monitorizare mai atentă a obiceiurilor zilnice. Domeniul Viață Activă și Asistată (Active and Assisted Living - AAL), care se preocupă de dezvoltarea tehnologiei pentru a sprijini nevoile populației în vârstă menționate mai sus, susține, prin urmare, în mod activ dezvoltarea capacităților roboților însoțitori.

Predicția traiectoriei oamenilor este metoda care estimează căile posibile urmate de fiecare persoană urmărită. Fiecare traiectorie estimată este calculată pe baza observațiilor anterioare, care sunt reprezentate de pozițiile persoanei în momentele anterioare de timp.

Formularea matematică a problemei de predicție a traiectoriei este reprezentată

de capacitatea de a estima X_{gt} pe baza X_{obs} . X_{obs} reprezintă traiectoria observată a unei persoane, de la momentul 1 până la o , unde 1 denotă primul element al secvenței și o este numărul de poziții observate pe baza cărora se face predicția. X_{gt} reprezintă traiectoria reală urmată de individ, de la momentul de timp $o+1$ până la $o+p$, unde p este numărul de poziții prezise. Sistemul propus urmărește să genereze X_{pred} , o estimare a lui X_{gt} , astfel încât erorile să fie minime.

Traietoriile viitoare prezise ale oamenilor din imagini oferă o informație critică pentru aplicațiile autonome. Predicția traiectoriei oamenilor este integrată în arhitecturile dispozitivelor autonome din motive de siguranță, mai ales când ne referim la mașinile cu conducere autonomă care merg la viteze mai mari. Informațiile furnizate de acest sistem pot fi un element decisiv în prevenirea situațiilor nedorite.

Reidentificarea și urmărirea persoanelor este o tehnică care permite sistemelor să identifice aceeași persoană într-o secvență de imagini. Tehnica atribuie un număr unic de identificare fiecărei persoane detectate într-o imagine. Urmărind o persoană pentru mai multe cadre, un sistem poate înțelege mai bine comportamentul unei persoane, poate asocia acțiuni sau poate planifica o interacțiune mai bună.

Din punct de vedere teoretic, problema reidentificării oamenilor presupune atribuirea unui număr unic de identificare $x_{i:n}$ fiecărei detecții $y_{i:n}$ reprezentând o persoană, detectată la momentul respectiv. n . Dacă în momentul de timp n , o detecție $y_{j:n}$ conține aceeași persoană ca și o detecție $y_{k:n-r}$ de la momentul de timp $n-r$, unde r reprezintă un număr oarecare de cadre, atunci numărul de identificare asociat lui $x_{j:n}$ trebuie să se potrivească cu cel care a fost asociat anterior, și anume $x_{k:n-r}$.

Informațiile furnizate de un sistem de reidentificare a persoanelor sunt esențiale pentru multe aplicații autonome. Roboții sociali trebuie să facă diferența între oameni pentru a-și ajusta comportamentul în funcție de comportamentul fiecărei persoane. Mașinile cu conducere autonomă trebuie să urmărească oamenii pentru a le putea prezice mișcarea și a conduce în siguranță. Sistemele de securitate, cum ar fi dronele de supraveghere, trebuie să urmărească oamenii pentru a analiza posibile situații periculoase. Sistemul este fundamental în multe scenarii, așa că sunt necesare rezultate precise.

Capitolul 3

Lucrări din Domeniu

Având în vedere îmbunătățirea actuală în domeniul inteligenței artificiale, zona de cercetare se concentrează pe dezvoltarea sistemelor inteligente alături de dispozitive autonome care pot ajuta oamenii în viața de zi cu zi. Unul dintre obiectivele noastre este să dezvoltăm un sistem general potrivit pentru orice platformă robotică capabilă să interacționeze cu oamenii. O parte importantă a acestei cercetări se concentrează asupra problemelor de vedere computațională care sunt necesare pentru un astfel de sistem. Analiza cercetării existente a fost realizată luând în considerare aceste domenii.

Roboții sociali sunt proiectați să interacționeze cu oamenii în mod natural. Mai multe proiecte au dezvoltat sisteme care cuprind un set divers de funcționalități, de la unele mai specifice [2, 3], la sisteme generale [4, 5, 6], incluzând inițiative precum [7] care sprijină dezvoltarea tehnologiilor pentru roboți autonomi.

Sisteme precum NAOqi oferă capabilități diverse, dar sunt limitate la anumiți roboți. Sisteme precum [3, 8] se concentrează mai mult pe interacțiunile sociale, dar nu oferă o soluție pentru o implementare determinată de obiective sau pentru compoziția comportamentului roboților. Proiectul EnrichMe [5] a vizat sprijinirea vieții persoanelor în vârstă prin tehnologii robotice și web. Deși este în general similar cu proiectul nostru ca principii de proiectare, module și proceduri de testare, acestuia îi lipsește un test mai cuprinzător al capabilităților de navigare, precum și mijloacele de detectare a unor acțiuni mai diverse ale utilizatorilor. Proiectul SocialRobot [6] s-a concentrat pe dezvoltarea unui robot personalizat pentru îngrijirea persoanelor vârstnice, dar autorii nu raportează o performanță mai sistematică a modulelor de funcționalitate individuale (nici în implementarea live, nici în cadrul de laborator).

Deoarece predicția traiectoriei este o metodă importantă care îmbunătățește experiența cu utilizatorul atunci când ne referim la dispozitive autonome,

subiectul este cercetat pe scară largă în domeniul inteligenței artificiale. Lucrarea *Social LSTM* [9] a fost punctul de plecare pentru multe direcții în cercetarea predicției traiectoriei persoanei, prin integrarea unui strat de influență socială între rețelele neurale recurente. Ideea a fost ulterior dezvoltată și îmbunătățită în multe lucrări, cum ar fi [10, 11, 12], prin combinarea unor rețele mai complexe, cum ar fi rețelele generative adversariale sau rețelele neurale adânci. Principalul dezavantaj al acestor sisteme este că predicțiile nu iau în considerare informațiile din mediu, care pot influența mișcarea unui individ.

Lucrări precum [13, 14, 15, 16] au îmbunătățit rezultatele sistemelor anterioare introducând informații din scenă în sistemele lor. Abordări mai recente, cum ar fi sistemele din [17, 18], folosesc metode de predicție bazate pe obiective, care aleg din mai multe obiective candidat pentru a alege cea mai bună potrivire. Principalul dezavantaj al acestor metode este că au nevoie de informații globale despre scenă obținute din imagini văzute de sus. Metoda propusă elimină aceste limitări și generează traiectorii folosind imagini preluate de la nivelul ochilor unui robot autonom, în medii dinamice cu schimbări ale unghiurilor și aspectului vizual.

Problema urmăririi persoanelor este o problemă complexă care a fost cercetată intens pentru o perioadă semnificativă de timp [19]. Una dintre primele abordări pentru reidentificarea persoanei datează din 1997 [20]. Abordările actuale propun sisteme complexe pentru a rezolva problema reidentificării oamenilor. În cercetări precum [21, 22], arhitecturile integrează date temporale pentru a reidentifica oamenii, în timp ce sisteme precum [23, 24] realizează o metodă complexă de potrivire a caracteristicilor pentru reidentificare.

Aplicațiile sistemelor de reidentificare a persoanelor necesită de obicei procesare în timp real. Unul dintre sistemele populare de urmărire în timp real a persoanelor este *DeepSort* [25, 26]. Sistemul combină informațiile vizuale extrase folosind o rețea neuronală convoluțională cu o tehnică simplă de estimare a poziției. *Tracktor* [27] efectuează urmărirea multiplă în timp real folosind o tehnică simplă bazată pe un regresor al unui detector de obiecte. Principalul dezavantaj al acestor tehnici, în special pentru sistemul *Tracktor*, este că necesită o frecvență mare a imaginilor.

În metoda propusă am combinat avantajele sistemului *DeepSort* cu avantajele unui sistem de predicție a traiectoriei, pentru a crea o tehnică mai puternică. Sistemul înlătură limitările legate de frecvențele mici ale imaginilor, deoarece modulul de predicție a traiectoriei introdus poate ajusta estimarea traiectoriilor în funcție de viteza persoanei țintă.

Capitolul 4

Seturi de Date Relevante

Pentru a proiecta o arhitectură adecvată pentru urmărirea persoanelor, a fost necesară o analiză a seturilor de date existente. Întrucât unul dintre scopurile cercetării este de a urmări oamenii pe baza informațiilor vizuale, seturile de date investigate conțin imagini ale oamenilor, atât în medii interioare, cât și în aer liber.

Seturile de date au fost selectate având în vedere scopul final al acestei teze, care este urmărirea persoanelor în aplicații autonome. Caracteristicile relevante examinate ale seturilor de date sunt unghiul de vizualizare al camerei și mobilitatea acesteia (fixă sau în mișcare). În ceea ce privește unghiul de vizualizare, fluxurile de imagini sunt captate dintr-un unghi de vizualizare de sus sau dintr-un unghi de vizualizare de la nivelul ochilor. În ceea ce privește mobilitatea, există seturi de date cu camere statice și seturi de date cu camere care sunt în mișcare.

În teză am integrat 6 seturi de date: *MOT17: Multi-Object Tracking* [28], *JRDB* [29], *ETH - BIWI Walking Pedestrians* [30], *UCY - Crowds by example* [31], *Caltech-Pedestrians* [32] și *Stanford Drone* [33].

MOT17: Multi-Object Tracking [28] și *JRDB* [29] sunt folosite pentru a evalua comportamentul sistemului în scenarii robotice. *MOT17: Multi-Object Tracking* este un set de date mare, care prezintă unghiuri de vizualizare și medii mixte. În unele scene camera se mișcă conform platformei robotice pe care se află. Numărul de persoane și obstacole din imagini variază în funcție de scenă. Setul de date *JRDB* este înregistrat de o platformă robotică aflată în mișcare cu mai multe camere și senzori. Setul de date conține imagini atât în interior, cât și în aer liber, cu un număr variabil de persoane. În funcție de scenă, imaginile pot include obstacole, cum ar fi scaune, mese, copaci, stâlpi. Seturile de date *ETH - BIWI Walking Pedestrians* [30] și *UCY - Crowds by example* [31] conțin imagini colectate în diferite scenarii în aer liber, cu imagini preluate dintr-un punct fix, plasat deasupra ambelor scene. Imaginile prezintă situații aglomerate cu puțini obstacole, precum copaci, stâlpi, bănci. Caracteristica acestor seturi de date este

că pozițiile oamenilor sunt reprezentate ca valori reale exprimate în metri și nu reprezintă coordonatele în imagini.

Setul de date *Caltech-Pedestrians* [32] este un set de date mare care conține imagini înregistrate de o mașină aflată în mișcare. Scopul acestui set de date este de a fi integrat în aplicații care utilizează vehicule autonome. Camera este orientată spre benzile drumului, pietonii urmăriți fiind în mare parte în lateralul imaginilor. Camera cu care au fost preluate imaginile se mișcă cu o viteză variabilă, iar unghiul acesteia se schimbă în funcție de mișcarea mașinii.

Stanford Drone [33] este un set de date mare care conține mai multe videoclipuri din diferite zone în aer liber dintr-un campus universitar. Imaginile sunt preluate de la o altitudine foarte mare. Caracteristica acestui set de date este că detecțiile asociate persoanelor urmărite sunt mici, având în vedere înălțimea. Setul de date este utilizat pentru a valida abordarea propusă în imagini înregistrate cu ajutorul unei drone.

Pentru a evalua abordarea pe mai multe scenarii particulare, am colectat date suplimentare pentru a include situații mai dificile. Am înregistrat mai multe videoclipuri pentru două contexte autonome diferite: platformă robotică și mașină cu conducere autonomă. Pentru a genera detecțiile persoanelor din imagini, am folosit un sistem existent de detectare a persoanelor, respectiv YOLO [34], pe care l-am aplicat pe fiecare imagine din setul nou de date.

Videoclipurile colectate din punctul de vedere al unui robot sunt preluate atât în medii interioare, cât și în aer liber. Am colectat 19 videoclipuri în total, cu o rezoluție de 640x480 pixeli și o frecvență de 30 de cadre pe secundă. Datele sunt înregistrate cu o cameră externă și includ foarte puțină mișcare în timpul înregistrării. Pentru acest caz, am înregistrat și un videoclip suplimentar folosind platforma noastră robotică într-un mediu interior. Videoclipul are o rezoluție de 320x240 pixeli și un framerate de 5 cadre pe secundă. Pentru contextul conducerii autonome am colectat 12 videoclipuri diferite. Camera este fixată în interiorul unei mașini, iar mișcarea acesteia este determinată de deplasarea vehiculului. Viteza variază de la 0 la 50 km/h, iar unghiul camerei se modifică în funcție de direcția mașinii. Videoclipurile au o rezoluție de 640x480 pixeli și o frecvență de 30 de cadre pe secundă.

Pe lângă cele două contexte, am înregistrat 5 videoclipuri dintr-un punct mai înalt pentru a simula perspectiva unei drone. Camera este fixă, iar datele sunt achiziționate în două scene, una în interior și una în exterior.

Aplicația Robotică AMIRO pentru Roboți Sociali

Platformele robotice autonome sunt integrate într-o varietate de aplicații care necesită intervenție umană minimă spre deloc. Având în vedere spectrul vast de funcții pe care un astfel de sistem le poate îndeplini, este de la sine înțeles că există o multitudine de platforme diferite construite cu niște sisteme specifice. Roboții sociali în special, deși au caracteristici similare în ceea ce privește obiectivele finale, pot avea implementări foarte diferite. Totalitatea diferențelor dintre sistemele existente produce două dezavantaje majore: complică procesul de proiectare pentru ingineri, deoarece aceștia trebuie să învețe și să se adapteze la fiecare sistem nou întâlnit și elimină posibilitatea de a migra cu ușurință comportamentul sau funcția unui robot de la o platformă la alta.

Unul dintre scopurile acestei cercetări este acela de a dezvolta o aplicație de robotică socială, potrivită pentru robotul Pepper [35], dar aplicabilă în mod necesar oricărui robot compatibil cu sistemul de operare robot (ROS) [36]. Sistemul AMIRO (Ambient Robotics) [37] este o aplicație robotică care își propune să ofere o platformă care poate fi integrată cu ușurință cu o varietate de roboți. Designul modular al platformei permite posibilitatea de a integra noi module sau de a le actualiza pe cele existente cu un efort minim. Platforma dezvoltată urmează tendințele recente ale roboticii, folosind procesare distribuită și servicii externe. Aceasta pune la dispoziție un set cuprinzător funcționalități care facilitează crearea de comportamente complexe.

5.1. Prezentare Generală și Arhitectură

Sistemul abordează cerințele majore pentru un robot de asistență socială, menținând în același timp un grad ridicat de modularitate în construirea și integrarea de noi module. Componentele principale sunt detașate și fiecare

componentă poate fi rulată ca o aplicație separată. Această caracteristică asigură robustețea sistemului, acesta putând funcționa chiar și în cazul în care unul dintre module se defectează. Componentele principale ale sistemului sunt enumerate mai jos.

- Înțelegerea vizuală - integrează mai multe tehnici de vedere computațională, cum ar fi recunoașterea oamenilor și a obiectelor, reidentificarea oamenilor, predicția traiectoriei și componenta de recunoaștere a activității. Atunci când un obiect este detectat, modulul calculează poziția 3D a obiectului în raport cu robotul, care este apoi trimis către componenta *Navigare* pentru a-l plasa pe hartă.
- Navigarea și evitarea obstacolelor - este responsabil cu mișcările robotului în interiorul mediului. Acest modul realizează achiziția de date și este responsabil cu procesarea algoritmilor SLAM care generează căile de urmat.
- Interpretarea vorbirii și dialogul - este responsabil cu interpretarea vorbirii utilizatorului și pune la dispoziție capabilități Text-to-Speech (transformarea textului în vorbire). Modulul poate trimite sarcini noi robotului pe baza dialogului curent cu utilizatorul.
- Integrarea cu sisteme inteligente - adună date despre sănătatea utilizatorului (tensiune arterială, ritm cardiac, pași, somn) și mediu (temperatura camerei, umiditate, luminozitate) și oferă instrumentele necesare pentru a acționa diferite dispozitive inteligente (de exemplu, iluminat inteligent, jaluzele).
- Generarea de comportamente - este în centrul sistemului și trimite comenzi către toate celelalte module. Acesta generează comportamente bazate pe comenzile primite. Se caracterizează prin ușurința și robustețea cu care agregă celelalte module.

Arhitectura sistemului este construită pe baza sistemului de operare ROS. Fiecare componentă principală a arhitecturii oferă un set de canale ROS, care permite ca datele să fie schimbate între module în mod continuu și asincron. Din perspectiva implementării modulelor, arhitectura AMIRO își distribuie serviciile pe mașinile care rulează în cloud (servicii externe) sau care constituie cloud-edge (servicii distribuite). Avantajul unei astfel de arhitecturi este faptul că fiecare nod poate fi rulat pe mașini separate, prevenind propagarea erorilor și facilitând implementarea sistemului. Ca atare, robotul devine doar un mijloc de achiziție de date și de executare a unor acțiuni. Modul în care este proiectată arhitectura este

direct influențat de cerințele de procesare în timp real ale proiectului. Folosind modulele cloud-edge, eliminăm cerințele de calcul impuse robotului integrat și oferim capacități de calcul mult mai puternice, pentru tehnici mai complexe. Am ales să folosim procesarea locală a datelor în locul procesării în cloud, luând în considerare cantitatea necesară de informații transmise prin internet. Pentru răspunsuri rapide în cazul informațiilor vizuale, lățimea de bandă a conexiunii la internet ar trebui să fie suficient de mare pentru a nu adăuga latență.

5.2. Experimente

Pentru a evalua funcționarea generală a aplicației am testat mai multe scenarii experimentale, care sunt plauzibile în contextul unei aplicații de robotică socială. Am implementat sistemul folosind robotul social Pepper [35] pentru a valida aplicația AMIRO în condiții reale de funcționare. Am propus două scenarii de evaluare care ilustrează o parte din funcționalitățile de bază ale aplicației: recunoașterea și înțelegerea vocală, găsirea vizuală a persoanei, estimarea poziției unei ținte detectate, recunoașterea acțiunii umane, explorarea și navigarea mediului, interacțiunea cu sistemul inteligent, planificarea acțiunilor și execuție.

Scenariile alese dovedesc aplicabilitatea sistemului în două situații separate: scenariu cu o singură persoană și scenariu cu mai multe persoane. În primul scenariu, robotul interacționează cu o singură persoană pe care o asistă în activitățile de bază. În al doilea scenariu, robotul este utilizat în medii cu mai multe persoane, necesitând tehnici de recunoaștere și planificare mai avansate. Scenariile au fost evaluate folosind platforma robotică Pepper în condiții de laborator. Un avantaj important al sistemului propus este că este independent de platformă, ceea ce permite integrarea cu o varietate de platforme robotice. Experimentele pe care le-am efectuat pentru a evalua performanța fiecărui modul se bazează pe intrarea care vine direct de la robotul Pepper. Acest lucru ne-a permis să evaluăm în mod obiectiv limitările de funcționalitate care se datorează platformei robotizate Pepper, sau care pot fi atenuate prin îmbunătățiri viitoare ale modulelor noastre.

Capitolul 6

Sistemul de Predicție a Traiectoriei

Predicția traiectoriei oamenilor este capacitatea de a analiza comportamentul unei persoane și de a deduce mișcările viitoare pe baza observațiilor. Deși pare simplu, modelarea procesului complex din spatele comportamentului uman este o sarcină dificilă. Predicția traiectoriei poate fi integrată într-o varietate de aplicații: robotică, conducere autonomă, urmărirea persoanelor, supraveghere etc. Odată cu dezvoltarea dispozitivelor autonome, rezultatele precise sunt imperative în realizarea unui obiectiv specific sau pentru prevenirea evenimentelor nedorite. Sistemele de predicție a traiectoriei trebuie să ia în considerare mai mulți factori pentru a obține un rezultat precis. Întrucât traiectoria unei persoane este influențată de mișcarea celorlalți oameni din vecinătate, de obstacolele din mediu și de punctele de interes din scenă, tehnica propusă încorporează influențele sociale dintre oameni și informațiile contextuale din mediu, împreună cu pozițiile persoanei, pentru generarea rezultatelor finale.

6.1. Prezentare Generală și Arhitectură

Arhitectura sistemului constă în principal în două module: un modul pentru înțelegerea scenei și un modul pentru generarea traiectoriei prezise. Structura arhitecturii este prezentată în Figura 6.1. Modulul *Scene Understanding* extrage datele vizuale asociate cu imaginea de intrare primită, generând informații despre scenă, cum ar fi obstacole sau căi posibile de urmat, alături de poziția corespunzătoare a persoanei. Modulul *Trajectory Generation* integrează datele vizuale generate pentru a estima cea mai plauzibilă traiectorie, pe baza mediului, a celorlalți participanți și a mișcării persoanei.

Pentru a genera predicțiile, arhitectura propusă integrează trei factori care pot influența traiectoria unui individ: comportamentul persoanei, influența socială și forma scenei. Comportamentul unei persoane se referă la pozițiile

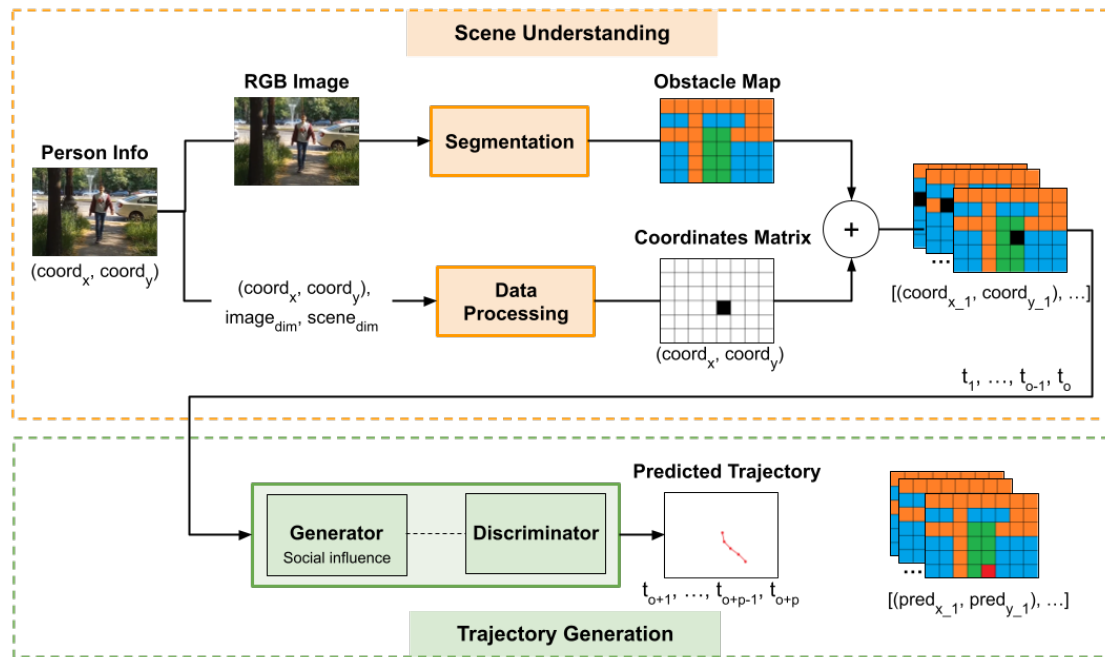


Figura 6.1: Arhitectura Social-GAN cu hărți ale obstacolelor. Datele agregate sunt calculate pentru fiecare persoană din imagini.

anterioare în care persoana a fost observată. În Figura 6.1, comportamentul este modelat de coordonatele 2D alături de matricele de coordonate corespunzătoare. Influența socială se referă la mișcarea celorlalți participanți din mediu, care poate influența traiectoria unei persoane, prin evitarea coliziunilor. În arhitectură, influența socială este codificată de un strat de conexiune în modulul *Trajectory Generation*, care transferă informații între oamenii din vecinătate. Forma scenei este se referă la modul în care arată mediul, în ceea ce privește obstacolele și căile posibile. Aceste informații sunt codificate prin harta obstacolelor, care este generată pe baza măștii de segmentare a imaginii de intrare.

Informațiile asociate cu o persoană care sunt transmise ca intrare în sistem sunt definite de coordonatele reprezentând poziția persoanei la fiecare moment de timp, alături de imaginile RGB ale scenei în exact acel moment. Pe baza acestor informații, sistemul calculează două informații suplimentare: matricea coordonatelor persoanei și harta obstacolelor din mediu. Sistemul generează un volum de date pentru fiecare persoană detectată în imagine și apoi transmite volumele combinate către rețeaua generativă adversarială.

6.2. Evaluare și Validare

Sistemul de predicție a traiectoriei a fost evaluat prin analiza rezultatelor obținute pe seturi de date relevante existente și prin interpretarea performanței componente atunci când este integrată în cadrul aplicației AMIRO. Pentru seturile de date integrate, erorile raportate pentru un subset au fost obținute prin antrenarea rețelei pe baza celorlalte subseturi, cu secvențe de predicție de lungime 8.

Evaluarea obiectivă a fost realizată pe baza setului de date JRDB, deoarece acesta este relevant pentru contextul roboticii sociale. Am raportat valorile în pixeli obținute pentru două metrici de eroare standard, ADE, eroare medie de deplasare și FDE, eroare finală de deplasare. Erorile medii pe care le-am obținut sunt de 23.69 pixeli pentru metrica ADE și 36.34 pentru metrica FDE, care sunt suficient de mici pentru un comportament satisfăcător al platformei robotizate. În plus, am evaluat sistemul pe seturile de date ETH și UCY, pentru a ne poziționa metoda față de alte sisteme existente, deși metoda a fost concepută pentru a lucra pe coordonate în imagini, detectate dintr-o perspectivă de la nivelul ochilor.

Pentru a analiza impactul sistemului de predicție a traiectoriei în contextul aplicațiilor de robotică socială, am definit și câteva scenarii care au fost testate folosind aplicația AMIRO. Am analizat comportamentul robotului în mai multe situații posibile, după cum urmează:

1. Robotul ajută o persoană furnizând indicații către o anumită locație.
2. Robotul lucrează ca ghid turistic într-un muzeu.
3. Robotul ajută oamenii cu îndrumări în timpul unei evacuări de urgență.

Fiecare scenariu descris combină capacitățile existente din cadrul aplicației AMIRO cu funcțiile oferite de modulul de predicție a traiectoriei. Pe lângă funcția elementară de estimare a traiectoriilor viitoare, aplicația poate detecta mai devreme dacă o persoană se apropie de o anumită țintă, ceea ce poate evita situațiile periculoase sau poate accelera procesul de interacțiune.

Chiar dacă prezintă unele limitări în cazuri particulare, informațiile furnizate de modulul de predicție a traiectoriei pot ajuta înțelegerea comportamentului oamenilor și pot îmbunătăți comportamentul robotului. În ceea ce privește viteza de procesare, modulul este potrivit pentru aplicații cu rezultate în timp real, generând rezultate în aproximativ 110 milisecunde, indiferent de numărul de traiectorii prezise.

Capitolul 7

Sistemul de Urmărire și Reidentificare în Timp Real a Persoanelor

Un sistem de reidentificare a persoanelor atribuie aceiași identificatori exact acelorași oameni într-o succesiune de imagini. Caracteristicile unei astfel de componente nu se limitează doar la aplicațiile de robotică socială, ci pot fi extinse la orice dispozitiv autonom. Fie că vorbim de vehicule autonome, roboți, supraveghere sau medii inteligente, capacitatea de a detecta și recunoaște aceeași persoană într-o secvență de imagini, precum și abilitatea de a distinge între mai mulți indivizi, este un element cheie.

Problema reidentificării persoanelor are o natură complexă, deoarece performanța unui astfel de sistem poate fi puternic influențată de mai mulți factori. A proiecta un sistem care funcționează bine într-o varietate de scenarii implică crearea unui mecanism care este capabil să depășească probleme precum ocluziile sau variația aspectului. Aceste probleme apar din circumstanțe naturale, cum ar fi oamenii care trec prin spatele obstacolelor în timp ce se mișcă, stiluri similare în ceea ce privește îmbrăcămintea și coafura, sau interacțiunile dintre indivizi. Pentru a gestiona aceste tipuri de probleme, metoda noastră integrează informații generate de un sistem de predicție a traiectoriei.

7.1. Prezentare Generală și Arhitectură

Metoda propusă introduce un sistem modular care combină o tehnică standard de urmărire a persoanelor cu o metodă de predicție a traiectoriei persoanelor. Arhitectura ia ca intrare un flux de imagini și prezice un număr de identificare pentru fiecare persoană, mai precis pentru fiecare detecție generată pentru imaginea de intrare. Figura 7.1 prezintă arhitectura propusă pentru sistemul de reidentificare în timp real. Arhitectura propusă poate fi integrată într-un sistem conceput pentru roboți sau pentru orice dispozitiv autonom,

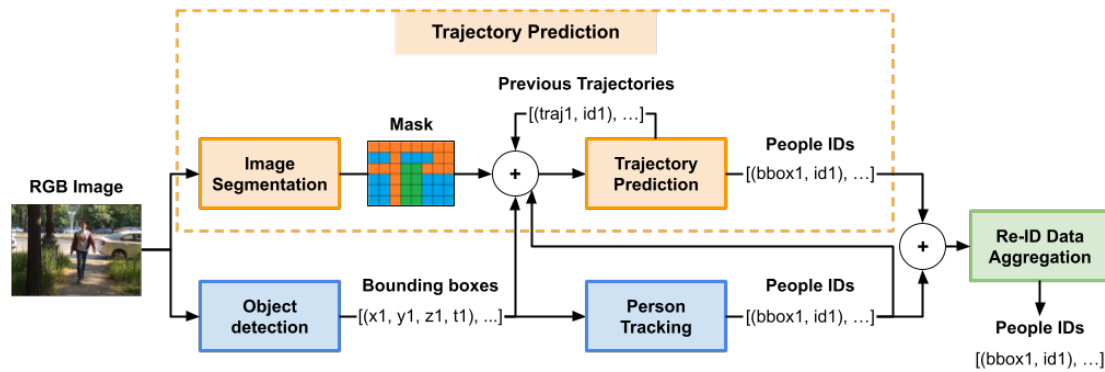


Figura 7.1: Arhitectura de reidentificare și urmărire în timp real a persoanelor, bazată pe o componentă de predicție a traiectoriei.

deoarece necesită doar date RGB pentru a genera rezultate.

Arhitectura pentru reidentificarea și urmărirea persoanelor combină cinci module pentru a genera rezultatul final, după cum urmează:

- **Detectia obiectelor** - Modulul de detecție a obiectelor identifică persoanele din imagini. Pentru fiecare imagine de intrare modulul generează dreptunghiuri reprezentând pozițiile oamenilor.
- **Segmentare semantică** - Modulul de segmentare semantică face parte din modulul de predicție a traiectoriei oamenilor și e responsabil cu extragerea informațiilor de mediu din imagini. Rețeaua este rulată în paralel cu detecția obiectelor pentru a reduce și mai mult timpul de procesare.
- **Predicția traiectoriei** - Modulul de predicție a traiectoriei este utilizat pentru a estima mai bine mișcarea oamenilor. Ajută sistemul să atribuie mai precis identificatorii potriviți luând în considerare traiectoriile viitoare estimate.
- **Urmărirea persoanei** - Modulul de urmărire a persoanei este utilizat de sistemul de predicție a traiectoriei pentru a aduna datele inițiale observate înainte de a genera o traiectorie. Sistemul trebuie să poată distinge între detecțiile generate astfel încât să construiască succesiunea de poziții observate care aparțin unei singure persoane. Utilizează sistemul introdus în [25] și [26].
- **Re-ID Data Aggregation** - Modulul de agregare a datelor conectează informațiile din modulul de urmărire a persoanelor și din modulul de predicție a traiectoriei pentru a genera identificatorii finali.

7.2. Evaluare si Validare

Sistemul de reidentificare și urmărire a persoanelor a fost testat luând în considerare scenariile de robotică socială. Am efectuat o evaluare pe baza unui set de date existent și o analiză folosind imaginile colectate. Evaluarea compară rezultatele unei tehnici standard de urmărire în timp real a persoanelor [25] cu sistemul nostru de reidentificare bazat pe predicția traiectoriei. Evaluarea demonstrează performanța mai bună a tehnicii propuse. Tabelul 7.1 raportează valorile obținute pe setul de date MOT17 pentru metricile de evaluare a sistemelor de urmărire a persoanelor. Sistemul propus obține o valoare mai mare pentru metrica MOTA și valori similare pentru IDF și HOTA. Aceste valori au fost obținute prin utilizarea unui model de predicție a traiectoriei antrenat pe setul de date JRDB.

Metrică	Sistemul standard	Sistemul propus
MOTA [38] (%)	55.05	61.04
IDF1 [39] (%)	55.62	52.24
HOTA [40] (%)	45.57	43.79

Tabela 7.1: Valorile metricilor de evaluare a sistemelor de urmărire a persoanelor, calculate pentru sistemul propus pe setul de date MOT17.

Am efectuat o analiză a metodei pe anumite scenarii particulare folosind datele colectate, cu una sau mai multe persoane în imagini. Scenariile variază în ceea ce privește condițiile de mediu, precum fondul, lumina, distanța sau obstacolele care pot provoca ocluzii totale sau parțiale, dar și în ceea ce privește mișcarea oamenilor, precum interacțiunile, ocluziile, schimbarea bruscă a direcției. Metoda noastră este capabilă să facă față situațiilor mai complexe și să prezică identificatorii corecți chiar și după ocluzii îndelungate, în timp ce metoda DeepSort prezintă probleme.

Sistemul poate funcționa în medii dinamice datorită informațiilor vizuale despre mediu, deoarece masca de segmentare este calculată pentru fiecare imagine de intrare. Informațiile procesate continuu ajută și în cazul camerelor care se mișcă, întrucât sistemul ajustează traiectoriile mai mult pe baza datelor recente. Problema de scalare este gestionată de reprezentarea sub formă de coordonate 2D ale oamenilor. Această abordare, totuși, poate prezenta unele probleme, luând în considerare situațiile în care detecțiile nu sunt consistente și variază semnificativ în ceea ce privește suprafața detectată.

Capitolul 8

Validarea Sistemelor în Alte Contexte

Contextul roboticii autonome este o direcție foarte populară în domeniul cercetării, având în vedere varietatea de aplicații în care poate fi integrat și interesul crescut al oamenilor. Cu toate acestea, accentul este pus nu numai pe roboții sociali, ci și pe alte dispozitivele autonome existente, cum ar fi mașinile sau dronele autonome.

Principiile pe care le-am aplicat în sistemele de predicție a traiectoriei oamenilor și reidentificarea și urmărirea oamenilor pot fi aplicate și în scenarii diferite. Pe lângă robotica socială, am testat metodele în alte două contexte, și anume mașini cu conducere autonomă și drone zburătoare. Ambele contexte presupun mișcarea camerei, variația aspectului și ocluziile, principala diferență fiind unghiul camerei. Aplicațiile pentru mașini autonome analizează imaginile dintr-o perspectivă de la nivelul ochilor, ca în cazul roboticii sociale, în timp ce dronele procesează imagini dintr-un unghi de vizualizare înalt.

8.1. Validarea în Contextul Mașinilor Autonome

Contextul conducerii autonome presupune detectarea pietonilor de pe marginea drumului și estimarea pozițiilor lor viitoare pentru a preveni eventualele coliziuni. Cele mai importante diferențe dintre contextul roboticii sociale și cel al conducerii autonome sunt că viteza unei mașini autonome este mult mai mare decât în cazul unui robot, iar dimensiunea detecțiilor reprezentând pietonii sunt mai mici decât cele din aplicațiile de robotică socială.

Pentru o evaluare obiectivă a componentei de predicție a traiectoriei oamenilor pentru aplicațiile cu mașini cu conducere autonomă, am integrat setul de date Caltech-Pedestrians [32], deoarece include videoclipuri înregistrate de o mașină în mișcare într-un mediu urban. Am calculat valorile ADE (eroare medie de deplasare) și FDE (eroare finală de deplasare) pentru fiecare subset, obținând

valori medii de 7.00 pixeli pentru ADE și respectiv 12.85 pixeli pentru FDE. Valorile obținute sunt mici, demonstrând că funcționarea sistemului este robustă și fiabilă în aplicațiile de conducere autonomă.

Reidentificarea în contextul conducerii autonome este o problemă mai provocatoare decât în contextul roboticii sociale. Pentru o performanță bună, sistemul necesită un model care poate codifica mișcarea camerei, variația dimensională și ocluziile multiple. Mișcarea camerei este o constrângere importantă în problema reidentificării, deoarece implică nu numai modificări ale unghiului de vedere, ci și fluctuații ale vitezei de mișcare. Am analizat mai multe scenarii particulare folosind datele colectate, și am comparat valorile raportate pentru metricile standard de evaluare a sistemelor de urmărire a persoanelor pe setul de date Caltech-Pedestrians, pentru a demonstra îmbunătățirea pe care o aduce sistemul propus în comparație cu cel standard. Tabelul 8.1 demonstrează o îmbunătățire semnificativă a sistemului propus în comparație cu DeepSort pentru valorile tuturor metricilor standard.

Metrică	Sistem standard	Sistemul propus
MOTA [38] (%)	89.22	94.92
IDF1 [39] (%)	78.36	90.10
HOTA [40] (%)	78.96	89.15

Tabela 8.1: Valorile metricilor de evaluare pentru sistemele de urmărire a persoanelor pentru sistemul propus pe baza setului de date Caltech-Pedestrians [32].

8.2. Validarea în Contextul Aplicațiilor cu Vedere de Sus

Scenariile prezentate anterior presupun că imaginile sunt înregistrate dintr-o perspectivă de la nivelul ochilor. Atât un robot social, cât și un vehicul autonom au camerele plasate pe platformă astfel încât imaginile înregistrate să fie văzute în perspectivă. Aplicații precum dronele sau sistemele de supraveghere folosesc imagini obținute dintr-un punct de vedere înalt.

Pentru a obține o evaluare obiectivă a performanței sistemelor în condițiile în care imaginile au o vedere de sus a scenei, am integrat setul de date Stanford

Drone [33], care conține o colecție de imagini înregistrate de la mare altitudine într-un campus universitar. Am calculat ADE (eroare medie de deplasare) și FDE (eroare finală de deplasare), ca și în fazele de evaluare ale celorlalte două cazuri, pentru fiecare subset. Erorile raportate sunt relativ mici, cu valori medii de 13.26 pixeli pentru ADE și 22.02 pixeli pentru FDE. Erorile medii validează faptul că sistemul poate fi integrat într-o aplicație care utilizează imagini preluate dintr-un unghi de vizualizare înalt.

Sistemul de reidentificare și urmărire a persoanelor a fost testat folosind date colectate dintr-un punct de vedere înalt. Experimentele noastre au demonstrat că sistemul funcționează satisfăcător pentru o aplicație de urmărire a persoanelor, deși prezintă mai multe schimbări de identificatori asociați persoanelor în comparație cu celelalte două contexte. Rezultatele obținute în acest scenariu sunt puternic influențate de rețeaua de segmentare semantică integrată în sistem, deoarece aceasta trebuie antrenată să segmenteze imagini cu vedere de ansamblu, pentru a se potrivi cerințelor contextului dronelor autonome.

Aceste scenarii de validare, alături de cel de robotică socială, demonstrează că sistemul poate fi utilizat într-o varietate de aplicații. Mai mult, proprietatea de a fi potrivit chiar și pentru scenarii cu un unghi de vizualizare de sus este importantă pentru aplicațiile în care există mai multe fluxuri de imagini care provin din unghiuri diferite, care pot fi combinate pentru a obține un sistem mai stabil.

Concluzii

Teza prezintă o abordare pentru un sistem de reidentificare și urmărire a persoanelor în timp real pentru aplicații autonome bazat pe o componentă de predicție a traiectoriei oamenilor. Validează abordarea în trei contexte autonome diferite, și anume roboți sociali, mașini cu conducere autonomă și drone autonome. În plus, introduce o aplicație pentru scenarii robotice care integrează sistemul de urmărire a persoanelor și componenta de predicție a traiectoriei oamenilor cu alte capacități robotice pentru a crea un sistem robotic social complet autonom pentru diferite scenarii de interacțiune cu utilizatorul.

Sistemul de urmărire și reidentificare are un rol esențial în contextele care presupun funcționarea în medii care includ mișcarea oamenilor sau chiar interacțiuni cu aceștia. Dispozitivele autonome necesită informații despre comportamentul persoanelor pentru a preveni situațiile potențial periculoase. Componenta de predicție a traiectoriei estimează mișcarea viitoare pe baza datelor observate pentru a îmbunătăți stabilitatea și siguranța dispozitivelor autonome. Faza de evaluare a celor două sisteme dovedește aplicabilitatea acestora pe diverse situații și scenarii. Sistemele au fost testate pe date proprii și pe seturi de date relevante pentru a avea atât o evaluare cantitativă, cât și una calitativă. Evaluarea efectuată a demonstrat aplicabilitatea sistemelor pentru aplicații în timp real în condiții de funcționare produse de dispozitivele autonome.

Aplicația de robotică socială combină informațiile menționate mai sus cu capacitățile generate pe baza senzorilor robotici, pentru a dezvolta o platformă potrivită pentru aplicații sociale generale care utilizează roboți. Aceasta încorporează funcționalități pentru roboți în ceea ce privește înțelegerea vizuală, navigarea, interacțiunea vocală și planificarea. Cadrul oferă posibilitatea de a combina diferite funcții din fiecare capacitate pentru a crea comportamente potrivite pentru un scenariu robotic. Sistemul a fost testat în condiții de laborator folosind un robot social existent. Faza de evaluare include mai multe

scenarii, pe care robotul trebuie să le poată executa corect, indiferent de utilizatorul cu care interacționează. Scenariile au testat diferite situații, implicând atât un singur utilizator, cât și mai mulți utilizatori, atât cu un fir de execuție simplu, cât și pe baza unor comportamente complexe. Am realizat mai multe videoclipuri¹ pentru a oferi un exemplu de funcționare a sistemelor prezentate în teză.

9.1. Contribuții

Teza introduce câteva tehnici noi legate de extragerea informațiilor vizuale pentru aplicații autonome. Contribuțiile originale ale lucrării prezentate în această teză sunt următoarele:

- Un sistem pentru roboții sociali generali care poate fi aplicat pentru aplicații de asistență socială. Combinăm datele obținute de la mai mulți senzori pentru a crea o platformă omogenă care poate oferi acțiuni grupate în comportamente complexe. Sistemul oferă capacități de navigare, bazate pe laserele robotului, pentru înțelegere vizuală, bazată pe camere RGB și de adâncime, pentru interacțiune vocală, bazată pe microfoane și difuzoare. Senzorii specificați sunt la baza oricărui dispozitiv robotizat social. Unele părți ale sistemului au fost dezvoltate în colaborare cu Mihai Nan, Alex Awada și Alexandru Sorici.
- O arhitectură pentru planificarea succesiunii sarcinilor. Arhitectura permite schimbări dinamice atunci când noi comenzi sunt declanșate de diverse surse: comenzi vocale ale utilizatorului, interfață vizuală și platformă externă. Sarcinile sunt schimbate în funcție de prioritățile asociate. Fiecare comandă primită de sistem declanșează un comportament specific definit, care constă într-o succesiune de sarcini. Dispozitivul robotizat execută o sarcină la un moment dat, permițând posibilitatea de a întrerupe și de a relua comportamentele în funcție de importanța acestora.
- O arhitectură pentru un sistem de predicție a traiectoriei oamenilor în timp real. Sistemul încorporează trei informații pentru a genera predicțiile finale: mișcarea individuală, informațiile de mediu și influența socială. Mișcarea individuală este colectată pe baza datelor observate, reprezentând pozițiile

¹https://ctipub-my.sharepoint.com/:f:/g/personal/stefania_a_ghita_upb_ro/EoCy0oZ--RZ0re5ktqPXioAB2DjgsNExNk2kn0mAvQWzgw?e=fM4cEa.

în care persoana se afla înainte de predicție, informațiile de mediu sunt generate folosind o rețea de segmentare semantică care procesează imaginile RGB corespunzătoare, iar influența socială este integrată printr-un strat de conexiune care face schimb de informații între persoanele vecine.

- O analiză a avantajelor integrării unui sistem de predicție a traiectoriei oamenilor într-un sistem robotizat. Analiza include experimente efectuate folosind un robot social. Metoda este integrată în cadrul sistemului robotic ca o nouă componentă care expune capacități suplimentare pentru robot. Scenariile experimentale definite demonstrează îmbunătățirea performanței unui robot de asistență socială în diverse situații critice.
- O metodă pentru schimbarea unghiului de vizualizare al unei imagini de la o vedere în perspectivă la o vedere de ansamblu. Metoda aplică o transformare matematică a unei imagini folosind o serie de parametri generați pe baza unei metode de segmentare semantică.
- O arhitectură pentru un sistem de urmărire și reidentificare a persoanelor în timp real. Arhitectura combină o tehnică simplă de urmărire cu un sistem de predicție a traiectoriei oamenilor pentru a reidentifica aceleași persoane într-o secvență de imagini. Metoda este concepută pentru fluxuri de imagini înregistrate de un dispozitiv autonom, deoarece poate codifica mișcarea camerei, modificări ale unghiului de vizualizare și variații de viteză. Atribuie numere de identificare fiecărei persoane detectate într-o imagine într-un timp scurt, făcându-l potrivit pentru aplicații autonome. Faza de evaluare a dovedit validitatea metodei în trei contexte diferite: robotică socială de asistență, conducere autonomă și drone autonome.
- O colecție de date proprii, utilizată pentru a valida sistemul de urmărire și reidentificare a persoanelor. Colectarea datelor include videoclipuri înregistrate în mai multe medii văzute în perspectivă. Datele conțin fluxuri de imagini RGB înregistrate în două scenarii diferite: din punctul de vedere al unei platforme robotice și dintr-o mașină în mișcare. Videoclipurile includ ocluzii, modificări ale unghiului camerei și variații ale mișcării vitezei. Detecțiile corespunzătoare pentru fiecare imagine au fost calculate folosind un sistem extern de detectare a persoanelor.

9.2. Perspectivă

Scopul acestei teze este de a proiecta un sistem fiabil pentru urmărirea persoanelor în contexte autonome și de a-l integra într-un sistem dezvoltat pentru roboții de asistență socială. Sistemele nou introduse pot fi îmbunătățite luând în considerare obiectivele pe termen scurt și dezvoltarea pe termen lung.

Sistemul robotic implică mai multe aspecte atunci când se iau în considerare posibilele îmbunătățiri în ceea ce privește modulele funcționale individuale. Setul de module de funcționalitate poate fi extins cu abilitatea de a percepe emoții din imaginile RGB. O astfel de caracteristică poate îmbunătăți comportamentul general al unui robot social, făcându-l mai conștient de ceea ce are nevoie utilizatorul. Modulul de comandă vocală poate fi extins pentru a include limbi suplimentare și un set mai bogat de comenzi. Această îmbunătățire ar permite o testare mai amplă care implică utilizatori din alte țări. În plus, modulul de estimare a coordonatelor 3D poate fi îmbunătățit în continuare prin înlocuirea informațiilor provenite de la un senzor de adâncime cu o sursă de date mai fiabilă. Deoarece valorile tuturor componentelor coordonatei 3D sunt calculate pe baza adâncimii înregistrate, este de așteptat ca eliminarea erorilor din datele de adâncime să conducă la estimări mai precise. În plus, modulul de planificare a comportamentului robotului poate fi extins pentru a îmbina secvențe de acțiuni predefinite cu rezultatele de planificare, bazate pe metoda ROSPlan, permițând astfel o funcționalitate flexibilă și mai extensibilă de compoziție a comportamentului.

Modulul de predicție a traiectoriei poate fi îmbunătățit prin integrarea de informații suplimentare în sistem. O posibilă dezvoltare ar presupune includerea influenței venite de la toate obiectele care se mișcă în imagini care ar putea influența mișcarea unei persoane. Mișcarea mașinilor sau a animalelor poate influența semnificativ traiectoria unei persoane, provocând schimbări sau opriri bruște. În plus, influența socială poate fi mai bine distribuită între participanți. În forma actuală a arhitecturii, oamenii din imagini schimbă informații despre traiectoriile lor, indiferent de destinația sau orientarea acestora. O posibilă îmbunătățire constă în analiza destinațiilor oamenilor și distribuirea informațiilor doar între persoanele care se află sau vor fi în apropiere. Mai mult, putem suplimenta informațiile de mediu extrase de sistem prin includerea unor hărți de probabilitate care reprezintă posibilele destinații ale oamenilor în scenă. Având în vedere acest sistem, unul dintre obiectivele noastre este de a genera și mai multe informații pentru un sistem autonom, cum ar fi detectarea coliziunilor și analiza

trajectoriei grupurilor, pe baza rezultatelor modulului de predicție a trajectoriei.

Modulul de urmărire și reidentificare a persoanelor este o componentă esențială care trebuie îmbunătățită în continuare pentru reidentificări și mai fiabile. O îmbunătățire principală pe care o poate experimenta sistemul este înlocuirea tehnicii simple de urmărire care generează primele trei poziții observate ale unei persoane. Prin înlocuirea acestei metode cu una mai precisă, performanța generală a sistemului ar fi mult îmbunătățită. O altă posibilă dezvoltare constă în integrarea caracteristicilor vizuale extrase de o rețea neurală pentru a diferenția mai bine cazurile în care pozițiile viitoare precise ale diferitelor persoane sunt în proximitate. În arhitectura actuală, în cazul în care trajectoriile a două persoane se suprapun și doar o singură persoană este vizibilă în imagine, sistemul atribuie persoanei vizibile numărul de identificare al persoanei care anterior era considerată a fi mai aproape de cameră, mai precis cu o dimensiune mai mare a detecției. Prin integrarea unor indicatori vizuali care să reprezinte aspectul general al oamenilor, astfel de probleme ar fi rezolvate cu o acuratețe mai mare. În plus, putem îmbunătăți sistemul combinând informațiile extrase prin metoda de predicție a trajectoriei oamenilor cu detecția posturilor acestora, pentru a valida trajectoriile în funcție de activitățile realizate. Această îmbunătățire presupune două direcții: una în care validăm trajectoriile pe baza acțiunilor recunoscute și una în care integrăm posturile direct în sistemul de predicție a trajectoriei pentru rezultate mai informate.

Referințe

- [1] D. Feil-Seifer and M. J. Mataric, “Defining socially assistive robotics,” in *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*, pp. 465–468, IEEE, 2005.
- [2] H. Nap, S. Suijkerbuijk, D. Lukkien, S. Casaccia, R. Bevilacqua, G. Revel, L. Rossi, and L. Scalise, “A social robot to support integrated person centered care,” *International Journal of Integrated Care*, vol. 18, no. s2, 2018.
- [3] C. Antonopoulos, G. Keramidas, N. S. Voros, M. Hübner, D. Goehringer, M. Dagioglou, T. Giannakopoulos, S. Konstantopoulos, and V. Karkaletsis, “Robots in assisted living environments as an unobtrusive, efficient, reliable and modular solution for independent ageing: The radio perspective,” in *International Symposium on Applied Reconfigurable Computing*, pp. 519–530, Springer, 2015.
- [4] D. Fischinger, P. Einramhof, K. Papoutsakis, W. Wohlkinger, P. Mayer, P. Panek, S. Hofmann, T. Koertner, A. Weiss, A. Argyros, *et al.*, “Hobbit, a care robot supporting independent living at home: First prototype and lessons learned,” *Robotics and Autonomous Systems*, vol. 75, pp. 60–78, 2016.
- [5] S. Coșar, M. Fernandez-Carmona, R. Agrigoroaie, J. Pages, F. Ferland, F. Zhao, S. Yue, N. Bellotto, and A. Tapus, “Enrichme: Perception and interaction of an assistive robot for the elderly at home,” *International Journal of Social Robotics*, pp. 1–27, 2020.
- [6] D. Portugal, P. Alvito, E. Christodoulou, G. Samaras, and J. Dias, “A study on the deployment of a service robot in an elderly care center,” *International Journal of Social Robotics*, vol. 11, no. 2, pp. 317–341, 2019.
- [7] N. Hawes, C. Burbridge, F. Jovan, L. Kunze, B. Lacerda, L. Mudrova, J. Young, J. Wyatt, D. Hebesberger, T. Kortner, *et al.*, “The strands project: Long-term autonomy in everyday environments,” *IEEE Robotics & Automation Magazine*, vol. 24, no. 3, pp. 146–156, 2017.

- [8] G. Wilson, C. Pereyda, N. Raghunath, G. de la Cruz, S. Goel, S. Nesaei, B. Minor, M. Schmitter-Edgecombe, M. E. Taylor, and D. J. Cook, “Robot-enabled support of daily activities in smart home environments,” *Cognitive Systems Research*, vol. 54, pp. 258–272, 2019.
- [9] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social lstm: Human trajectory prediction in crowded spaces,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 961–971, June 2016.
- [10] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, “Social gan: Socially acceptable trajectories with generative adversarial networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2255–2264, June 2018.
- [11] Y. Xu, Z. Piao, and S. Gao, “Encoding crowd interaction with deep neural network for pedestrian trajectory prediction,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5275–5284, June 2018.
- [12] J. Amirian, J. Hayet, and J. Pettré, “Social ways: Learning multi-modal distributions of pedestrian trajectories with gans,” *CoRR*, vol. abs/1904.09507, 2019.
- [13] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, and S. Savarese, “Sophie: An attentive GAN for predicting paths compliant to social and physical constraints,” *CoRR*, vol. abs/1806.01482, June 2018.
- [14] D. Ridel, N. Deo, D. Wolf, and M. Trivedi, “Scene compliant trajectory forecast with agent-centric spatio-temporal grids,” *IEEE Robotics and Automation Letters*, vol. PP, pp. 1–1, 02 2020.
- [15] J. Liang, L. Jiang, J. C. Niebles, A. G. Hauptmann, and L. Fei-Fei, “Peeking into the future: Predicting future person activities and locations in videos,” *CoRR*, vol. abs/1902.03748, 2019.
- [16] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. S. Torr, and M. K. Chandraker, “DESIRE: distant future prediction in dynamic scenes with interacting agents,” *CoRR*, vol. abs/1704.04394, 2017.
- [17] P. Dendorfer, A. Osep, and L. Leal-Taixé, “Goal-gan: Multimodal trajectory prediction based on goal position estimation,” *CoRR*, vol. abs/2010.01114, 2020.

- [18] J. Gu, C. Sun, and H. Zhao, “Densetnt: End-to-end trajectory prediction from dense goal sets,” *CoRR*, vol. abs/2108.09640, 2021.
- [19] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: A survey,” *ACM Computing Surveys*, vol. 38, pp. 1–45, 01 2006.
- [20] T. Huang and S. Russell, “Object identification in a bayesian context,” in *IJCAI*, vol. 97, pp. 1276–1282, Citeseer, 1997.
- [21] J. Yang, W.-S. Zheng, Q. Yang, Y.-C. Chen, and Q. Tian, “Spatial-temporal graph convolutional network for video-based person re-identification,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3289–3299, 2020.
- [22] P. Voigtlaender, J. Luiten, P. H. Torr, and B. Leibe, “Siam r-cnn: Visual tracking by re-detection,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6578–6588, 2020.
- [23] Y. Zhong, X. Wang, and S. Zhang, “Robust partial matching for person search in the wild,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6827–6835, 2020.
- [24] S. Gao, J. Wang, H. Lu, and Z. Liu, “Pose-guided visible part matching for occluded person reid,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11741–11749, 2020.
- [25] N. Wojke, A. Bewley, and D. Paulus, “Simple online and realtime tracking with a deep association metric,” in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649, IEEE, 2017.
- [26] N. Wojke and A. Bewley, “Deep cosine metric learning for person re-identification,” in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 748–756, IEEE, 2018.
- [27] P. Bergmann, T. Meinhardt, and L. Leal-Taixé, “Tracking without bells and whistles,” *CoRR*, vol. abs/1903.05625, 2019.
- [28] A. Milan, L. Leal-Taixé, I. Reid, and S. Roth, “Mot16: A benchmark for multi-object tracking,” *CoRR*, vol. abs/1603.00831, March 2016.
- [29] R. Martín-Martín, H. Rezatofighi, A. Sheno, M. Patel, J. Gwak, N. Dass, A. Federman, P. Goebel, and S. Savarese, “Jrdb: A dataset and benchmark

- for visual perception for navigation in human environments,” *CoRR*, vol. abs/1910.11792, Oct. 2019.
- [30] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, “You’ll never walk alone: Modeling social behavior for multi-target tracking,” in *2009 IEEE 12th International Conference on Computer Vision*, pp. 261–268, Sep. 2009.
- [31] A. Lerner, Y. Chrysanthou, and D. Lischinski, “Crowds by example,” *Computer graphics forum*, vol. 26, pp. 655–664, Sep. 2007.
- [32] P. Dollár, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: A benchmark,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 304–311, IEEE, 2009.
- [33] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, “Learning social etiquette: Human trajectory understanding in crowded scenes,” in *European conference on computer vision*, pp. 549–565, Springer, 2016.
- [34] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” *CoRR*, vol. abs/1804.02767, 2018.
- [35] “Pepper Robot.” <https://www.softbankrobotics.com/emea/en/pepper>. [Online; accessed 11 July 2022].
- [36] “Robot Operating System.” <https://www.ros.org/>. [Online; accessed 11 July 2022].
- [37] A. Ș. Ghiță, A. F. Gavril, M. Nan, B. Hoteit, I. A. Awada, A. Sorici, I. G. Mocanu, and A. M. Florea, “The amiro social robotics framework: deployment and evaluation on the pepper robot,” *Sensors*, vol. 20, no. 24, p. 7271, 2020.
- [38] K. Bernardin and R. Stiefelhagen, “Evaluating multiple object tracking performance: the clear mot metrics,” *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–10, 2008.
- [39] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, “Performance measures and a data set for multi-target, multi-camera tracking,” in *European conference on computer vision*, pp. 17–35, Springer, 2016.
- [40] J. Luiten, A. Osep, P. Dendorfer, P. Torr, A. Geiger, L. Leal-Taixé, and B. Leibe, “Hota: A higher order metric for evaluating multi-object tracking,” *International Journal of Computer Vision*, pp. 1–31, 2020.