Zbornik radova Fakulteta tehničkih nauka, Novi Sad

UDK: 621.39
DOI: https://doi.org/10.24867/22BE28Tomic

JEDNOSTEPENI I DVOSTEPENI TREKERI VIŠE OBJEKATA

MULTIPLE OBJECT TRACKING WITH END-TO-END TRACKERS AND TRACKING-BY-DETECTION

Katarina Tomić, *Fakultet tehničkih nauka, Novi Sad*

Oblast – ELEKTROTEHNIČKO I RAČUNARSKO INŽENJERSTVO

Kratak sadržaj – *Opisan je problem praćenja objekata sa dvostepenim i jednostepenim trekerima. Performanse trekera merene su sa više različitih metrika: MOTA, IDF1, HOTA i pod-metrika HOTA. Parametri sistema su varirani u cilju izbora optimalnog rešenja.*

Ključne reči: *Praćenje objekata, YOLOv5, DeepSORT, FairMOT*

Abstract – *Multiple object tracking using tracking-by-detection and end-to end trackers is described. Tracker performances were evaluated by MOTA, IDF1, HOTA scores and the submetrics of HOTA. System paramateres were varied with the goal of optimal solution choice.*

Keywords: *Object tracking, YOLOv5, DeepSORT, FairMOT*

1. UVOD

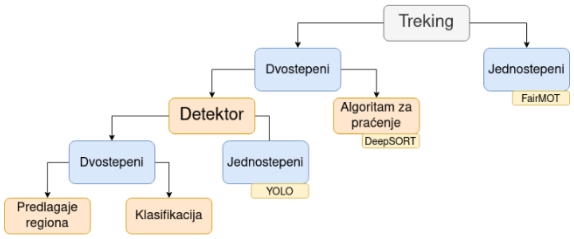
Praćenje objekata (treking) u video zapisima je jedan od klasičnih problema u kompjuterskoj viziji. Predstavlja problem estimacije trajektorije jednog ili više objekata koji se kreću u sceni.

Treking može da se podeli u dva koraka – detekcija i asocijacija. Prvo, algoritam za detekciju određuje lokaciju objekata u frejmovima, a zatim algoritam za praćenje vrši asocijaciju objekata u susednim frejmovima. U klasičnoj kompjuterskoj viziji, detekcija i treking su odvojene, kaskadne operacije, gde asocijacija frejmova zavisi od dobre detekcije objekata. Zato se veliki napor ulaže u unapređivanje algoritama za detekciju.

Trekere koji imaju odvojene module za detekciju i asocijaciju nazivamo dvostepenim. Sa druge strane, u polju dubokog učenja, ovu zavisnost je moguće prevazići. Neuralne mreže mogu istovremeno da rešavaju oba problema i dele naučene informacije.

Kada postoji jedinstven algoritam (model) koji na ulazu prima sekvencu slika, a na izlazu pruža informaciju o lokaciji objekta i njegov identitet kroz frejmove, taj model nazivamo jednostepenim *end-to-end* trekerom.

Cilj ovog rada jeste razmatranje klasičnog, dvostepenog treking rešenja i novijeg, jednostepenog rešenja, slika 1.



Slika 1. *Podela trekera*

2. ALGORITMI ZA PRAĆENJE OBJEKATA

Inicijalno, metode praćenja bile su fokusirane na vizuelna obeležja u pokušaju da reše globalni optimizacioni problem pronalaženja trajektorije objekta. Međutim, za rad ovih algoritama potrebna je obrada grupe podataka (eng. *batch*) odjednom, zbog čega nisu adekvatan izbor za *real-time* korišćenje. Neke klasične metode za asocijaciju objekata nemaju potrebu da obrađuju grupe podataka odjednom već mogu da vrše asocijaciju frejm po frejm. Zbog toga su se istraživači okrenuli ka korišćenju jednostavnijih algoritma za asocijaciju, a fokus je na učenju obeležja na kojima se zasniva asocijacija.

2.1. Dvostepeni trekeri

Detekcija i praćenje su odvojeni moduli, gde se za svaki frejm prvenstveno radi detekcija pravougaonih regiona od interesa, a zatim se rezultati prosleđuju na ulaz modula za asocijaciju koji povezuje pravougaone regione kroz frejmove. U literaturi ovaj tip trekera naziva se eng. *tracking-by-detection*, i za potrebe ovog rada nazivamo ih dvostepenim trekerima.

Pod pretpostavkom da (i) detektor koji se koristi ima veoma dobre performanse, odnosno da će tačno proizvesti detekciju za svaki objekat u svakom frejmu, (ii) i da video sekvence imaju veliki broj kadrova u sekundi (eng. *frame rate)*, odnosno da je razlika između dva susedna frejma dovoljno mala, treking postaje jednostavan problem koji se može rešiti IOU metrikom [1]. IOU metrika (eng. *intersection over union score*) meri količinu poklapanja dve površine, i daje vrednost $IOU = 1$ ukoliko se dve površine sasvim poklapaju, a $IOU = 0$ ako nemaju dodirne tačke. Algoritam inicijalno bira region sa najvećom verovatnoćom detekcije (a) i računa IOU vrednost za sve regione iz prethodnog frejma (b). Ukoliko nema preklapanja, onda se region (a) zapisuje kao novi objekat:

$$IOU_{(a,b)} = \frac{Area(a) \cap Area(b)}{Area(a) \cup Area(b)} \quad (1)$$

NAPOMENA:
Ovaj rad proistekao je iz master rada čiji mentor je bio dr Branko Brkljač, docent.

Napomenimo da se IOU metrika koristi i u kontekstu samih detektora. Naime, NMS (eng. *non maximum suppression*) algoritam računa IOU metriku za sve predložene regione i odbacuju one koji imaju višu vrednost od izabranog praga (eng. *threshold*). Odbacivanjem regiona koji imaju veoma visoku IOU vrednost (visoko preklapanje) omogućava se eliminacija višestrukih detekcija.

Za ranije pomenuti IOU algoritam usvojena je pretpostavka o kvalitetu detektora i frekvenciji frejmova. Međutim, u realnom sistemu ne možemo da računamo na savršen detektor. U koraku asocijacije teži se kompenzaciji grešaka koje pravi detektor, ali ne samo grešaka već i drugih artefakata u videu, poput okluzija objekta. Na primer SORT algoritam [2], asocira detekcije proizvedene od strane detektora pomoću Hungarian algoritma i Kalman filtera. Ovaj jednostavni pristup postiže dovoljno veliku brzinu izvršavanja da se koristi u *real-time* sistemima i daje zadovoljavajuću tačnost.

Iako SORT algoritam daje dobru tačnost, javlja se drugi problem – identiteti objekata se usled okluzija prebacuju sa objekta na objekat. DeepSORT metoda rešava navedeni problem dodavanjem informacija o izgledu na postojeće podatke o pomeraju objekta. Konkretno, implementirana je konvoluciona neuralna mreža koja povećava robusnost na okluzije i prebacivanje identiteta, a održava brzinu izvršavanja. Ažurirana verzija ovog algoritma je nazvana StrongSORT algoritam, u kome su starije metode zamenjene novijim.

Još jedan pristup rešavanju problema asocijacije jeste reidentifikacija (ReID) [3]. ReID se postavlja kao dodatak na osnovnu mrežu (engl. *backbone*), i postoje dve vrste u odnosu na funkciju cene: klasifikacioni i klasterizacioni tip. Ako se pristupa kao problemu klasifikacije, funkcija cene koristi *ID-discrimnative embedding* (IDE). Da bi model mogao da napravi predikciju o identitetu, dodaju se slojevi na postojeću mrežu koji imaju diskriminativnu funkciju. U suprotnom, ako se ReID posmatra kao klasterizacija, za funkciju cene uzima se neka metrika, poput *triplet loss* metrike.

2.2. Jednostepeni trekeri

U razvoju detektora i algoritama za asocijaciju (ReID) napravljen je značajan napredak i postignute su dobre performanse. Međutim, dvostepeni trekeri imaju problem u skaliranju, odnosno kada postoji veliki broj objekata u sceni oni ne mogu da postignu *real-time* vreme izvršavanja. Ovo se dešava zato što asocijacija treba da se primeni na svaki region koji detektor generiše u svakom frejmu. Primećeno je da ukoliko se napravi model u kom detekcija i asocijacija koriste isti model ili mrežu, tj. dele dele informacije, vreme izvršavanja se znatno skraćuje, ali performanse trekera mogu i da opadaju. Jednostepeni trekeri, u literaturi eng. *end-to-end* (ili *one-shot*) trekeri, jesu trekeri koji imaju jedinstveni model ua izdvajanje obeležja za oba zadatka [4]. Primer koji će biti analiziran u nastavku je FairMOT algoritam [5].

3. DETEKCIJA OBJEKATA

Na ovom mestu ćemo se ukratko osvrnuti i na problem detekcije objekata, koji se takođe može razmatrati u jednostepenom i dvostepenom obliku, slika 1.

Detekcija objekata u slici podrazumeva rešavanje dva zadatka: 1) detekcije postojanja objekta u slici (problem klasifikacije), i 2) određivanja relativne pozicije objekta (problem lokalizacije). Ovi zadaci mogu da se rešavaju odvojeno, kada takve sisteme nazivamo dvostepenim, ili istovremeno, kada ih nazivamo jednostepenim, slika 1.

Primer dvostepenog detektora je R-CNN model. R-CNN model prvo koristi algoritam za predlaganje regiona od interesa koji potencijalno sadrže objekte, a zatim na svaki pravougaoni region primenjuje klasifikator. Dobijeni sistem daje visoku tačnost, ali je zbog ponovnog izdvajanja obeležja za svaki od potencijalnih regiona od interesa veoma spor, što dvostepene detektore često čini nepogodnim za rad u realnom vremenu. Generalno, najjednostavniji algoritam lokalizacije je metod klizajućeg prozora, gde se cela slika pretraži sa prozorima različitih veličina, a u drugom koraku biraju se regioni koji imaju najveću verovatnoću da sadrže objekat i klasifikuju se kategorije objekata unutar njih. Dakle, predlaganje regiona od interesa predstavlja rešavanje problem lokalizacije objekata u slici, koji se nakon potvrđene detekcije mogu još dodatno lokalizovati u odnosu na prvobitni položaj regiona od interesa.

Jednostepeni detektori lokalizaciju i klasifikaciju realizuju u istom koraku, tj. 1) predviđaju pozicije i dimenzije pravougaonih regiona od interesa i 2) kategorije objekta za celu sliku u jednoj iteraciji. Time se izbegava nezavisno izračunavanje obeležja za svaki od regiona u slici, što dovodi do znatnog povećanja brzine, ali generalno i manje tačnosti. Primer jednostepenih detektora je familija YOLO detektora, koji problem detekcije i lokalizacije formulišu kao jedinstven regresioni problem. Prva verzija YOLO detektora pokazivala je veliki porast u brzini detekcije, ali je tačnost detekcije opala, naročito za male objekte. Vremenom su dodata različita poboljšanja na osnovni YOLO model [6].

Za drugu verziju modela YOLOv2 fokus je bio na poboljšaju tačnosti, stoga ovaj model ima bolju sposobnost generalizacije, veću preciznost za male objekte, manju grešku loalizacije i povećanu rezoluciju. U trećoj verziji modela, YOLOv3, fokus je bio na ubrzanju ali je tačnost opala. Slični pristupi primenjeni su i za narednu verziju detektora YOLOv4, ali sa više uspeha. Detektor YOLOv5 nije napravljen od strane originalnih autora i zato se generalno smatra da je ime detektora neadekvatno, jer tehnički nije nova verzija, već reimplementacija. Bez obzira na to, YOLOv5 model ima veću tačnost od prethodnih, održanu *real-time* detekciju i uz to, olakšano korišćenje modela.

4. FAIRMOT ALGORITAM

Autori FairMOT algoritma [4] ukazali su na tri razloga pada u performansama jednostepenog trekera. Faktori koji utiču na pad su (i) postojanje unapred definisanih dimenzija za pravougaone regione (eng. *anchors*), (ii) nepravilno deljenje obeležja za zadatak detekcije i ReID - ReID zahteva mnogo više detalja od detekcije. Treći problem (iii) nastaje usled dimenzija obeležja, jer detekcije zahtevaju mnogo manju dimenzionalnost nego reidentifikacija.

FairMOT arhitektura sadrži dve homogene grane nadograđene na osnovnu mrežu. Grana za detekciju koristi CenterNet arhitektu [7] koja nema unapred definisane dimenzije pravougaonih regiona, već estimira pozicije centralne tačke svakog objekta. Grana za reidentifikaciju estimira obeležja objekta opisanog centralnom tačkom. Ove dve grane su sasvim homogene, za razliku od dvostepenih trekera koji reidentifikaciju postavljaju kao sekundarni problem nakon detekcije. FairMOT eliminiše pristrasnost ka detekciji, i uči visoko kvalitetna obeležja za reidentifikaciju.

FairMOT za osnovnu mrežu (engl. *backbone*) koristi ResNet-34 arhitekturu na koju su primenjene određene korekcije (DLA, DCN) kako bi dobijena obeležja bila bolja. CenterNet arhitektura, koja se koristi za granu detekcije, ima tri paralelna izlaza koji procenjuju: mape odziva (eng. *heatmaps*), pomeraj centra (eng. *offset*) i visinu i širinu regiona objekta. NMS algoritam se primenjuje na mape odziva i čuva tačke čiji je odziv iznad definisanog praga. Ove tačke se proglašavaju centrima objekta, a zatim se regresijom procenjuju visina i širina regiona objekta. Izlaz sa informacijom o pomeraju centra služi za precizniju lokalizaciju objekta. Grana za reidentifikaciju generiše obeležja, tako da su razlike u obeležjima između objekata različitih klasa veće nego unutar klasne razlike. Asocijacija se zasniva na ovim obeležjima. Naime, prvo se inicijalizuje N identiteta (eng. *tracklets*), gde je N broj objekata u prvom frejmu. U narednom frejmu, identiteti pravougaonih regiona se povezuju sa prethodnim pomoću Kalman i Hungarian algoritma (DeepSORT algoritam). Detekcijama koje nisu asocirane inicijalizuje se novi identitet, a identiteti koji nisu asocirani sa nekim regionom se čuvaju *M* broj frejmova pa se ponovo asociraju ako se pojave njihovi regioni, u suprotnom se brišu (starost identiteta).

5. REZULTATI

5.1 Metrike

Oblast praćenja objekata je doživela veliku ekspanziju, delom zahvaljujući industriji autonomnih vozila. Usled porasta interesovanja, pojavljuju se i nove ideje, novi doprinosi u načinima evaluacije. Praćenje objekata je složen zadatak za koji je potrebna tačna detekcija, lokalizacija i asocijacija kroz vreme. Za kvantitativno merenje trekera, korišćene su sledeće metrike: MOTA, IDF1 i HOTA.

MOTA (eng. *Multiple Object Tracking Accuracy*) je najčešće korišćena metrika za evaluaciju trekera. Meri se poklapanje detekcija i anotiranih objekata (*engl. Ground Truth – GT)*. Greške koje se javljaju pri poklapanju su lažni negativi (FN) i lažni pozitivi (FP), koji su mere detekcije. Još jedna greška koja ulazi u računanje MOTA metrike je prebacivanje identiteta (IDSW), što predstavlja meru asocijacije. MOTA (2) ne meri grešku lokalizacije, i naglađava performanse detekcije preko asocijacije.

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IDSW_t)}{\sum_t (GT)_t} \quad (2)$$

IDF1 (eng. *Identity F1 score*) akcentuje merenje tačnosti asocijacije više nego detekcije i koristi se uglavnom kao propratna metrika uz MOTA. Meri poklapanje anotiranih trajektorija objekta sa predikovanim trajektorijama, i računa odnos tačno identifikovanih detekcija prema ukupnom broju izračunatih detekcija. IDF pruža estimaciju o broju jedinstvenih objekata u sceni, takođe kao i MOTA ne meri grešku lokalizacije.

$$IDF1 = \frac{|IDTP|}{|IDTP| + 0.5|IDFN| + 0.5|IDFP|} \quad (3)$$

MOTA i IDF1 prenaglašavaju bitnost detekcije i asocijacije, respektivno. HOTA (eng. *Higher Order Tracking Accuracy*) eksplicitno meri obe vrste greške i kombinuje ih da bi pružila jedinstvenu vrednost za evaluaciju trekera. Pored toga, HOTA meri i grešku lokalizacije što nije slučaj kod MOTA i IDF1 metrike. Ovakav pristup evaluacije pruža detaljan uvid u greške koje treker pravi i tako omogućava precizno korigovanje.

HOTA se dekomponuje na pod-metrike koje opisuju zasebne vrste problema. Na primer, tačnost detekcije (DetA) (4) je procenat poklapanja detekcija, dok je tačnost asocijacije (AssA) (5) poklapanje asociranih trajektorija, uprosečeno kroz sve detekcije.

$$DetA = Det - IoU = \frac{|TP|}{|TP| + |FN| + |FP|} (4) \quad (4)$$

$$AssA = \frac{1}{|TP|} \sum Ass - IoU(c) \quad (5)$$

Tačnost lokalizacije (LocA) (6) računa se kao prosečno poklapanje anotiranih detekcija sa procenjenim, nad celim skupom podataka.

$$LocA = \frac{1}{|TP|} \sum Loc - IoU(c) \quad (6)$$

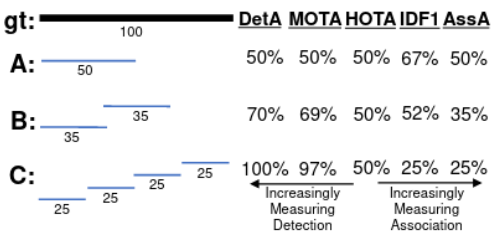
Konačni HOTA indeks (7) je geometrijska srednja vrednost DetA i AssA za različite pragove lokalizacije.

$$HOTA_\alpha = \sqrt{DetA_\alpha \cdot AssA_\alpha} \quad (7)$$

5.2. Diskusija

Testirane su performanse dva modela za praćenje objekata: dvostepeni YOLO model sa StrongSORT algoritmom i jednostepeni FairMOT model. Sa njihovih repozitorijuma preuzete su težine već obučenih modela i optimalna podešavanja. Kao što je opisano, ovi modeli su dva sasvim različita pristupa istom problemu, ali performanse oba se rangiraju pri samom vrhu u njihovim kategorijama. Korišćen je MOT16 skup [8] i oba modela su testirana za samo jednu klasu – ‘pešak‘.

Na slici 2 mogu se uočiti glavne razlike između MOTA, IDF1 i HOTA metrika. Ilustrovana su tri različita trekera poređana po porastu tačnosti detekcije i smanjenju tačnosti asocijacije. Dok HOTA balansirano uzima u obzir oba zadatka, MOTA je očigledno naklonjena ka detekciji, a IDF1 ka asocijaciji.



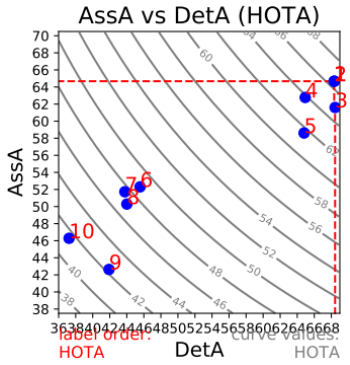
Slika 2. HOTA i druge metrike za tri trekera (A, B, C) sa sve boljom detekcijom i lošijom asocijacijom [9].

Parametri YOLO trekera koji su menjani u svrhu analize njihovog uticaja na celokupnu tačnost jesu: skor pouzdanosti detektora, odnosno donji prag verovatnoće detekcije, zatim maksimalna starost neasociranog identiteta, prag za asocijaciju, i IOU prag za NMS algoritam. Parametri koji utiču na performanse FairMOT modela su: NMS prag, prag pouzdanosti detekcije i dimenz. ReID modela. Sekvence slika propuštene su kroz svaki od ovih modela i poređeni su rezultati. Modeli su numerisani kao u tabeli 1. Originalna podešavanja odgovaraju „Modelu 2.0“ i „Modelu 1.0“.

Tabela1. Promene parametara u odnosu na polazne modele

Model 2.1	Model 2.0	Model 2.2	Model 2.3	Model 2.4
Prag pouzdanosti +0.2	Osnovni dvostepeni model	Max starost -15	Prag za asocijaciju +0.3	IOU +0.2
Model 1.1	**Model 1.0**	**Model 1.2**	**Model 1.3**	**Model 1.4**
NMS -0.2	Osnovni jednostepeni model	ReID 1024	Prag pouzdanosti -0.2	ReID 512

Performanse trekera na HOTA pod-metrikama detekcije i asocijacije prikazane su na slici 3. Plave tačke su trekeri sa različitim podešavanjima, numerisani kao u tabeli 1. Na slici 2 može se primetiti da se za različite kombinacije tačnosti detekcije i asocijacije dobijaju isti HOTA indeksi. Zato su HOTA indeksi ilustrovani sivim izolinijama.



Slika 3. T*ačnost trekera u detekciji i asocijaciji. Crvena linija predstavlja Pareto optimalni front opisan u [9].*

Povećanje praga pouzdanosti detekcije dovodi do skoka u tačnosti YOLO detektora (Model 2.1) u odnosu na osnovni model (Model 2.0). Ovo ponašanje je očekivano, jer povećavanjem praga eliminišemo detekcije koje su manje sigurne. Menjanje drugih YOLO parametara dovodi do pada tačnosti. Autori FairMOT modela sugerišu da prevelika dimenzionalnost ReID modela negativno utiče na tačnost. Na slici 3. možemo videti i da u odnosu na osnovni model (Model 1.0), promene parametara jednostepenog trekera dovode do pogoršanja rezultata. Svi YOLO modeli imaju manju HOTA metriku u odnosu na FairMOT modele. U tabeli 2 prikazane su i vrednosti drugih metrika za najbolje modele iz obe grupe.

Tabela 2. Poređenje metrika za najbolje modele.

	MOTA	IDF1	HOTA
YOLO Model 2.1	49.69	61.33	48.46
FairMOT Model 1.0	83.36	80.76	66.33

Primećujemo da u slučaju ova dva trekera, FairMOT daje bolje rezultate na svim metrikama.

6. ZAKLJUČAK

Rad analizira dva rešenja za praćenje objekata u realnom vremenu. Dvostepeni YOLO+StrongSORT model kombinuje različite algoritme, dok je jednostepeni FairMOT model složenije strukture i rešava problem praćenja objekata na celovit način. Na osnovu eksperimenata pokazano je da FairMOT daje bolje rezultate, ali je opšti zaključak da je dvostepeni model na bazi YOLO detektora intuitivniji za korišćenje.

7. LITERATURA

[1] Bochinski, Erik, Volker Eiselein, and Thomas Sikora. "High-speed tracking-by-detection without using image information." *2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS)*. IEEE, 2017.

[2] Wojke, Nicolai, Alex Bewley, and Dietrich Paulus. "Simple online and realtime tracking with a deep association metric." *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017.

[3] Zheng, Zhedong, Liang Zheng, and Yi Yang. "A discriminatively learned CNN embedding for person reidentification." ACM Transactions on multimedia computing, communications, and applications (TOMM) 14.1 (2017): 1-20.

[4] Munjal, Bharti, et al. "Joint detection and tracking in videos with identification features." *Image and Vision Computing* 100 (2020): 103932.

[5] Zhang, Yifu, et al. "FairMOT: On the fairness of detection and re-identification in multiple object tracking." *International Journal of Computer Vision* 129.11 (2021): 3069-3087.

[6] Zaidi, Syed Sahil Abbas, et al. "A survey of modern deep learning based object detection models." *Digital Signal Processing* (2022): 103514.

[7] Zhou, Xingyi, Dequan Wang, and Philipp Krähenbühl. "Objects as points." arXiv preprint arXiv:1904.07850 (2019).

[8] https://motchallenge.net/data/MOT16/ (10. okt. 2022).

[9] Luiten, Jonathon, et al. "Hota: A higher order metric for evaluating multi-object tracking." *International journal of computer vision* 129.2 (2021): 548-578.

Kratka biografija:

Katarina Tomić rođena je u Novom Sadu 1996. god. Upisala je master studije na Fakultetu tehničkih nauka na studijskom programu Energetika, elektronika i telekomunikacije – Obrada signala. Osnovne akademske studije završila je 2019. godine na studijskom programu Biomedicinsko inženjerstvo.