

PETRA

Detecția și urmărirea persoanelor pentru roboți sociali și mașini autonome

(People detection and tracking for social robots and autonomous cars)

Etapa III - 2022

Raport științific și tehnic

Cuprins

1. Introducere	1
2. Metoda dezvoltată pentru urmărirea și re-identificarea persoanelor	3
3. Integrarea soluției dezvoltate în două tipuri de aplicații	5
3.1. Urmărirea persoanelor în contextul Roboților Sociali	5
3.2. Urmărirea pietonilor în contextul Mașinilor Autonome	13
4. Predicția traiectoriilor vehiculelor înconjurătoare	17
5. Concluzii	18
Bibliografie	19

1. Introducere

Detecția, urmărirea și recunoașterea persoanelor este o capacitate valoroasă pentru aplicațiile de vedere computerizată. Cu toate acestea, aceste sarcini sunt destul de dificil de realizat în mod autonom și, deși s-au obținut rezultate semnificative, acestea sunt încă o provocare tehnologică majoră. Urmărirea persoanelor, spre deosebire de alte sarcini de recunoaștere și interpretare, este dificilă atât din punctul de vedere al recunoașterii și predicției traiectoriei, cât și din cel al identificării instanțelor reale. Persoane parțial vizibile, reflexii în oglinzi sau ferestre și obiecte care seamănă foarte mult între ele, toate impun ambiguități intrinseci, astfel încât chiar și oamenii ar putea să nu fie de acord cu o aceeași soluție. Mai mult, stabilirea unor metrici de evaluare cu parametri liberi și definiții ambigue (cum se întâlnește des în literatură) duc adesea la rezultate cantitative contradictorii.

Obiectivul principal al proiectului PETRA este dezvoltarea unei platforme software pentru dezvoltarea de aplicații care necesită detecția și urmărirea persoanelor în medii reale. Scopul este acela de a realiza o platformă și un set de algoritmi asociați care să poată fi ușor utilizați și integrați în aplicații care necesită atât detecția și urmărirea persoanelor de către roboții sociali în spații închise, cât și efectuarea aceleiași sarcini în spații deschise, în particular detecția și urmărirea pietonilor.

În cadrul primei etape a proiectului s-au analizat diferite arhitecturi de rețele neurale adânci pentru detecția de persoane și pentru urmărirea de persoane, atât pentru spații închise (indoor) – caz potrivit pentru aplicații ale roboților sociali, cât și spații deschise (outdoor) – caz potrivit

pentru pietoni în aplicații de conducere autonomă. S-au comparat și evaluat aceste arhitecturi și s-au propus arhitecturi pentru detecția și urmărirea persoanelor. S-au investigat diferite seturi de date existente în domeniul public, s-au comparat aceste seturi de date din punct de vedere al performanțelor realizate pe diferite arhitecturi adânci și s-a colectat propriul set de date pentru detecția pietonilor și a mașinilor.

Obiectivul Etapei II (2021) a proiectului a fost acela de a realiza soluții noi pentru predicția traiectoriei persoanelor în scopul urmăririi mișcării acestora și utilă pentru a putea estima unde se vor duce persoanele în perioada următoare, ce vor face, și eventual a lua diverse decizii în cazurile care impun efectuarea unor acțiuni (de exemplu în cazul pietonilor dintr-un scenariu de conducere autonomă, ce decizii trebuie să ia șoferul automat). Pentru realizarea acestui obiectiv am propus, implementat și evaluat *două abordări*: (i) o metodă de extragere a informațiilor globale ale scenei din imaginile de vizualizare 2D de la nivelul liniei orizontului, prin combinarea unei transformări de perspectivă de la vedere la nivelul ochiului (*eye-level view*) la vedere la nivel de pasăre (*bird-level view*), cu o tehnică de segmentare care determină parametrii pentru transformare, și (ii) un model generic de predicție a secvențelor video, din care deducem traiectoria. În ambele metode, au fost incluse metode de detecție a persoanelor din secvențele video utilizate.

În cadrul Etapei III (2022) ne-am focalizat pe consolidarea soluției de predicție a traiectoriei persoanelor în urmărire, inclusiv soluții de re-identificare în cazurile mai dificile în care există mai multe persoane care se deplasează și în care, în timpul deplasării, persoanele sunt acoperite de diferite ocluzii (obiecte). Am validat soluția noastră în două tipuri de aplicații: aplicație de urmărire a persoanelor pentru roboți sociali și aplicație de urmărire a persoanelor pentru cazul pietonilor, sub-componentă a unui modul de conducere autonomă a unui autovehicul. Astfel am dezvoltat o tehnologie proprie de reidentificare și urmărire a persoanelor în timp real bazat pe predicția traiectoriei. Rezultatele Etapei III sunt prezentate în secțiunile următoare.

Funcția unui robot social asistiv [1] este de a ajuta oamenii în activitățile lor prin implicarea în interacțiuni cu aceștia. Fie că ne referim la oferirea de informații, îndrumări, notificări, instrucțiuni etc., fie doar la a ține companie și a menține o conversație simplă, un robot social trebuie să fie capabil să identifice și să recunoască corect persoana cu care interacționează. În contextul roboticii sociale, o componentă de detectare, reidentificare și urmărire a persoanei este indispensabilă. Pentru ca un robot să fie capabil să analizeze și să înțeleagă comportamentul oamenilor cu care interacționează, acesta trebuie să fie capabil să cartografieze mediul și să modeleze cu precizie pozițiile acestora.

Un vehicul cu conducere autonomă este capabil să se deplaseze în siguranță prin împrejurimi, fără intervenția umană. Poate naviga cu sau fără pasageri între două locații. Mișcarea unui autoturism autonom trebuie să fie fiabilă, asigurând atât protecția pasagerilor mașinii, cât și siguranța celorlalți participanți la trafic, fie că ne referim la alți șoferi sau la pietoni. Pentru a construi un sistem stabil este necesar un sistem de urmărire a persoanei și de predicție a traiectoriei, pentru ca vehiculul să poată analiza și preveni situațiile riscante.

Un sistem de reidentificare a persoanei atribuie aceiași identificatori exact acelorași oameni într-o succesiune de imagini. Caracteristicile unei astfel de componente nu se limitează doar la aplicațiile ce folosesc roboți sociali sau mașini autonome, ci pot fi extinse la orice dispozitiv autonom. Fie că vorbim de vehicule autonome, roboți, supraveghere sau sisteme inteligente pentru case, 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 unei persoane are o natură complexă, deoarece performanța unui astfel de sistem poate fi puternic influențată de 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 vestimentația sau coafura, sau interacțiunile dintre indivizi. Pentru a gestiona aceste tipuri de probleme, metoda propusă de noi folosește informații suplimentare generate de un sistem 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 poate părea simplu, modelarea procesului complex din spatele comportamentului uman este o sarcină dificilă. Traiectoriile urmate de oamenii din mediu sunt influențate de mișcarea celorlalte persoane din vecinătate, obstacolele din mediu și punctele de interes din scenă. Componenta de predicție a traiectoriei adăugată sistemului de reidentificare este necesară pentru a integra factorii care pot influența mișcarea persoanei pentru a obține un rezultat precis.

Sistemul de reidentificare și urmărire în timp real a oamenilor propus are la bază o metodă de predicție a traiectoriei. Scopul acestei cercetări este de a îmbunătăți o tehnică standard de urmărire în timp real pe care o folosim în cadrul unei arhitecturi robotice prin încorporarea informațiilor dintr-un sistem de predicție a traiectoriei. Sistemul de predicție a traiectoriei include atât influența socială care vine din partea oamenilor din jur, cât și informații vizuale din mediu pentru a estima pozițiile viitoare. Ne-am validat abordarea efectuând mai multe experimente pe două aplicații posibile: roboți sociali și mașini autonome. Pentru a putea valida sistemul în contextul unei aplicații de robotică socială, am integrat sistemul nostru în cadrul platformei AMIRO [2] pentru a îmbunătăți comportamentul robotului Pepper [3]. Aplicația AMIRO este o platformă modulară folosită pentru a programa orice robot compatibil cu sistemul de operare robot ROS [4]. Considerând contextul unei aplicații de conducere autonomă, am validat abordarea noastră folosind videoclipuri pe care le-am înregistrat pentru a simula un scenariu din viața reală. Am raportat rezultatele obținute de sistemul de reidentificare și urmărire a persoanelor atât pe seturi de date existente, cât și pe fluxuri de imagini. Am prezentat câteva exemple din experimentele noastre, scenarii în care sistemul propus rezolvă probleme complexe, în timp ce tehnicile existente de urmărire în timp real nu ar obține aceleași rezultate precise.

2. Metoda dezvoltată pentru urmărirea și re-identificarea persoanelor

Abordarea noastră propune un sistem modular care abordează problema reidentificării și urmăririi persoanelor în timp real, combinând o tehnică standard de urmărire a persoanelor cu o metodă de predicție a traiectoriei persoanelor. Metoda de predicție a traiectoriei oamenilor combină modelul mișcării persoanei cu informații contextuale externe pentru a prezice pozițiile viitoare posibile ale persoanei din imagini. Informațiile generate de metoda de predicție a traiectoriei sunt apoi combinate cu informațiile generate de un sistem standard de urmărire pentru a genera un sistem mai precis de reidentificare în timp real.

Scopul final al sistemului nostru este de a putea urmări și reidentifica oamenii într-un flux de imagini continuu. Sistemul trebuie să poată recunoaște (reidentifica) oamenii asociind aceleași numere de identificare exact acelorași persoane printr-o succesiune de imagini.

Arhitectura pe care o propunem ia ca intrare un flux de imagini și prezice un număr de identificare pentru fiecare persoană detectată în imagine. Figura de mai jos prezintă arhitectura propusă pentru sistemul de reidentificare și urmărire în timp real. Arhitectura propusă poate fi integrată atât pentru un robot social, cât și în sistemul unui vehicul autonom, deoarece necesită doar date RGB pentru a genera rezultate.

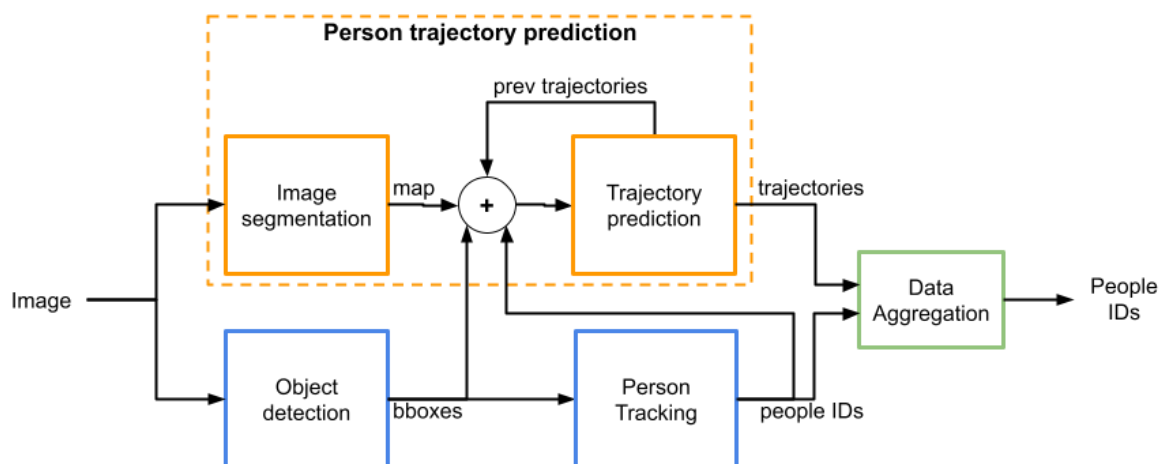


Figura 1. Arhitectura de reidentificare și urmărire a persoanelor în timp real bazată pe o componentă de predicție a traiectoriei persoanelor.

Am combinat cinci module pentru sistemul de reidentificare și urmărire a persoanei în timp real: *deteția obiectelor*, *urmărirea standard a persoanelor*, *segmentarea imaginii*, *predicția traiectoriei* și *agregarea datelor*.

Modulul *deteția obiectelor* generează dreptunghiuri de delimitare reprezentând pozițiile oamenilor pentru fiecare imagine primită ca intrare. Există mai multe rețele care obțin rezultate precise pentru problema detecției obiectelor. În experimentele noastre am folosit arhitectura YOLO [5], deoarece se potrivește cerințelor noastre de performanță și viteză. Pentru restul experimentelor, am folosit detecțiile pre-calculat disponibile în seturile de date pe care le-am inclus în evaluare.

Modulul *urmărirea standard a persoanelor* este o tehnică standard de urmărire care asociază un set de identificatori pentru fiecare persoană dintr-o imagine. Acesta utilizează sistemul introdus în [6] și [7]. Combină o tehnică simplă de estimare a pozițiilor viitoare cu o rețea convoluțională care calculează caracteristicile vizuale pentru persoanele detectate. Modulul generează rezultate în timp real pentru problema de reidentificare.

Modulele *segmentarea imaginii* și *predicția traiectoriei* sunt componentele metodei de predicție a traiectoriei persoanelor descrise în raportul anterior. Componenta *segmentarea imaginii* încorporează o rețea de segmentare semantică pre-antrenată pentru a extrage informații semantice din imagini, pentru predicții mai precise ale traiectoriilor. Am folosit rețeaua ResNet50 [8] combinată cu un modul Pyramid Pooling [9], deoarece calculează rezultate destul de precise într-un interval scurt de timp, necesitate introdusă de cerințele în timp real ale aplicației. Pentru a reduce și mai mult timpul de calcul pentru aplicații în timp real, modulul de segmentare este rulat în paralel cu detectorul de obiecte.

Modulul *predicția traiectoriei* generează viitoare poziții posibile pentru persoanele urmărite în imagini. Pentru a calcula rezultatele, este nevoie de secvențe consecutive de poziții observate ale oamenilor din imagini, alături de date vizuale din scenă. Pozițiile necesare sunt determinate de dreptunghiurile de delimitare extrase de detectorul de obiecte, iar informațiile despre mediu sunt extrase de o rețea de segmentare semantică. În arhitectura de reidentificare și urmărire în timp real, modulul *predicția traiectoriei* necesită informații suplimentare pentru a putea genera traiectorii. Pentru pozițiile inițiale observate ale persoanelor, modulul utilizează datele generate de un modul de urmărire, iar pentru restul procesului, completează informațiile lipsă cu rezultatele generate anterior.

Pentru fiecare imagine preluată de platforma autonomă, sistemul calculează identificatorii și traiectoriile viitoare pentru persoanele detectate. Pentru a putea transforma traiectoriile prezise în identificatori, se combină informațiile legate de detecțiile curente ale oamenilor (dreptunghiurile de delimitare) și traiectoriile prezise la pasul anterior. Fiecare traiectorie viitoare estimată în pasul anterior este asociată în pasul curent uneia dintre dreptunghiurile de delimitare extrase de detectorul de obiecte. Asocierea se face prin calculul distanțelor euclidiene dintre pozițiile dreptunghiurilor de delimitare și prima poziție estimată a fiecărei traiectorii generate. Dacă distanța minimă euclidiană se află într-un prag predefinit, atunci persoana respectivă, care este reprezentată de cel mai apropiat dreptunghi de delimitare, este asociată cu numărul de identificare care a corespuns acelei traiectorii la pasul anterior. Dacă există o traiectorie care nu are un dreptunghi de delimitare corespunzător, ceea ce poate fi cazul în care o persoană este blocată de un obstacol, atunci modulul completează informațiile curente și adaugă ca poziție curentă a persoanei blocate poziția estimată din traiectoria prezisă anterior. Prin completarea informațiilor pentru persoanele care nu apar în imagini cu pozițiile estimate, modulul le poate prezice în continuare posibilele poziții viitoare, iar atunci când oamenii reapar în imagine le poate reasocia identificatorul pe care l-au avut anterior.

Modulul *agregarea datelor* combină rezultatele obținute de *predicția traiectoriei* și *urmărirea standard a persoanelor* pentru a genera ID-uri de reidentificare mai precise pentru persoanele din imagini. Fiecare dintre cele două module, *predicția traiectoriei* și *urmărirea standard a persoanelor* estimează identificatori diferiți pentru persoanele din imagine la un moment dat. Modulul de agregare a datelor calculează ID-urile finale pe baza încrederii și a rezultatelor celor două module. Cu fiecare imagine, modulul de predicție a traiectoriei câștigă mai multă încredere pe măsură ce dobândește mai multe informații despre traiectoriile anterioare. În cazul în care modulul de predicție a traiectoriei nu reușește să asocieze ID-uri fezabile persoanelor detectate, ID-urile estimate sunt exclusiv cele obținute de modulul *urmărirea standard a persoanelor*.

3. Integrarea soluției dezvoltate în două tipuri de aplicații

Am testat și analizat performanța sistemului propus în diverse scenarii și aplicații. Scopul proiectului nostru este de a aplica sistemul de reidentificare pe o platformă autonomă, fie că vorbim despre roboți sau mașini autonome. Toate experimentele executate au fost realizate din punctul de vedere al unei platforme autonome. Arhitectura și rezultatele procesate au fost obținute având în vedere cerințele în timp real ale unei astfel de aplicații. Am prezentat rezultatele obținute luând în considerare două posibile cazuri de utilizare: roboți sociali și vehicule cu conducere autonomă. Am evaluat metoda propusă atât pe seturile de date existente, cât și pe date înregistrare de noi, pentru a obține o evaluare obiectivă a performanței sistemului de reidentificare în diferite scenarii.

Componenta de reidentificare și urmărire a persoanelor în timp real este evaluată pe studii de caz relevante pentru proiectul nostru, robotică socială și conducere autonomă. Unele părți ale informațiilor prezentate în secțiunea dedicată roboților sociali, și anume timpul de procesare, setările experimentelor, abordarea informațiilor lipsă, sunt aplicabile și pentru secțiunea dedicată mașinilor autonome. Am obținut evaluarea componentei și pe un caz auxiliar, luând în considerare imaginile achiziționate de sus, pentru a demonstra generalitatea soluției.

3.1. Urmărirea persoanelor în contextul Roboților Sociali

Pentru a evalua rezultatele luând în considerare o aplicație pentru roboți sociali, am integrat sistemul în cadrul platformei AMIRO [2], o platformă pentru operarea roboților în contextul

aplicațiilor sociale, cu scopul de a îmbunătăți performanța comportamentului general al robotului Pepper. Arhitectura sistemului propus a fost creată având în considerare modul în care se poate integra în cadrul platformei AMIRO, a cărei arhitecturi este prezentată în figura de mai jos.

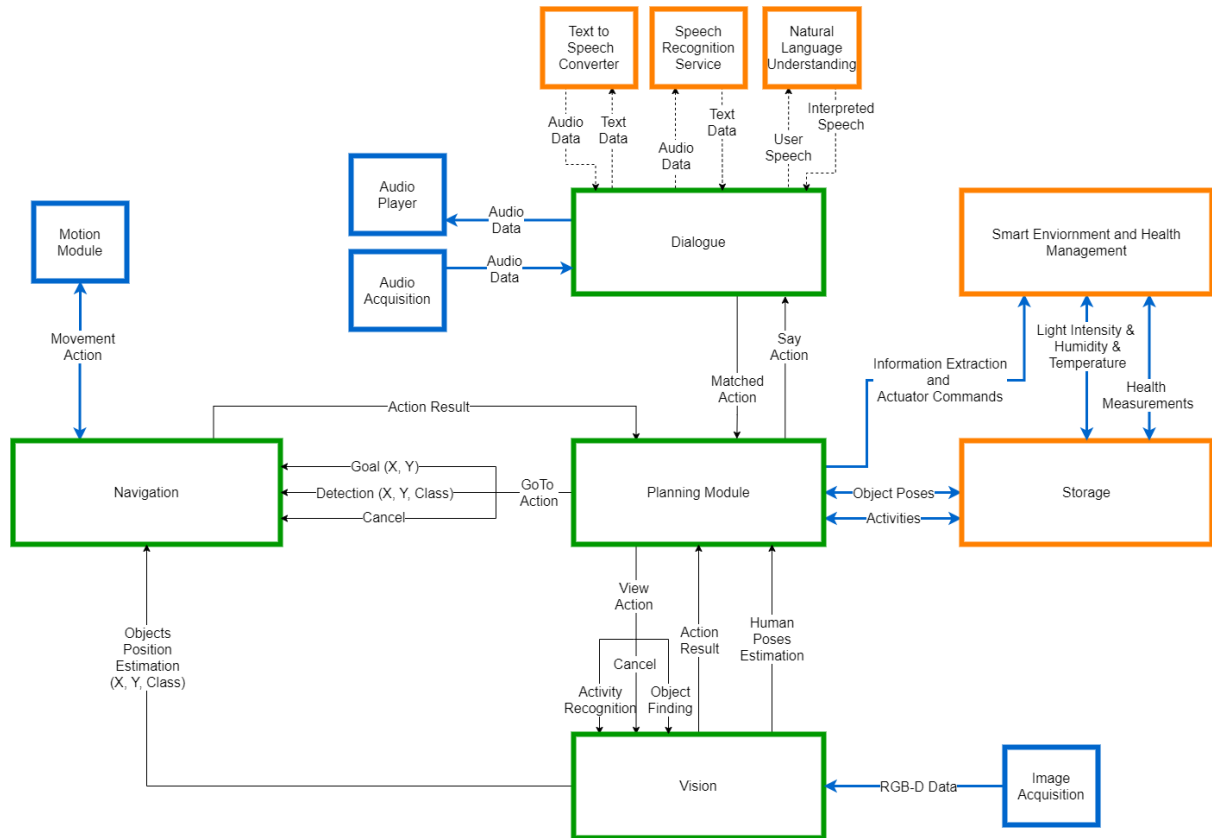


Figura 2. Arhitectura platformei AMIRO pentru roboți sociali [2]. Arhitectura propusă este integrată în interiorul componentei *Vision*.

Arhitectura platformei AMIRO, așa cum este prezentată în figură, este construită ca o compoziție de module independente care comunică în timp ce funcționează în paralel. Datele generate de fiecare componentă principală sunt procesate independent și apoi trimise la modulul de bază. Sistemul propus este integrat în componenta *Vision* a platformei AMIRO și generează identificatori asociați fiecărei persoane dintr-o imagine. Scopul sistemului nostru este de a reidentifica oamenii într-o succesiune de imagini, asociind aceleași numere exact aceluiași persoane. Reidentificarea prin identificator a unei persoane este mai fiabilă decât recunoașterea pe față, deoarece nu are nevoie ca fața să fie vizibilă pe parcursul întregii interacțiuni. Îmbunătățirea componentei de reidentificare ar trebui să îmbunătățească considerabil comportamentul general al robotului, luând în considerare sub-module precum recunoașterea activității sau găsirea unei persoane.

Având în vedere schema sistemului propus, integrarea în cadrul platformei AMIRO este simplă. Imaginile transmise sistemului sunt achiziționate de la robot prin modulul *Image Acquisition*. Folosim modulul de detectare existent în interiorul componentei *Vision* pentru a calcula dreptunghiurile de delimitare ale persoanelor și asociem fiecărui dreptunghi un număr de identificare generat de sistemul propus. Restul modulelor din cadrul platformei AMIRO continuă să funcționeze în același mod, cu îmbunătățirea recunoașterii persoanelor după numărul de identificare.

Unul dintre obiectivele proiectului nostru este de a îmbunătăți robustețea comportamentului robotului Pepper în contextul aplicațiilor de robotică socială. Unul dintre obiectivele noastre este de a proiecta o arhitectură robustă care să poată aborda cu succes provocările vieții asistate ambientale. Intenționăm să folosim robotul Pepper în genul de aplicații în care mediul nu este foarte aglomerat, astfel încât robotul să își poată concentra atenția asupra unei singure persoane odată, pentru interacțiuni semnificative. Robotul poate interacționa cu o persoană de acasă sau poate funcționa și ca punct de informare sau ghid într-un muzeu. Având în vedere scenariile posibile pentru un astfel de robot, am evaluat sistemul de urmărire și reidentificare propus prin reproducerea acelorași condiții posibile. Am analizat rezultatele obținute de sistem în diferite medii. Sistemul a fost testat atât în medii interioare, cât și în exterior, cu una sau mai multe persoane în imagini.

Experimentele pe care le-am efectuat au testat diverse scenarii care pot apărea în timp ce un robot urmărește o persoană. 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 prin mișcare, schimbarea bruscă a mișcării oamenii.

Atât urmărirea standard, cât și urmărirea bazată pe predicția traiectoriei gestionează bine circumstanțele simple, fără schimbarea numerelor asociate pentru persoanele urmărite. Situațiile care nu implică ocluzii sau nicio modificare a aspectului vizual al oamenilor nu sunt greu de gestionat de ambele sisteme. Cu toate acestea, sistemul actual de urmărire utilizat în cadrul AMIRO prezintă probleme atunci când o persoană este în spatele unui obstacol pentru un număr de cadre sau aspectul fizic se schimbă. Figurile de mai jos sunt exemple de situații în care tehnica standard de urmărire este nu funcționează conform așteptărilor. În toate cele patru figuri, dreptunghiurile albastre reprezintă dreptunghiurile de delimitare ale persoanelor detectate generate de modulul de detecție de obiecte, numărul galben de deasupra unei casete de delimitare reprezintă numărul de reidentificare asociat unei persoane prin tehnica standard de urmărire și numărul roșu în interiorul unei casete de delimitare reprezintă numărul de reidentificare asociat de sistemul propus. Fiecare imagine conține și traiectoriile viitoare prezise pentru persoanele care au fost observate pentru un număr minim de poziții. Figurile de mai jos conțin doar o parte din imaginile experimentului, respectiv cele care sunt relevante pentru a înțelege îmbunătățirile aduse.

Figura 3 arată un experiment în care o persoană este complet blocată pentru mai multe cadre consecutive. Persoana se deplasează de la stânga imaginii la dreapta, în timp ce trece prin spatele unui obstacol. Persoana, care este identificată inițial ca p1 de către sistemul de urmărire standard, este schimbată la p2 când este re-detectată după cadrele lipsă. Sistemul nostru corectează acest comportament și atribuie același identificador p1, în ciuda cadrelor lipsă. Traectoria prezisă asociată cu persoana înainte de ocluzie, care este prezentată în a doua imagine din figură, estima că p1 va merge drept și va trece de obstacole. Când persoana a fost re-detectată, deoarece detecția se potrivea cu poziția estimată la momentul, i s-a atribuit același număr de identificare.

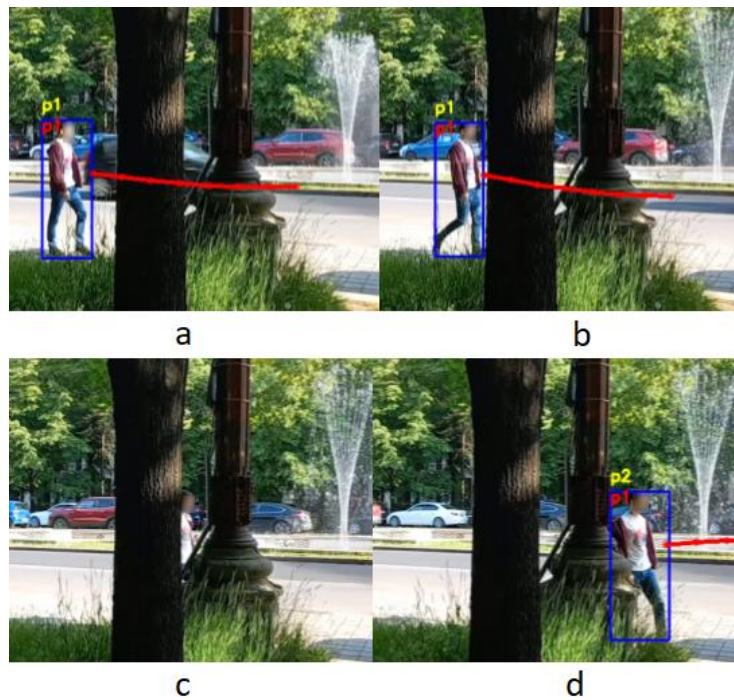


Figura 3. Reidentificarea persoanei pe baza predicției traiectoriei în cazul în care o persoană este complet ascunsă de un obstacol pentru un număr de imagini.

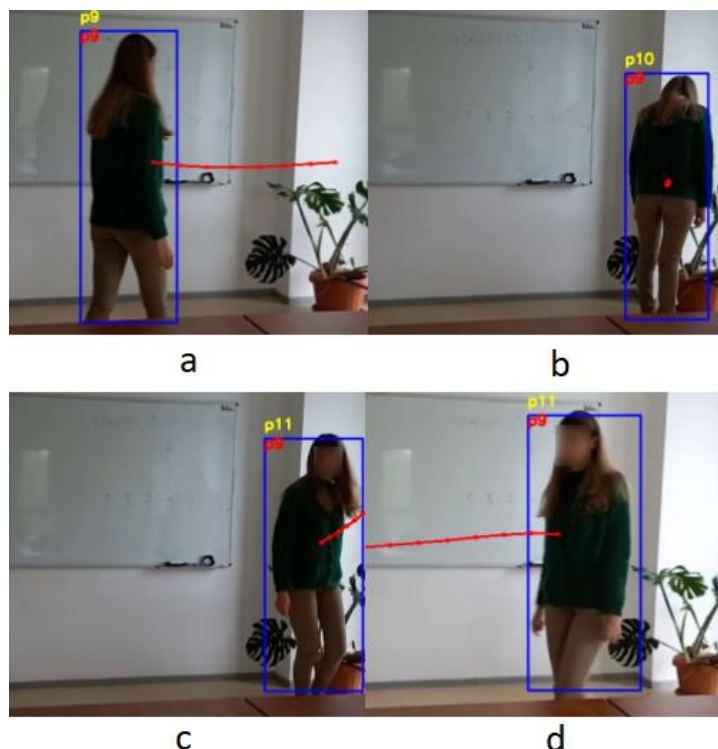


Figura 4. Reidentificarea persoanei pe baza predicției traiectoriei în cazul unei persoane care își schimbă direcția inițială.

Figura 4 prezintă un scenariu în care o persoană merge într-o anumită locație, apoi își schimbă direcția și revine la poziția inițială. Acest scenariu produce schimbări bruște în traiectoria urmată de persoana urmărită și în informațiile vizuale extrase de la persoană. În acest scenariu, persoana care merge la ghiveciul din dreapta imaginii și apoi se întoarce, i se atribuie inițial numărul de identificare p9. Tehnica standard de urmărire trece la p10 când persoana se oprește

din mișcare și se aplecă spre plantă și trece din nou la p11 când persoana se rotește și se îndreaptă spre cameră. Sistemul propus rezolvă aceste probleme și păstrează același identificator inițial, p9, pentru întreaga secvență de imagini. Imaginile prezintă și traiectoriile viitoare prezise în timpul experimentului, pentru o mai bună înțelegere a modului în care acestea se adaptează dinamic.

Figurile următoare prezintă imagini din experimente pe care le-am efectuat cu mai multe persoane din imagini. În imagini, pe măsură ce oamenii se mișcă, există ocluzii și variații ale traiectoriilor și aspectului vizual.

Figura 5 este un exemplu de experiment cu patru persoane urmărite. Persoanele p3, p11, care se află în spatele imaginii și p5 sunt inițial urmărite corect de sistemul standard de urmărire. Când persoana p11 nu mai este vizibilă deoarece este ascunsă de p5, sistemul de urmărire schimbă identificatorul pentru p3 la p11. Când p11 devine din nou vizibil, identificatorul pentru p3 trece la valoarea inițială, p5 trece la p11 și lui p11 i se atribuie un nou identificator, care este p16. Mai mult, bebelușul din cărucior, căruia i se atribuie inițial identificatorul p14, este schimbat în p17 în timp ce se deplasează în dreapta imaginii. Performanța sistemului de urmărire în acest scenariu ar avea un impact extrem de mare asupra performanței unui sistem robotic, deoarece acesta nu ar putea fi capabil să se concentreze corect asupra unei persoane. Sistemul nostru bazat pe predicția traiectoriei reidentifică corect toți oamenii din scenariu, indiferent dacă apar ocluzii sau nu. Deoarece traiectoria prezisă a persoanei p11 estima că se va deplasa către poziția inițială a lui p5 după un număr de cadre, atunci când a fost re-detectată, i s-a reatribuit corect același număr de identificare. Trajectoriile viitoare prezise pentru p3, p5 și p14 au oferit sistemului informațiile necesare pentru a menține numerele de identificare corecte pentru oameni în timpul experimentului.

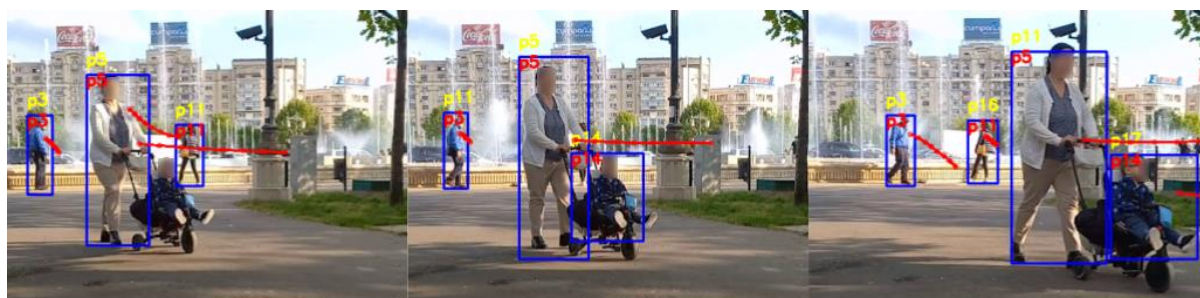


Figura 5. Reidentificarea persoanei pe baza predicției traiectoriei în cazul suprapunerii persoanelor.

Figura 6 prezintă un alt exemplu de urmărire multiplă. Acest exemplu particular prezintă modul în care sistemul funcționează pentru un scenariu mai lung și mai complex. Persoana p4 merge spre stânga imaginii, după ce a evitat un copac. În timp ce p4 urmează acest traseu, un grup de oameni trec prin fața camerei care acoperă p4 pentru un număr de cadre. În acest scenariu, p4 este blocat atât de un obstacol, cât și de alți participanți ai mediului. Sistemul de urmărire standard atribuie inițial numărul de identificare corect atunci când se intersectează cu p3, dar după ce a fost blocat pentru mai multe cadre, sistemul schimbă ID-ul la p14. Mai mult, putem observa că persoana p3 este schimbată în p5 după intersectarea cu p4. Sistemul nostru propus urmărește corect atât p3 cât și p4, având în vedere traiectoriile prezise. Acest exemplu arată, de asemenea, cum se adaptează traiectoriile oamenilor cu fiecare detecție, de la deplasarea inițială către centrul imaginii pentru persoana p4, până la ajustarea traiectoriei spre stânga imaginii.



Figura 6. Reidentificarea persoanei pe baza predicției traiectoriei în cazul suprapunerilor persoanelor și a ocluziilor generate de mediu.

Precizia componentei de reidentificare și urmărire a persoanelor depinde foarte mult de celelalte componente ale platformei AMIRO: achiziție de imagini, detecția obiectelor, urmărirea standard a persoanelor. Platforma robotică pe care o folosim oferă imagini la o rezoluție de 640x480 cu număr variabil de cadre pe secundă. Pentru a avea o evaluare mai obiectivă a performanței sistemului, am folosit setul de date MOT17 [10] pentru a calcula [11] mai multe valori pentru problema urmăririi persoanelor. Setul de date MOT17 conține imagini achiziționate din punctul de vedere al unui robot în mai multe medii, atât în interior, cât și în exterior. Am folosit acest set de date deoarece este cel mai potrivit atunci când vorbim despre urmărirea persoanelor. Valorile pe care le-am calculat sunt MOTA (Multiple Object Tracking Accuracy) [12], IDF1 (Identificare F1) [13] și HOTA (Higher Order Tracking Accuracy) [14]. Tabelul de mai jos raportează metricele pentru problema urmăririi persoanelor obținute pe baza fișierelor cu detecții disponibile calculate folosind rețeaua SDP [15]. 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 [16]. Setul de date JRDB este o colecție de imagini înregistrate de o platformă robotică, atât în medii interioare, cât și în aer liber. Conține mai multe subseturi, cu setări diferite de scenă, lumini diferite și număr diferit de participanți. Am ales să integrăm acest set de date în evaluarea noastră, deoarece simulează condiții mai apropiate de cele reale. Setul de date JRDB a fost preprocesat pentru a simula o rată de cadre de 2,5 cadre pe secundă, pentru a se potrivi cu rata de cadre a robotului nostru. Dacă arhitectura ar fi fost antrenată pe setul de date MOT17, folosind viteza corectă a imaginilor, așa cum sunt subseturile din MOT17, rezultatele obținute prin metoda propusă ar fi fost și mai bune.

Tabela 1. Metricile pentru problema urmăririi persoanelor pentru sistemul de reidentificare bazat pe predicția traiectoriei pentru setul de date MOT17.

Metrică	Metodă standard de urmărire	Sistemul propus
MOTA (%)	55,05	61,04
IDF1 (%)	55,62	52,24
HOTA (%)	45,57	43,79

Rezultatele obținute mai sus sunt direct influențate de rezultatele sistemului de predicție a traiectoriei. Pentru a avea o evaluare obiectivă, am raportat și erorile obținute de componenta de predicție a traiectoriei. Tabela 2 prezintă un raport detaliat al rezultatelor obținute de sistem pentru subseturile conținute de setul de date JRDB. Valorile reprezintă erorile obținute pentru un subset, atunci când sistemul este antrenat pe restul subseturilor. Am luat în considerare doar seturile de date care includeau informații 3D despre persoanele din imagini. Tabela prezintă valorile în pixeli obținute pentru două metrici, ADE, eroarea medie a pozițiilor și FDE, eroarea poziției finale. Valorile calculează diferența dintre coordonatele estimate ale pixelilor și coordonatele persoanelor din imagini. Valorile au fost raportate pentru traiectorii prezise de lungime 8. Fiecare subset a fost preprocesat pentru a simula o rată de cadre de 2,5 cadre pe secundă. Subseturile din tabel sunt grupate în două clase, corespunzătoare celor două setări posibile de scenă, interior și exterior. După cum se poate observa, nu există nicio diferență semnificativă între rezultatele subseturilor de interior și exterior. Ceea ce se poate observa este că în unele subseturi valorile erorilor sunt foarte mici, cu o eroare finală de deplasare de 7 pixeli, în timp ce altele au o eroare finală de deplasare de aproximativ 60 pixeli. Diferența vine din tipul imaginilor din subseturi. De exemplu, în cazul subsetului *cubberly-auditorium*, imaginile includ o mulțime de oameni care sunt urmăriți în timp ce platforma robotică se mișcă prin mediu. Acest caz particular face dificilă o predicție mai precisă a traiectoriei, deoarece mișcarea ego-ului platformei trebuie luată în considerare. Indiferent de tipul scenei, valorile medii de eroare obținute pentru traiectorii sunt suficient de mici pentru un comportament satisfăcător al platformei robotice.

Tabela 2. Rezultatele metricilor ADE (Average Displacement Error) și FDE (Final Displacement Error) pentru fiecare subset din setul de date JRDB [16].

Tipul scenei	Subset JRDB	Erori obținute	
		ADE	FDE
interior	bytes-cafe_0	22.38	36.66
	clark-center_0	23.62	34.88
	cubberly-auditorium_0	42,96	65,64
	forbes-cafe_0	29.75	40.44
	gates-159-grup-meeting_0	32.21	52.37
	gates-ai-lab_0	12.24	20.95
	porti-subsol-ascensoare_1	42.54	69.33
	huang-2_0	26.46	39.48
	huang-basement_0	31,97	49,85
	jordan-hall_0	32.44	48.30
	nvidia-aud_0	30.53	45.81
	packard-poster-session_0	25.59	35.53
	packard-poster-session_1	7.77	13.61
	packard-poster-session_2	4.61	6.45

exterior	stlc-111_0	21.72	38.18
	svl-meeting-gates-2_0	15.22	23.62
	svl-meeting-gates-2_1	16.08	25.39
	tressider_2	15.07	21.67
	clark-center_1	35.77	53.31
	clark-center-intersection_0	24.19	34.64
	gates-to-clark_1	27.45	33.03
	hewlett-packard-intersection_0	10.60	20.24
	huang-lane_0	21.97	40.58
	memorial-court_0	29.97	36.70
	meyer-green_0	29,78	49,26
	tressider_0	17.23	28.76
	tressider_1	9.71	16.70
Eroare medie		23,69	36,34

Tabela 3 prezintă timpul mediu de procesare al sistemului luând în considerare un număr diferit de persoane urmărite. Timpul este măsurat din momentul în care modulul de predicție a traiectoriei primește datele până în momentul în care modulul de agregare a datelor generează rezultatul. După cum se poate observa, sistemul obține valori similare ale vitezei, indiferent de numărul de persoane urmărite. Acest lucru demonstrează că sistemul poate efectua urmărirea mai multor persoane în paralel în același timp de procesare pe care l-ar necesita pentru urmărirea unei singure persoane. Mai mult, timpul de procesare este rezonabil de mic, ceea ce îl face potrivit pentru aplicații în timp real.

Tabela 3. Timpul mediu de procesare al sistemului de reidentificare bazat pe traiectorie, luând în considerare un număr diferit de persoane urmărite.

Numărul de persoane urmărite	Viteza de procesare a sistemului propus (milisecunde)
1	102.06
2	106,55
3	106,63
4+	108,22

Toate experimentele pe care le-am efectuat au folosit predicții de traiectorii de 8 poziții viitoare. Ne-am hotărât asupra acestei constante luând în considerare frecvența de imagini generate de robotul Pepper și observațiile pe care le-am făcut despre cât de des oamenii își schimbă traiectoria. Rețeaua pe care am antrenat-o și am integrat-o în cadrul platformei AMIRO prezice traiectorii pe baza observațiilor a 8 poziții consecutive. Având în vedere că robotul Pepper integrat în cadrul nostru AMIRO are o viteză a numărului de imagini mai mică, pentru a putea estima rapid traiectoriile am decis să simulăm inițial traiectoria observată. A unei persoane plecând de la 3 poziții observate. Când o persoană este observată timp de 3 cadre consecutive, completăm restul datelor lipsă cu prima poziție detectată, creând o traiectorie care rămâne pe aceeași prima poziție observată timp de 6 cadre, apoi merge în direcția reprezentată de ultimele 2 poziții observate. Pe măsură ce sistemul primește din ce în ce mai multe date, pentru fiecare detecție primită de sistem pentru o persoană, una dintre pozițiile simulate este eliminată. Folosind acest mecanism, sistemul poate genera traiectorii mai rapid și nu necesită așteptarea a 8 detecții consecutive pentru aceeași persoană să genereze o traiectorie.

Sistemul abordează problema informațiilor lipsă având în vedere necesitatea unor observații continue utilizând pozițiile traiectoriilor prezise anterior. Dacă o persoană care a fost urmărită nu este detectată într-un cadru, se presupune că persoana este în spatele unui obstacol și că poziția curentă a persoanei coincide cu prima poziție a traiectoriei prezise la pasul anterior. Această metodă permite sistemului să completeze informațiile lipsă timp de 8 cadre în care persoana nu apare, care corespund celor 8 poziții viitoare prezise de modulul de predicție a traiectoriei. Dacă persoana nu este re-identificată pe parcursul a 8 cadre, atunci informațiile asociate cu acea persoană sunt eliminate.

3.2 Urmărirea pietonilor în contextul Mașinilor Autonome

Reidentificarea și urmărirea persoanelor în contextul conducerii autonome este o problemă mai dificilă decât în contextul roboticii sociale. Pentru a putea reidentifica corect o persoană într-o succesiune de imagini, sistemul necesită un model complex care poate codifica mișcarea camerei, variația dimensiunii persoanelor în imagini și ocluziile multiple. Mișcarea camerei este o constrângere importantă în problema reidentificării, deoarece implică nu numai modificări ale unghiului camerei, ci și fluctuații ale vitezei de mișcare. Acest factor reprezintă o mare provocare pentru aplicațiile care necesită reidentificări fiabile.

Experimentele pe care le-am efectuat au demonstrat că sistemul standard de urmărire este foarte susceptibil la erori atunci când este testat pe fluxuri de imagini înregistrate dintr-o mașină în mișcare. Sistemul de urmărire are o performanță acceptabilă în situațiile în care mașina se mișcă foarte încet și pietonii sunt aproape de cameră. Cu toate acestea, există o mulțime de schimbări ale identificatorilor asociați persoanelor atunci când mașina se mișcă cu o viteză normală, sau pietonii sunt la o distanță mai mare, sau mașina intră într-o curbă și schimbă unghiul camerei. Figurile de mai jos sunt trei exemple extrase din experimentele noastre pentru a demonstra îmbunătățirea pe care sistemul nostru o aduce.

Figura 7 prezintă un scenariu în care vehiculul virează la dreapta și pietonii traversează strada în fața mașinii. Vehiculul are o viteză normală în timpul deplasării, astfel încât vehiculul se apropie rapid de pietoni. Pe măsură ce mașina avansează rapid în direcția pietonilor, sistemul standard de urmărire comută numărul de identificare asociat pietonului detectat de la p4 la p7 între prima și a doua imagine. Această problemă apare ca urmare a dimensiunii mici pe care persoana o are în prima imagine și a creșterii rapide dimensiunii acestuia în a doua imagine. Sistemul nostru prezice corect că persoana detectată este aceeași ca în prima imagine. Mai mult, în cea de-a treia imagine se poate observa că sistemul nostru păstrează același număr de identificare pentru persoană, chiar și după ce a fost blocat de stâlpul de iluminat stradal și după ce camera și-a schimbat unghiul de vizualizare. Putem observa că între a doua și a treia imagine identificatorul asociat de metoda standard de urmărire s-a schimbat din p7 în p9.

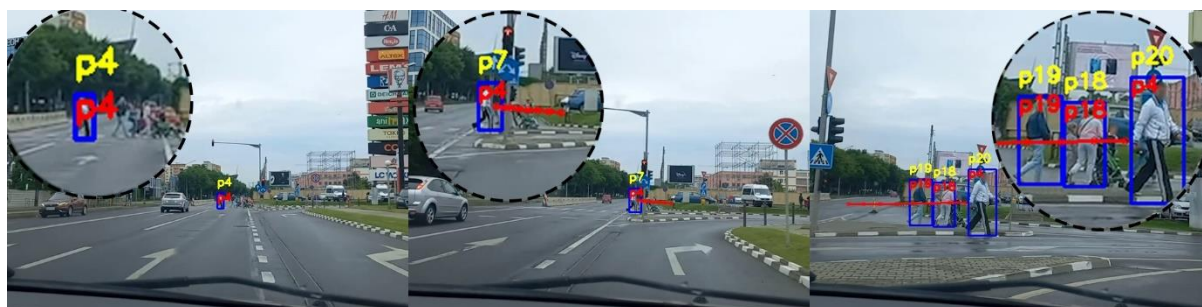


Figura 7. Reidentificarea persoanelor într-un scenariu în care mașina se mișcă la dreapta și pietonii traversează strada.

Figura 8 este un exemplu de situație în care vehiculul se deplasează spre stânga și pietonii merg în două direcții diferite pe trotuar. În imaginea inițială îi putem observa pe cei doi pietoni p4 și p2 mergând în stânga și respectiv în dreapta imaginii. Pe măsură ce mașina se apropie de pietoni și face stânga, pietonul p2 este reidentificat în mod greșit în a doua imagine de tehnica standard de urmărire a persoanelor ca fiind p11, în timp ce sistemul nostru atribuie corect același identificador. În a treia imagine, pietonul p4 este urmărit corect de sistemul nostru și incorrect de sistemul standard. Limitările în ceea ce privește scalarea și unghiurile camerei care apar în cadrul tehnicii standard de urmărire sunt rezolvate de metoda propusă de noi. Ceea ce putem observa în acest exemplu este modul în care modulul de predicție de traiectorie se adaptează în funcție de mișcarea vehiculului. În prima imagine, traiectoria pietonului p4 indică faptul că pietonul se va deplasa în stânga imaginii. Traiectoria din prima imagine este generată pe baza observațiilor făcute în timp ce vehiculul se deplasa drept. Pe măsură ce vehiculul începe să vireze la stânga, în a doua imagine modulul de predicție a traiectoriei combină mișcarea camerei cu mișcarea persoanei și estimează că poziția în imagine a persoanei va rămâne aceeași. În a treia imagine putem observa că traiectoria prezisă implică că poziția pietonului va fi deplasată la dreapta imaginii, deoarece viteza vehiculului este mai mare decât viteza pietonului. Traiectoria prezisă estimează corect pozițiile viitoare ale persoanei din imagine, luând în considerare mișcarea atât a vehiculului, cât și a pietonului.

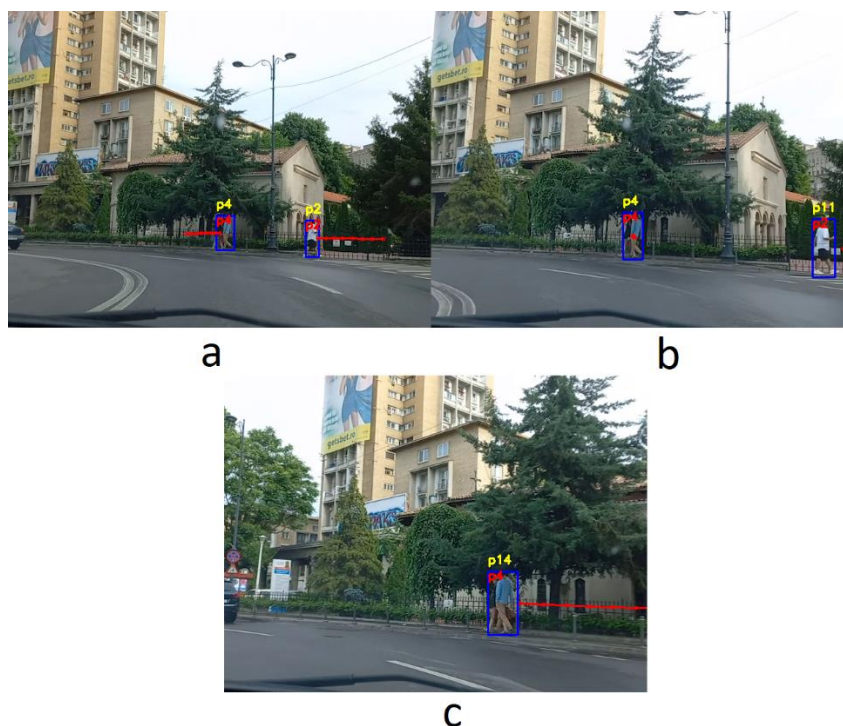


Figura 8. Reidentificarea persoanelor într-un scenariu în care mașina virează la stânga și pietonii se deplasează pe trotuar.

Figura 9 prezintă rezultatele unui experiment în care mașina se mișcă mai lent și urmează o direcție dreaptă. Caracteristica acestui experiment este că vehiculul se mișcă încet până când pietonul traversează strada, iar apoi începe să accelereze. Deoarece pietonii p76 și p77 sunt relativ aproape de camera și nu există o mișcare substanțială a vehiculului, tehnica standard de urmărire reușește reidentificarea corectă a pietonilor. Când pietonii vin în apropiere unul de altul și mașina începe să se miște puțin mai repede, ceea ce se întâmplă în cazul celei de-a treia imagini, ambelor persoane li se atribuie numere de identificare diferite prin metoda standard de urmărire. Sistemul propus reidentifică corect ambii pietoni, păstrând aceleași numere de identificare, fără schimbarea identității acestora.



Figura 9. Reidentificarea persoanelor într-un scenariu în care mașina se mișcă drept și pietonii se află în fața mașinii.

Pentru a avea, de asemenea, o fază de testare legitimă bazată pe valorile generate de metoda standard de urmărire, am calculat valorile MOTA [12], IDF1 [13] și HOTA [14] pentru a demonstra îmbunătățirea sistemului nostru. Am raportat valorile pe setul de date Caltech-Pedestrians [17], folosind aceeași tehnică de calcul [11] folosită în secțiunea anterioară. Setul de date Caltech-Pedestrians este o colecție de date care include 11 subseturi de videoclipuri înregistrate de o mașină în mișcare într-un mediu urban la o rată a imaginilor de 30 de cadre pe secundă. Pentru o mai bună vizualizare a modului în care arată setul de date am inclus imaginea din Figura 10.



Figura 10. Modulul de predicție a traiectoriei aplicat pe o imagine din setul de date Caltech-Pedestrians. Cu albastru este traiectoria observată, cu roșu traiectoria prezisă, iar cu verde traiectoria reală urmată de persoană.

Deoarece predicția traiectoriei necesită 8 poziții observate pentru a prezice traiectoria viitoare, o rată a imaginilor mare nu este potrivită pentru sistemul nostru. A observa mișcarea unei persoane timp de mai puțin de 300 de milisecunde nu este suficient de informativ pentru a face o predicție fiabilă. În consecință, am aplicat un pas de preprocesare a setului pentru a simula o rată mai mică a videoclipurilor din setul de date. Am folosit detecțiile persoanelor existente în setul de date, astfel încât valorile obținute ale indicilor sunt strict determinate de numerele de identificare atribuite detecțiilor. Valorile obținute sunt raportate în tabelul de mai jos. Modelul folosește modulul de predicție a traiectoriei oamenilor antrenat pe subseturile de la s00 la s05

și este testat pe subseturile de la s06 la s10. Setul de date a fost preprocesat pentru a simula o rată a imaginilor de 6 cadre pe secundă, pentru o predicție mai informată a traiectoriei.

Tabela 4. Metricile pentru problema urmăririi persoanelor pentru sistemul de reidentificare bazat pe predicția traiectoriei pentru setul de date Caltech-Pedestrians [17].

Metrică	Metodă standard de urmărire	Sistemul propus
MOTA (%)	89,22	94,92
IDF1 (%)	78,36	90,10
HOTA (%)	78,96	89,15

Valorile raportate demonstrează o îmbunătățire semnificativă a sistemului nostru în comparație cu tehnica de urmărire existentă. Sistemul propus obține valori mai mari pentru toate cele trei metrici. Acest rezultat se datorează naturii complexe a problemei de reidentificare și urmărire a persoanelor în contextul conducerii autonome. În cazul conducerii autonome, avem de integrat factori mai complecși care pot afecta comportamentul sistemului decât în cazul roboticii sociale. Pentru a putea avea o performanță robustă și fiabilă, sistemul trebuie să integreze schimbările de viteză și direcția camerei, alături de variațiile de aspect, cum ar fi dimensiunea și ocluziile. În timp ce aceste elemente sunt mai greu de prezis prin tehnica standard de urmărire, sistemul nostru propus rezolvă un număr considerabil de limitări și demonstrează o performanță satisfăcătoare. Integrarea unei predicții a traiectoriei oamenilor în sistemul de reidentificare produce un sistem mai informat și mai stabil.

La fel ca în cazul secțiunii anterioare, pentru a avea o înțelegere obiectivă a erorilor raportate pentru sistemul de reidentificare și urmărire a persoanelor, am calculat și erorile obținute de componenta de predicție a traiectoriei. Tabela 5 prezintă un raport detaliat asupra performanței sistemului în context unei mașini autonome. Am calculat ADE, eroarea medie a pozițiilor și FDE, eroarea poziției finale, pentru fiecare subset inclus în setul de date Caltech-Pedestrians, așa cum am procedat în secțiunea anterioară. Valorile erorilor sunt calculate în pixeli, pentru traiectorii observate și prezise de lungime 8. Erorile raportate în tabel dovedesc că nu există o diferență semnificativă între subseturi. Singurul FDE care are o valoare puțin mai mare în comparație cu restul subseturilor este obținut pentru *set00*, dar metrica ADE pentru același subset este comparabilă cu celelalte valori.

Evaluarea sistemului a obținut erori mici pentru componenta de predicție a traiectoriei, cu cel mai mare FDE de 21,35 pixeli și cel mai mic FDE de 7,20 pixeli. Ultima linie din tabel prezintă valorile obținute pentru metricile ADE și FDE folosind împărțirea între seturile de antrenare și testare recomandată de autorii setului Caltech-Pedestrians. Subsetul de antrenare recomandat este compus din subseturile de la *set00* la *set05*, iar subsetul de testare este format din subseturile de la *set06* la *set10*. Valorile nu se abat de la erorile medii calculate, demonstrând că funcționarea sistemului este robustă și fiabilă.

Tabela 5. Rezultatele metricilor ADE (Average Displacement Error) și FDE (Final Displacement Error) pentru fiecare subset din setul de date Caltech-Pedestrians [17].

Subseturi Caltech-Pedestrians	Erori obținute	
	ADE	FDE
set00	11.05	21.35
set01	7.19	13.17
set02	6.70	11.73
set03	9.68	17.79
set04	5.61	9.74
set05	7.40	13.85
set06	4.20	7.17
set07	7.03	12.82
set08	4.42	7.20
set09	7.01	13.52
set10	6.80	13.10
Eroare medie	7,00	12,85
Împărțire train-test recomandată	6.29	11.78

4. Predicția traiectoriilor vehiculelor înconjurătoare

Pornind de la cea de a 2-a soluție propusă în raportul din Etapa II, am dezvoltat de asemenea un modul de predicție a traiectoriilor vehiculelor înconjurătoare a vehiculului în mișcare autonomă. Pentru acest task am utilizat trei versiuni modificate ale unei arhitecturi PredNet, o arhitectura LSTM convoluțională pentru realizarea de predicții din video. La un nivel înalt, PredNet poate fi văzută ca având mai multe straturi convoluționale recurente, cu ieșiri care trec printr-o activare ReLU și un max-pooling cu stride 2.

Modificările pe care le-am adus arhitecturii de bază PredNet sunt descrise în cele ce urmează. Versiunea standard a arhitecturii utilizează 4 straturi cu 3x3 convoluții. Am modificat arhitectura de bază prin înlocuirea convoluțiilor 3x3 cu convoluții 5x5 obținând P_5_5, fără a adăuga straturi suplimentare (prima versiune), apoi am utilizat un model cu 6 straturi, având astfel 2 straturi suplimentare de convoluție de 3x3 și am înlocuit activarea ReLU cu una de tip PReLU care, în loc de a considera 0 valorile negative, învață un parametru care este multiplicat cu valoarea răspunsului (a 2-a versiune) conform următoarei ecuații

$$f(x) = \begin{cases} x, & x \geq 0 \\ a * x, & \text{otherwise} \end{cases}$$

În cea de a 3-a versiune, P_full, am utilizat în continuare un model cu 6 straturi și o activare de tip PReLU, descrisă mai sus.

Din punct de vedere al fluxului de lucru, cadrele de intrare din secvența video sunt inițial ajustate la dimensiunea necesară modelului apoi sunt introduse în rețea. Ieșirea rețelei este trecută printr-o rețea YOLO v4 pentru a detecta obiectele din cadru, informație care împreună cu o rețea FCN (Fully Convolutional Networks) reușește să facă predicții referitoare la mișcarea mașinilor înconjurătoare. Arhitectura generală a modului este prezentată în Figura 11.

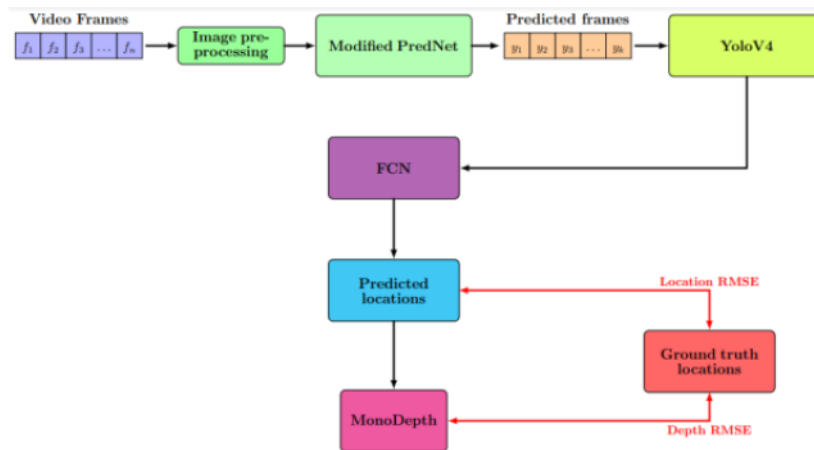


Figura 11. Fluxul de date și arhitectura modului de predicție a direcției de mișcare

Fiecare din cele 3 versiuni ale sistemului descris a fost antrenată și testată pe setul de date KITTI, rezultatele predicțiilor pentru fiecare din cele 3 versiuni fiind prezentate în Figura 12.

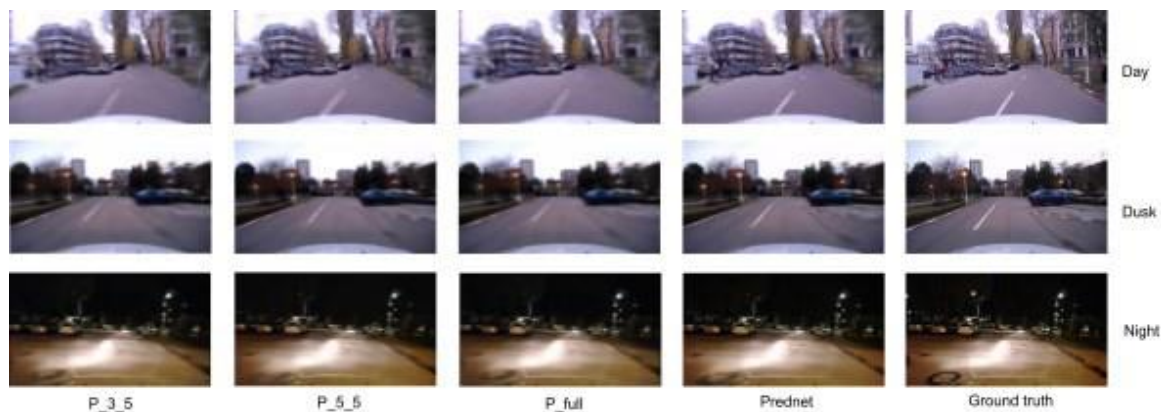


Figura 12. Rezultatul predicției traiectoriei vehiculelor înconjurătoare pentru cele 3 versiuni de PredNet

Considerăm că soluția propusă este de interes și vom investiga în continuare utilizarea ei atât pentru predicția mișcării autovehiculelor cât și a pietonilor.

5. Concluzii

Problema reidentificării și urmăririi persoanelor este o problemă foarte complexă, cu o aplicabilitate mare în diverse domenii. Dispozitivele autonome se bazează pe rezultate precise pentru o performanță sigură și stabilă. În special, roboții sociali și mașinile care se conduc singure depind de performanța acestui sistem pentru interacțiuni semnificative om-robot și navigare sigură.

Problema reidentificării și urmăririi persoanelor constă în capacitatea unui sistem de a recunoaște corect aceleași persoane pe o perioadă lungă de timp într-o succesiune de imagini. Pentru rezultate mai bune, sistemele de urmărire încorporează metode de predicție a traiectoriei pentru a estima mișcarea oamenilor. Predicția traiectoriei implică modelarea mișcării unui individ și prezicerea pașilor următori pe baza observațiilor anterioare. Modelul din spatele mișcării unei persoane trebuie să combine trei informații pentru a obține rezultate precise: starea interioară a persoanei, influența socială și înțelegerea scenei.

În acest raport am introdus un sistem de reidentificare și urmărire a oamenilor în timp real bazat pe predicția traiectoriei. Am proiectat o arhitectură modulară pentru partea de reidentificare, compusă dintr-un detector de obiecte, o rețea de segmentare, un modul de urmărire de persoane și un modul de predicție a traiectoriei. Pentru modulul de predicție a traiectoriei am folosit o metodă care generează o traiectorie în imagine bazată pe date vizuale preluate din mediu și coordonatele persoanei în imagine. Informația vizuală semantică asociată scenei este extrasă folosind o rețea de segmentare semantică. Abordarea noastră îmbunătățește capacitatea de reidentificare în comparație cu alte abordări similare, oferind predicții mai fiabile.

Am efectuat o analiză detaliată a rezultatelor obținute de abordarea noastră pentru sistemul de reidentificare și urmărire a persoanelor. Am evaluat performanța sistemului pe seturi de date relevante existente, cu imagini achiziționate din punctul de vedere al unei platforme autonome și pe date preluate de noi. Ne-am validat abordarea în două contexte diferite: robotică socială și conducere autonomă. Ne-am integrat munca în cadrul platformei pentru roboți AMIRO pentru a valida abordarea noastră prin îmbunătățirea comportamentului general al robotului Pepper. Cele două scenarii, roboți sociali și mașini autonome, pe care le-am folosit pentru a verifica abordarea noastră dovedesc aplicabilitatea acestora într-o varietate de aplicații. Am prezentat mai multe scenarii din experimentele noastre pentru a arăta comportamentul sistemului propus. Metoda prezentată obține reidentificări mai fiabile ale persoanelor detectate, în situații mai complexe, rezolvând probleme precum ocluziile, lipsa datelor sau mișcarea camerei.

În viitor, vom încerca să îmbunătățim sistemul pe care l-am propus prin includerea influenței tuturor obiectelor aflate în mișcare în imagini, cum ar fi mașini sau animale, asupra traiectoriei oamenilor. În continuare, intenționăm să extindem gama informațiilor semantice transmise sistemului, pentru a include date mai relevante, pentru reidentificări mai precise.

Bibliografie

- [1] Feil-Seifer, D.; Mataric, M.J. Defining socially assistive robotics. *Rehabilitation Robotics*, 2005. ICORR 2005. 9th International Conference on. IEEE, 2005, pp. 465–468.
- [2] Ghiță, A.Ș.; Gavril, A.F.; Nan, M.; Hoteit, B.; Awada, I.A.; Sorici, A.; Mocanu, I.G.; Florea, A.M. The AMIRO Social Robotics Framework: Deployment and Evaluation on the Pepper Robot. *Sensors* 2020, 20. doi:10.3390/s20247271.
- [3] <https://www.softbankrobotics.com/emea/en/pepper/>
- [4] <https://www.ros.org/>
- [5] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788. doi:10.1109/CVPR.2016.91.
- [6] Wojke, N.; Bewley, A.; Paulus, D. Simple Online and Realtime Tracking with a Deep Association Metric. 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017, pp. 3645–3649. doi:10.1109/ICIP.2017.8296962.
- [7] Wojke, N.; Bewley, A. Deep Cosine Metric Learning for Person Re-identification. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018, pp. 748–756. doi:10.1109/WACV.2018.00087.
- [8] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
- [9] Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. *CVPR*, 2017.
- [10] Milan, A.; Leal-Taixé, L.; Reid, I.; Roth, S.; Schindler, K. MOT16: A Benchmark for Multi-Object Tracking. arXiv:1603.00831 [cs] 2016. arXiv: 1603.00831.
- [11] Jonathon Luiten, A.H. TrackEval. <https://github.com/JonathonLuiten/TrackEval>, 2020.
- [12] Bernardin, K.; Stiefelhagen, R. Evaluating multiple object tracking performance: the clear mot metrics. *EU-RASIP Journal on Image and Video Processing* 2008, 2008, 1–10.
- [13] Ristani, E.; Solera, F.; Zou, R.; Cucchiara, R.; Tomasi, C. Performance measures and a data set for multi-target, multi-camera tracking. *European conference on computer vision*. Springer, 2016, pp. 17–35.

- [14] Luiten, J.; Osep, A.; Dendorfer, P.; Torr, P.; Geiger, A.; Leal-Taixé, L.; Leibe, B. HOTA: A Higher Order Metric for Evaluating Multi-Object Tracking. *International Journal of Computer Vision* 2020, pp. 1–31.
- [15] Yang, F.; Choi, W.; Lin, Y. Exploit All the Layers: Fast and Accurate CNN Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2129–2137. doi:10.1109/CVPR.2016.234.
- [16] Martín-Martín, R.; Rezaeifighi, H.; Shenoi, A.; Patel, M.; Gwak, J.; Dass, N.; Federman, A.; Goebel, P.; Savarese, S. JRDB: A Dataset and Benchmark for Visual Perception for Navigation in Human Environments. *CoRR* 2019, abs/1910.11792.
- [17] Dollár, P.; Wojek, C.; Schiele, B.; Perona, P. Pedestrian detection: A benchmark. *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 304–311.

Prof. dr. ing. Adina Magda Florea