

Pregled osnovnih tehnika kompjuterske vizije u autonomnim vozilima

Zlatko Veličković, Jelena Kocić, Nenad Jovičić

Elektrotehnički fakultet
Univerzitet u Beogradu
Beograd, Srbija

Sažetak—Cilj ovog rada je pregled osnovnih tehnika kompjuterske vizije u autonomnim vozilima. Istraživanje na temu algoritama mašinskog učenja i robotike sa primenom u domenu autonomnih vozila predstavljaju izuzetno aktuelnu temu istraživanja u nauci i industriji. Iz ovog razloga, od velike važnosti je da se ima jasan uvid u trenutna istraživanja ovog brzo rastućeg domena. Ovaj rad daje pregled osnovnih tehnika kompjuterske vizije u autonomnim vozilima. U radu su diskutovane ključne reference u prepoznavanju objekata, semantičkoj segmentaciji, estimaciji pokreta i pozicije, praćenju objekata i *end-to-end* učenju. Pored toga, opisane su najznačajnije baze podataka za treniranje algoritama mašinskog učenja.

Ključne riječi—kompjuterska vizija; mašinsko učenje; autonomna vozila; detekcija objekata; praćenje objekata; baze podataka;

I. UVOD

U poslednjoj deceniji došlo je do ekspanzije u oblasti obrade slike i kompjuterske vizije. Razvoj i usavršavanje savremenih elektro-optičkih sistema kao što su kamere visoke rezolucije, stereoskopski vid, 3D vizija i sl. su omogućili da se u oblasti obrade slike i kompjuterske vizije razvijaju novi algoritmi koji koriste prednosti napredka u hardverskim performansama sistema za viziju. Povećavanje mogućnosti kompjuterske vizije, kao i boljih hardverskih performansi sistema, dovodi do ponovnog korišćenja algoritama mašinskog učenja kao što su neuralne mreže, i u poslednje vreme veoma popularna tehnika višeslojnih neuralnih mreža, engl. *deep learning*.

Tvorci neuralnih mreža su Warren McCulloch i Walter Pitts, koji su 1943. godine objavili rad [1]. Međutim, do ekspanzije u korišćenju ovih mreža dolazi tek sredinom osamdesetih godina prošlog veka, sa razvojem paralelno distribuiranih procesa i poboljšanjima u hardverskim performansama sistema za obradu i procesiranje signala. Ovo je omogućilo treniranje velikih baza podataka, što je od velikog značaja za tačnost rešenja. Pedesetih godina prošlog veka dolazi se do prvih značajnijih pokušaja u razvoju autonomnih vozila. Prva potpuno autonomna vozila razvijena su 1984. na Univerzitetu Carnegie Mellon, [2]-[3], i 1987. od strane Mercedes-Benz-a i Univerziteta u Minhenu. Od tada, brojne velike kompanije i istraživačke organizacije su razvile prototipove autonomnih vozila, uključujući Mercedes-Benz,

General Motors, Continental Automotive Systems, Autoliv Inc., Bosch, Nissan, Toyota, Audi, Volvo, Vislab University of Parma, Oxford University i Google. Aktuelni domeni na kojima se trenutno radi su razvoj algoritama za mašinsko učenje, koji uključuju kompjutersku viziju, i razvoj kontrolne teorije i robotike, koja podrazumeva procesiranje informacija dobijenih sa senzora i donošenje odluka o ponašanju vozila na osnovu ovih informacija.

Cilj ovog rada je predstavljanje najznačajnijih algoritama iz oblasti kompjuterske vizije u autonomnim vozilima koji su razvijani u poslednje dve decenije. Predstavljene su postojeće baze podataka koje služe za obučavanje i testiranje algoritama, KITTI, [4], ISPRS, MOT i Cityscapes. Takođe, navedeno je nekoliko preglednih radova, [5]-[8]. Doprinos ovog rada je u povezivanju oblasti kao što su kompjuterska vizija, mašinsko učenje i robotika u domenu autonomnih vozila i predstavljanju akutelnih algoritama, rešenja, i otvorenih pitanja.

II. BAZE PODATAKA

Baze podataka igraju ključnu ulogu u razvoju mnogih oblasti istraživanja tako što obezbeđuju mogućnost obučavanja i testiranja raznih algoritama u toku njihovog stvaranja. Poznate baze podataka u oblasti kompjuterske vizije su KITTI, ISPRS, MOT i Cityscapes, i predstavljene su u radovima [4], [9]-[11]. Neke od pomenutih baza podataka postoje i na javno dostupnim internet stranicama. U oblasti autonomnih vozila najkompletnije i najupotrebljivije baze podataka su KITTI [4], i Cityscapes baza podataka [11], nad čijim podacima se mogu testirati algoritmi rekonstrukcije, estimacije pokreta i pozicije, i prepoznavanja objekata.

Baze podataka se grubo mogu podeliti na baze podataka sa realnim i sa sintetičkim podacima. Baze podataka sa realnim podacima su ključne za ponašanje sistema koji su namenjeni korišćenju u realnom svetu, na ulici, itd. Ove baze između ostalog, testiraju algoritme koji rešavaju probleme kao što su direktno upereno svetlo, refleksije sa određenih površina, magla ili kiša. Baze podataka sa sintetičkim podacima, često nerealne, predstavljaju prečišćene realne podatke, podatke sa kojih je uklonjen šum i koji su pojednostavljeni za primenu. Na primer, u radu [12] opisana je nova dizajnirana i snimljena baza podataka koja sadrži stereo i podatke protoka, gde je izdvojen početni referentni podskup koji sadrži 28504 stereo parova. Takođe, u ovom radu je data odličan pregled oblasti za

aktulene baze podataka. Neki od ključnih problema sa kojima se autori baza podataka sreću su konstrukcija složenih scena, pešaci u raznim pozicijama tela, dinamičnost scena, simulacije raznih saobraćajnih nezgoda, osvetljenje, promene vremena i godišnjih doba, dinamički objekti, dugoročne promene poludinatičkih objekata, itd. Takođe, mali je broj generisanih sintetičkih baza podataka, a njihova prednost je kompjutersko generisanje velikog broja instanci različitih objekata u različitim okolnostima, pa se često mogu koristiti i za treniranje i testiranje algoritama u simuliranim okruženjima.

III. DETEKCIJA OBJEKATA

Pouzdana detekcija objekata je ključna osobina za realizaciju autonomnih vozila. Kako automobil deli put sa mnogo drugih učesnika u saobraćaju, posebno u urbanim područjima, svesnost o ostalim učesnicima i preprekama je neophodna zarad izbegavanja potencijalnih saobraćajnih nezgoda. Detekcija i razlikovanje u urbanim područjima je kompleksno zbog mnoštva različitih objekata i sličnosti u percepciji različitih objekata. Senzori koji se koriste za detekciju objekata su senzori koji rade u vidljivom delu svetlosnog spektra, engl. *Visible Spectrum* (VS) za detekcije po dnevnom svetlu, i termalni senzori u infracrvenom spektru, engl. *Thermal Infrared* (TIR) za detekcije noću, omogućavaju razlikovanje toplih objekata (pešaka, životinja) od hladnih (put, objekti, vegetacija). Najbolje rezultate daje fuzija slike sa ova dva tipa senzora [13].

Postupak detekcije objekata se najčešće sastoji od sledećih koraka:

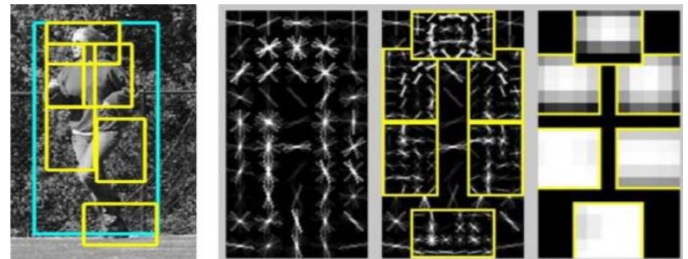
- a) *preprocesiranje*,
- b) *ekstrakcija regiona od interesa (ROI)*,
- c) *klasifikacija objekata i*
- d) *verifikacija*.

Ekstrakcija regiona od interesa se obično vrši korišćenjem klizećih prozora, engl. *sliding window*, koji se pomera po slici. Pretpostavljanjem veličine i pozicije potencijalnih prozora se smanjuje broj iteracija. Takođe, karakteristike slike se mogu iskoristiti za usmeravanje potrage na određene regione. U radu [14] autori filtriraju potencijalne pešake korišćenjem veličine, oblika i vertikalnih simetrija ljudskog oblika. Pored toga, informacije o udaljenosti dobijene iz stereo vizije koriste prilikom ekstrakcije regiona od interesa. Klasifikacija se najbolje vrši korišćenjem složenih konvolucionih neuralnih mreža u cilju ostvarivanja velike tačnosti i brzine donošenja odluke prilikom klasifikacije. U radu [15] autori su uveli konvolucione neuralne mreže za rešavanje problema detekcije pešaka, i to korišćenjem obučavanja bez supervizije za „grubu“ klasifikaciju obeležja i *end-to-end* učenje sa supervizijom za „finu“ klasifikaciju obeležja. U današnje vreme, sve veći broj pristupa detekcije se vrši *end-to-end* iz velikih baza podataka. Jedan od pristupa u klasifikaciji jeste podela objekata od interesa na karakteristične regione i jedinstvena reprezentacija tih regiona, kao što je prikazano na Sl. 1, [16].

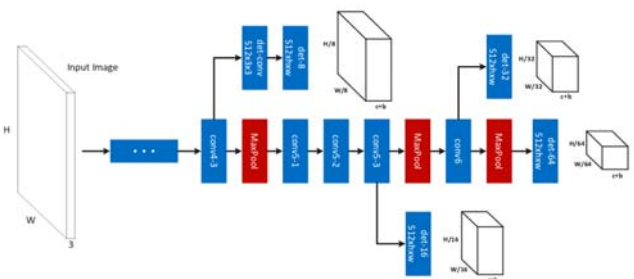
A. 2D detekcija objekata

Poslednjih godina projektovano je više značajnih algoritama za 2D detekciju objekata. U radu [17] je opisan algoritam koji je testiran na KITTI bazi podataka, gde su

uporedno prikazane karakteristike algoritama: tačnost u jednostavnim, srednje kompleksnim i kompleksnim situacijama, kao i vreme izvršavanja. Poređenja su vršena za detekciju tri ključne vrste objekata: automobili, pešaci i biciklisti. Ovo poređenje pokazuje da su isti algoritmi postigli mnogo veću tačnost pri detekciji automobila, jer se njihov oblik ne menja u jedinici vremena, za razliku od pešaka i biciklista, čija je tačnost detekcije trenutno veoma aktuelna tema. Jedno rešenje za brzo detekovanje objekata korišćenjem višeslojnih konvolucionalnih neuralnih mreža je opisano u radu [18], Sl. 2.



Slika 1. Primer detekcije osobe sa podelom na karakteristične regione [16]



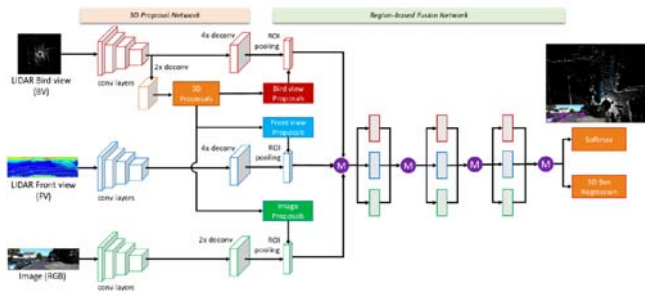
Slika 2. Detekcija više izlaznih slojeva koji odgovaraju objektima različite veličine [18]

B. 3D detekcija objekata

Jedna od metoda za trodimenzionalnu detekciju objekata je detekcija 3D objekata iz 2D slika. Trodimenzionalna reprezentacija klasa objekata može rekonstruisati znatno veći broj detalja nego dvodimenzionalni pravougaonici koji okružuju objekte na 2D slikama. U radu [19] koriste dostupne 3D CAD modele za rekonstrukciju iz 2D u 3D domen tako što iz modela dobijaju „grube“ 3D *wire-frame* modele.

Za 3D detekciju objekata iz 3D oblaka se koristi LiDAR engl. *Laser Imaging, Detecion and Ranging*, laserski senzor koji direktno obezbeđuje precizne 3D informacije što pojednostavljuje ekstrakciju objekata i daje realne informacije o obliku 3D objekata. Problem predstavlja mala prostorna rezolucija laserskih senzora, pa se oni ne mogu samostalno koristiti, već se koriste u kombinaciji sa nekim drugim sensorima, obično radarima ili kamerama. Takođe, algoritmi dizajnirani u poslednjih nekoliko godina su testirani i rezultati su takvi da je detekcija automobila slične tačnosti kao u algoritmima 2D detekcije objekata, ali je prednost to što su brži. Na Sl. 3 je prikazana neuralna mreža koja procesira informacije sa LiDAR-a predložena u radu [20]. Autori kombinuju podatke sa LiDAR-a sa RGB slikama za detekciju objekata. Proređen oblak tačaka je kodovan kompaktnom

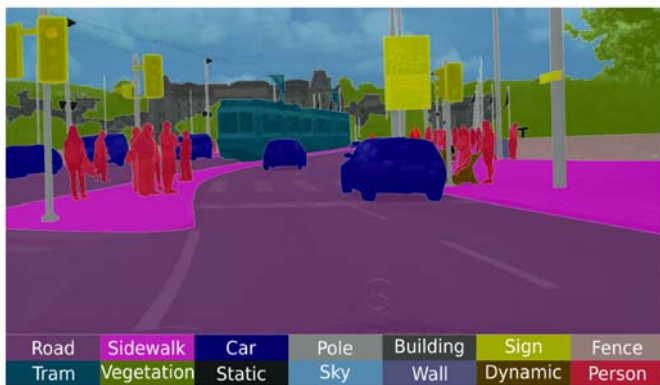
reprezentacijom i na osnovu reprezentacije iz „ptičje“ perspektive se generišu 3D kandidati. Na kraju se kombinuju obeležja iz više perspektiva u mreži.



Slika 3. Prednji pogled LiDAR-a u funkciji RGB slike kao ulaz u višeslojnu neuralnu mrežu [20]

IV. SEMANTIČKA SEGMENTACIJA

Semantička segmentacija je jedna od fundamentalnih oblasti kompjuterske vizije. Cilj je dodela svakog piksela slike grupi određene predefinisane kategorije. U domenu autonomne vožnje, segmentacija se radi na slikama okruženja na ulicama, na automobilima, pešacima, putu i pruža razumevanje okruženja. Na Sl. 4 je prikazana jedna scena na kojoj je izvršena semantička segmentacija, [11].

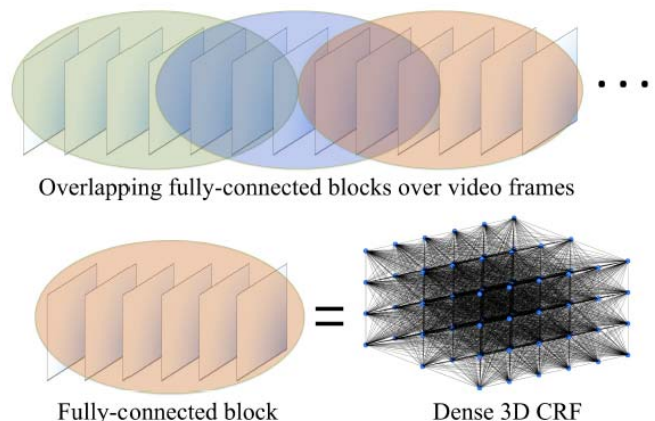


Slika 4. Semantička segmentacija scene iz Cityscapes baze podataka [11]

Metode koje su do sada uvedene uzimaju u obzir svaku klasu objekata nezavisno, a često se klase objekata međusobno korelišu odgovarajućim faktorom (npr. automobil će se verovatnije sresti na ulici nego u kancelariji). U poslednje vreme za semantičku segmentaciju se koriste i neuralne mreže, [21]. Dok moderne konvolucione neuralne mreže za klasifikaciju kombinuju informacije uzastopnim grupisanjem slojeva koji smanjuju rezoluciju, semantička segmentacija zahteva rezonovanje na osnovu konteksta i gustu visokorezolucionu predikciju. U radu [22] predlaže se konstrukcija Laplasove piramide koja je bazirana na potpuno konvolucionoj neuralnoj mreži. Ipak, višeslojne neuralne mreže dovode do saturacije i degradacije preciznosti. Pokušaj rešavanja ovog problema predložen je u radu [23] uvođenjem radnog okvira ResNet. ResNet se ponaša kao linearna kombinacija plitkih mreža, pa se smanjuje vremenska i resursna složenost. Specifičnost ove neuralne mreže je njena

dubina, od čak 152 lejera i uvođenje novih rezidualnih blokova. Rezidualni blokovi adresiraju problem treniranja veoma duboke arhitekture uvođenjem preskočenih konekcija tako da slojevi mogu da kopiraju svoj ulaz na sledeći sloj. Ideja ovog pristupa je da se osigura da sledeći sloj vidi nešto novo i različito od onoga što je njegov ulaz već enkodirao. Ovim je omogućeno da sledeći sloj dobija dva ulazna podatka, izlaz prethodnog sloja i nepromenjeni ulaz prethodnog sloja.

Cilj semantičke segmentacije instanci je istovremena detekcija, segmentacija i klasifikacija svakog objekta u slici. Takođe obezbeđuje informacije o poziciji, semantici, obliku i broju objekata, i ima veliku primenu u autonomnoj vožnji. Svaka instanca mora da bude precizno odvojena i zabeležena, za razliku od klasične semantičke segmentacije gde se grupe objekata iste kategorije mogu grupno obeležiti. Metod propagacije labela u poređenju sa obeležavanjem individualnih slika sa snimka, koji može imati veliku učestanost frejma, koristi prednost sličnosti uzastopnih frejmova tako što se, ukoliko se npr. ustanovi sličnost u boji, obeležje samo prepíše iz prethodnog frejma. Kod semantičke segmentacije sa više frejmova kratkotrajna korelacija između uzastopnih frejmova se može iskoristiti kako bi se poboljšala preciznost, efikasnost i robusnost segmentacije. U radu [24] se uvode grafički modeli koji procesiraju video sekvence kako bi se ustanovila privremena konzistentnost između frejmova. Na primer, 3D rekonstrukcija trenutno dobro funkcioniše u statičkim scenama, dok je još uvek otvoren problem kod dinamičkih scena, pa se veruje da se ovaj princip može primeniti i na taj problem. Mera udaljenosti između projektovanih objekata ne oslikava pravilno odgovarajuće prostorne ekvivalente. Takođe, kretanje kamere otežava asocijaciju objekata susjednih frejmova. Tako, euklidska razdaljina nije dobar indikator za asocijaciju objekata na više uzastopnih frejmova. Kako bi rešili ovaj problem, autori u radu [25] uvode metod za optimizaciju prostora obeležja, Sl. 5. Prostor obeležja je optimizovan, a rezultatno mapiranje se koristi za postizanje daleko dosežne karakterizacije snimanog prostora. Na Sl. 6 prikazana je vremenska struktura modela obeležja. Video je prekriven blokovima koji se preklapaju. Gust CRF je definisan preko svakog bloka u okviru kojih se vrši optimizacija prostornih karakteristika. Struktuuisano predviđanje se izvodi nad višestrukim blokovima.



Slika 5. Vremenska struktura modela obeležja [25]

Semantičko obeležavanje objekata je tema koja se veoma istražuje, međutim, većina algoritama radi u 2D domenu slika. 2D slike imaju manjak informacija u odnosu na 3D oblik i srazmerno skaliranje objekata koji treba da se obeleže. U radu [26] uvodi se pristup gde se od 2D slika kreiraju 3D scene iz ptičije perspektive, i nakon toga se obeležavaju objekti, sada u sve tri dimenzije. U radu [27] uvodi se brza semantička segmentacija za 3D skupove tačaka, tako da se u roku od nekoliko minuta procesira više miliona tačaka. Zanimljiv je i 3D pristup u dizajniranju konvolucionih neuralnih mreža (CNN), [28], gde se uvodi frejmwork za obeležavanje tačaka 3D oblaka, korišćenjem 3D konvolucionih neuralnih mreža, ali sa nedostatkom dobre rezolucije zbog memorijskih ograničenja računara. Kako bi njihov problem rešili, u radu [29] uvode se Oct-Nets (oktalne mreže), takođe 3D CNN, koje hijerarhijskim particionisanjem dele 3D prostor na oktante, čime se procesiranje pojednostavljuje. Gotovo svi algoritmi za semantičku segmentaciju puta su bazirani na mašinskom učenju gde se parametri modela estimiraju iz velikih setova podataka, [30].

V. ESTIMACIJA POKRETA I POZICIJE

Optički tok je definisan kao 2D kretanje osvetljenja između dve slike i pruža važne informacije o sceni, služi kao osnova za kretanje ego vozila i praćenje objekata. Glavni problem optičkog toka je da na promene intenziteta piksela, pored kretanja, mogu uticati i promene osvetljenja, refleksije i transparentnost objekata. Robusni algoritmi treba da dobro rešavaju situacije sa promenama intenziteta koje nisu izazvane kretanjem objekata. U radu [31] se ovi algoritmi sistematično analiziraju i nudi rešenje za problem osvetljenja testiran na KITTI bazi. Trenutno nerešeni problemi su upravo uzrokovani refleksijom, transparentnošću, nejasnoćama i zaklanjanjem objekata od interesa.

A. Istovremena lokalizacija i mapiranje (SLAM)

Istovremena lokalizacija i mapiranje, engl. *Simultaneous Localization and Mapping* (SLAM) je vrlo bitan pristup u planiranju puta i navigacije autonomnog vozila. Mapiranje bez lokalizacije nije izvodljivo, pogotovo što se i jedno i drugo moraju simultano izvršavati u realnom vremenu. Mapiranje i metrika su neophodni u rešavanju raznih zadataka, npr. metričke mape predstavljaju uslov za preciznu lokalizaciju dok semantičke mape obezbeđuju informacije o npr. oblastima parkiranja gde je moguće izvršiti autonomno parkiranje. Ipak, takvo mapiranje se može izvršiti i van realnog vremena i

kasnije uključiti u autonomno vozilo kao ugrađen element. U radu [32] je predstavljen engl. *Google Street View* projekat koji za lokalizaciju koristi estimiranje Kalmanovim filtrom sa GPS-om kao izvorom merenja.

B. Lokalizacija

Lokalizacija je opsežno razvijena oblast u robotici i kompjuterskoj viziji, i pokriva širok opseg tehnika, počev od lokalizacije lobota u unutrašnjem prostoru pomoću merenja sa senzora, pa sve do lokalizacije u spoljašnjem okruženju. Iz perspektive autonomne vožnje, zadatak je da se precizno odredi lokacija samog autonomnog vozila na mapi. Lokalizacija se može obaviti korišćenjem GPS senzora ili na osnovu vizuelnih informacija baziranih na slikama. Samo GPS senzori obezbeđuju preciznost oko 5 m, ali se kombinacijom informacija sa različitim senzora može se dobiti centimetarska preciznost. U realnom uslovima usled postojanja vegetacije, refleksije objekata i sl. lokalizacija se ne može osloniti samo na satelitske sisteme. U poslednje vreme je posebno interesantna lokalizacija korišćenjem višeslojnih neuralnih mreža u mašinskom učenju. U radovima [33] i [34] koriste se konvolucione neuralne mreže za određivanje pozicije kamere na osnovu jedne jedine RGB slike. Ovim je ponuđeno rešenje za eliminaciju problema koje unosi GPS, ali je i dalje potrebno rešavati problem preciznosti sistema.

VI. PRAĆENJE OBJEKATA

Cilj kod praćenju objekata, engl. *tracking*, je estimacija stanja jednog ili više objekata u jedinici vremena na osnovu merenja sa senzora. Obično se objekat opisuje sa tri stanja: pozicija, brzina i ubrzanje. Praćenje drugih učesnika u saobraćaju je veoma važno kako bi se izbegle saobraćajne nesreće i pravilno dolene odluke o daljem toku vožnje. Na osnovu trenutne trajektorije, sa velikom preciznošću se može predvideti sledeće stanje vozila, pa se u krajnjoj instanci mora na vreme prilagoditi brzina autonomnog vozila ili zakočiti, imajući u vidu i dužinu zaustavnog puta, koja je različita za svako vozilo. Praćenje u kombinaciji sa klasifikacijom omogućava adaptivnu promenu brzine u zavisnosti od situacije. Sistem za praćenje objekata mora da se izборi sa raznim izazovima. Često su objekti delimično ili potpuno zaklonjeni drugim objektima, a veliku poteškoću predstavlja i međusobna sličnost objekata iste klase. Problem predstavljaju i prevelika količina osvetljenja koje pada na senzore, kao i refleksija određenih objekata na staklima i prozorima.



Slika 6. Detekcija i praćenje artikuliranih ljudskih poza [40]

Filtriranje i predikcija su fundamentalni elementi svakog sistema za praćenje. To su metode korišćene za estimaciju sadašnjih i budućih stanja cilja, pozicije, brzine i ubrzanja. Postoje dva najčešće korišćena pristupa u filtriranju – filtriranje sa konstantnim koeficijentima i Kalmanovo filtriranje. Filtri sa konstantnim koeficijentima imaju prednost u vidu manje složenosti i manje potrebe za resursima, dok je prednost Kalmanovog filtra visoka preciznost praćenja tragova, pa je i daleko više primenjen u praksi. Oba pristupa ne zahtevaju puno memorije i mogu se implementirati rekurzivno. Najbolji algoritmi do sada objavljeni ostvaruju tačnost praćenja od 65-80% u slučaju praćenja automobila, a od 35-70% u praćenju pešaka, u različitim okolnostima, [35]. Takođe, u ovom kontekstu su korišćeni prošireni Kalmanov i čestični filter, [36]-[37]. U tim radovima se radi o rekurzivnim pristupima koji nisu otporni na greške u detekciji zbog šuma merenja i propuštenih detekcija. Tako su nerekurzivni pristupi dobili na značaju. Sa druge strane, veliki broj potencijalnih trajektorija cilja i veliki broj potencijalnih objekata u sceni vode do velikog prostora za pretragu. Jedan pristup za rešenje ovog problema je restrikcija skupa mogućih lokacija zarad rešavanja problema pridruživanja. U radu [38] autori imaju elegantno rešenje tako što problem svode na optimizacioni, sa zadatkom nalaženja minimuma kost funkcije. Rad [39] je fokusiran na primenu filtriranja u autonomnim vozilima i predlaže se „online“ praćenje.

Od posebnog značaja u autonomnim vozilima su praćenje i detekcija pešaka. U radu [40] se kombinuju prednosti detekcije i praćenja artikuliranih ljudskih poza, Sl. 6. Autori unapređuju detekciju ljudi i koriste hijerarhijski Gausov model, engl. *hierarchical Gaussian process latent variable model* (hGPLVM), koji im omogućava znatno pouzdaniju detekciju ljudi od drugih pristupa. U kombinaciji sa skrivenim Markovljevim modelom (HMM) mogu da prate ljude u relativno dugim vremenskim intervalima.

VII. END-TO-END UČENJE

Trenutno aktuelni pristupi u autonomnoj vožnji se sastoje od velikog broja modela, detekcije (saobraćajnih znakova, semafora, automobila, pešaka), segmentacije (puta, fasada), estimacije pokreta, praćenja učesnika u saobraćaju, rekonstrukcije. Rezultati ovih komponenata su uključeni u celokupan sistem upravljanja. Međutim, pouzdanost takvog sistema zahteva robusno rešenje kako bi se rešio veći problem, problem manipulisanja smerom kretanja i brzinom autonomnog vozila. Kao alternativa, uvedeno je učenje „od početka do kraja“, engl. *end-to-end* učenje. *End-to-end* autonomna vožnja se definiše kao vožnja koja koristi sistem koji se mapira na osnovu podataka sa senzora direktno na komande u vožnji, kao npr. ugao skretanja. U radu [41] uvodi se *end-to-end* višeslojna konvoluciona neuralna mreža koja direktno na osnovu podataka sa prednje kamere vozila prati put ispred sebe.

Trenutne *end-to-end* metode direktno mapiraju piksele u aktuatorne. Uspeh ovih metoda je za sada moguć samo u simuliranim uslovima, jer je za ove potrebe javno dostupan skup podataka limitiran. U radu [42] se uvodi alternativni pristup iskorišćavanja velikih javno dostupnih baza podataka kako bi trenirali modele. Oni formulišu autonomnu vožnju kao

problem predikcije pokreta ego vozila koristeći sasvim novu arhitekturu neuralnih mreža koje predviđaju pokrete na osnovu trenutnih stanja. U radu [43] uvodi se drugačija paradigma, zasnovana na Markovovim procesima odlučivanja, zarad postizanja boljih performansi kada se nailazi na scenarije koji su značajno različiti od onih koji se sreću prilikom treniranja. Ovde se implementira inverzno učenje kako bi ekstrahovali nepoznati parametri vožnje i time bolje rešavali scenariji koji se nisu sretali u procesu treniranja algoritama, tj. kreirani su argumentovani podaci.

VIII. ZAKLJUČAK

U ovom radu predstavljeni su osnovni algoritmi i rešenja iz oblasti kompjuterske vizije u autonomnim vozilima. Obradene su teme detekcije objekata, semantičke segmentacije, estimacije pokreta i pozicije, što uključuje SLAM algoritme i lokacizaciju, kao i teme praćenja objekata i *end-to-end* učenja. Autonomna vožnja je izuzetno aktuelna tema, domen koji u sebi okuplja mnoge oblasti, od obrade slike, kompjuterske vizije, mašinskog učenja, lokalizacije, fuzije informacija sa senzora i dr. U ovakvoj rapidno rastućoj oblasti izuzetno je važno biti u toku sa aktuelnim istraživanjima i razvijem. Stoga, doprinos ovog rada se ogleda u tome što je obuhvatio veoma široku temu kompjuterske vizije sa akcentom na pregled osnovnih algoritama i tehnika iz domena autonomne vožnje. Budući rad podrazumeva primenu predstavljenih algoritama, poređenje novih performansi na predefinisanoj setu podataka, i predlog novih rešenja. Takođe, planirano je detaljnije istraživanje u oblasti *end-to-end* učenja korišćenjem tehnike višeslojnih neuralnih mreža.

LITERATURA

- [1] W. McCulloch and P. Walter, "A Logical Calculus of Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, pp. 115–133, 1943.
- [2] T. Kanade, "Autonomous land vehicle project at CMU", CSC '86 Proceedings of the 1986 ACM fourteenth annual conference on Computer science, 1986.
- [3] R. Wallace, "First results in robot road-following", *JCAI'85 Proceedings of the 9th international joint conference on Artificial intelligence*, 1985.
- [4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset", *International Journal of Robotics Research (IJRR)*, 2013, pp. 1231–1237.
- [5] J. Janaia, F. Guney, A. Behla, and A. Geiger, "Computer Vision for Autonomous Vehicles: Problems, Datasets and State-of-the-Art", arXiv preprint arXiv:1704.05519, 2017.
- [6] H. Winner, S. Hakuli, F. Lotz, C. Singer, A. Geiger, et al., "Handbook of Driver Assistance Systems", Springer Vieweg, 2015.
- [7] R. Klette, "Vision-based Driver Assistance Systems", Technical Report CITR, Auckland, New Zealand, 2015.
- [8] H. Zhu, K. V. Yuen, L. Mihaylova, and H. Leung, "Overview of environment perception for intelligent vehicles", *IEEE Trans. on Intelligent Transportation Systems (TITS)*, pp. 1–18, 2017.
- [9] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense twoframe stereo correspondence algorithms", *International Journal of Computer Vision (IJCV)*, 47, pp. 7–42, 2002.
- [10] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski, "A database and evaluation methodology for optical flow", *International Journal of Computer Vision (IJCV)*, 92, pp. 1–31, 2011.
- [11] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset

- for semantic urban scene understanding”, Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016.
- [12] D. Kondermann et al., "The HCI Benchmark Suite: Stereo and Flow Ground Truth with Uncertainties for Urban Autonomous Driving", Proc. of IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, 2016, pp. 19-28.
- [13] M. Enzweiler and D. M. Gavrilu, "A Multilevel Mixture-of-Experts Framework for Pedestrian Classification," in IEEE Transactions on Image Processing, vol. 20, no. 10, pp. 2967-2979, Oct. 2011.
- [14] A. Broggi, M. Bertozzi, A. Fascioli, & M. Sechi, "Shape-based pedestrian detection", In Proc. IEEE Intelligent Vehicles Symposium (IV), 2000.
- [15] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun, "Pedestrian detection with unsupervised multi-stage feature learning", In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2013.
- [16] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2008.
- [17] J. Janai, F. Güney, J. Wulff, M. J. Black and A. Geiger, "Slow Flow: Exploiting High-Speed Cameras for Accurate and Diverse Optical Flow Reference Data," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 1406-1416, 2017.
- [18] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection", Proc. of the European Conf. on Computer Vision (ECCV), 2016.
- [19] M. Z. Zia, M. Stark, B. Schiele and K. Schindler, "Detailed 3D Representations for Object Recognition and Modeling," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 11, pp. 2608-2623, Nov. 2013.
- [20] X. Chen, K. Kundu, Y. Zhu, H. Ma, S. Fidler, and R. Urtasun, "3D object proposals using stereo imagery for accurate object class detection", arXiv.org, 1608.07711, 2016.
- [21] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640-651, 2017.
- [22] G. Ghiasi and C. C. Fowlkes, "Laplacian pyramid reconstruction and refinement for semantic segmentation", Proc. of the European Conf. on Computer Vision (ECCV), 2016.
- [23] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp. 770-778, 2016.
- [24] G. Floros and B. Leibe, "Joint 2D-3D temporally consistent semantic segmentation of street scenes", 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, pp. 2823-2830, 2012.
- [25] A. Kundu, V. Vineet and V. Koltun, "Feature space optimization for semantic video segmentation", In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016.
- [26] S. Sengupta, P. Sturgess, L. Ladicky, and P. H. S. Torr, "Automatic dense visual semantic mapping from street-level imagery", IROS, 2012.
- [27] T. Hackel, J. D. Wegner, and K. Schindler, "Fast semantic segmentation of 3d point clouds with strongly varying density", ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences (APRS), III-3, pp. 177 – 184, 2016,
- [28] J. Huang and S. You, "Point cloud labeling using 3d convolutional neural network", Proc. of the International Conf. on Pattern Recognition (ICPR), 2016.
- [29] G. Riegler, A. O. Ulusoy and A. Geiger, "OctNet: Learning Deep 3D Representations at High Resolutions", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 6620-6629, 2017.
- [30] A. Laddha, M. K. Kocamaz, L. E. Navarro-Serment, and M. Hebert, "Map-supervised road detection", 2016 IEEE Intelligent Vehicles Symposium (IV), Gothenburg, pp. 118-123, 2016.
- [31] C. Vogel, S. Roth, and K. Schindler, "An evaluation of data costs for optical flow", Proc. of the German Conference on Pattern Recognition (GCPR), 2013.
- [32] D. Anguelov et al., "Google Street View: Capturing the World at Street Level," in Computer, vol. 43, no. 6, pp. 32-38, June 2010.
- [33] A. Kendall, M. Grimes, and R. Cipolla, "PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization", 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, pp. 2938-2946, 2015.
- [34] F. Walch, C. Hazirbas, L. Leal-Taix'e, T. Sattler, S. Hilsenbeck, and D. Cremers, "Image-based localization with spatial lstm's", arXiv.org, 1611.07890, 2016.
- [35] B. Lee, E. Erdenee, S. Jin, M. Y. Nam, Y. G. Jung, and P. Rhee, "Multi-class multi-object tracking using changing point detection", Proc. of the European Conf. on Computer Vision (ECCV) Workshops, 2016.
- [36] J. Giebel, D. Gavrilu, and C. Schnorr, "A bayesian framework for multicue 3d object tracking", In Proc. of the European Conf. on Computer Vision (ECCV), 2004.
- [37] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. J. V. Gool, "Online multiperson tracking-by-detection from a single, uncalibrated camera", IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 33, 1820–1833, 2011.
- [38] L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows", In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2008.
- [39] B. Leibe, K. Schindler, N. Cornelis, and L. Van Gool, "Coupled detection and tracking from static cameras and moving vehicles", IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 30, 1683–1698, 2008.
- [40] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking", In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2008.
- [41] M. Bojarski et al., "End to end learning for self-driving cars", arXiv.org, 1604.07316, 2016.
- [42] H. Xu, Y. Gao, F. Yu, and T. Darrel, "End-to-end learning of driving models from large-scale video dataset", arXiv.org, 1612.01079, 2016.
- [43] S. Sharifzadeh, I. Chiotellis, R. Triebel, and D. Cremers, "Learning to drive using inverse reinforcement learning and deep q-networks", Neural Information Processing Systems (NIPS) Workshops, 2016.

ABSTRACT

This paper provides an overview of basic computer vision techniques used in autonomous vehicle domain. Research topics on machine learning algorithms and robotics with application in automotive vehicle domain are on the cutting edge of science and industry. For this reason it is of the utmost importance to have a clear insight into current research in this fast growing area. This paper provides insights into basic techniques in the domain of computer vision in autonomous vehicles. The paper discusses key references in object recognition, semantic segmentation, motion and position estimation, object tracking and end-to-end learning. In addition, the key databases for the training of machine learning algorithms are described.

AN OVERVIEW OF BASIC COMPUTER VISION TECHNIQUES IN AUTONOMOUS VEHICLES

Zlatko Veličković, Jelena Kocić, Nenad Jovičić