

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

# **Praćenje objekata u videozapisima**

*Marko Tutić*

*Voditelj: Izv. prof. dr. sc. Zoran Kalafatić*

Zagreb, siječanj 2022.

# SADRŽAJ

<b>1. Uvod</b>	<b>1</b>
1.1. O praćenju objekata . . . . .	1
1.2. Izazovi u praćenju objekata . . . . .	2
<b>2. Razine praćenja objekata</b>	<b>3</b>
2.1. Praćenje jednog objekta . . . . .	3
2.1.1. Mjerenje performansi algoritama za praćenje jednog objekta .	4
2.2. Praćenje više objekata . . . . .	4
2.2.1. Mjerenje performansi algoritama za praćenje više objekata . .	5
<b>3. Metode praćenja objekata</b>	<b>6</b>
3.1. Metode bazirane na značajkama . . . . .	6
3.1.1. Optički tok . . . . .	6
3.1.2. Mean-shift . . . . .	8
3.2. Metode bazirane na segmentaciji . . . . .	8
3.3. Metode baziranje na procjeni funkcije gustoće vjerojatnosti . . . . .	9
3.3.1. Kalmanov filter . . . . .	9
3.3.2. Čestični filter . . . . .	9
3.4. Metode baziranje na strojnom učenju . . . . .	10
3.4.1. Praćenje objekata pomoću sijamskih neuronskih mreža . . . .	10
3.4.2. DeepSORT . . . . .	10
3.4.3. ROLO . . . . .	11
3.4.4. MDNet . . . . .	12
<b>4. Zaključak</b>	<b>14</b>
<b>5. Literatura</b>	<b>15</b>

# 1. Uvod

## 1.1. O praćenju objekata

Praćenje objekata je grana računalnog vida kojoj je cilj detektirane objekte u nekom inicijalnom okviru videozapisa pratiti u nadolazećim okvirima. Praćenje objekta se također može definirati kao problem procjene putanje objekta na slici. Praćenje objekata se razlikuje od detekcije objekata po tome što je cilj detekcije objekata odrediti sve objekte i njihove pozicije na jednom okviru te se svaka detekcija tretira nezavisno od neke druge dok je cilj praćenja objekata detektirane ili označene objekte pratiti i razlikovati u svim nadolazećim okvirima na temelju prijašnjih pozicija i karakteristika objekta. Dodatno ako je objekt u okviru zaklonjen, detektor objekata će ga propustiti detektirati, dok će se praćenjem objekta potencijalno moći locirati na temelju povijesnih podataka o kretanju objekta. Praćenje objekata dodaje vremensku komponentu detekcijama objekata te omogućuje dodatnu analizu ponašanja objekata koja primjenom isključivo detekcije nije moguća.

Jedan od pristupa u praćenju objekata je praćenje detekcijom. U ovakvim pristupima koristi se detektor objekata za određivanje objekata na jednoj slici, a zadatak praćenja objekata je onda asocirati detektirane objekte s objektima na prijašnjim okvirima. Dobro praćenje objekata u takvom sustavu bi moralo moći se oporaviti od neuspješno detektiranih objekata i lažno detektiranih objekata na okviru[3].

Drugi pristup ne koristi detekcije objekata za praćenje već se u početnom okviru zadaju objekti koji se žele pratiti te se oni prate u nadolazećim okvirima na temelju značajki koje opisuju izgled i poziciju objekta. Problem ovakvog pristupa je što se ne mogu detektirati objekti koji se u međuvremenu pojave u okvirima.

Neke od primjena praćenja objekata su[14]:

- Prepoznavanje pokreta
- Automatiziran nadzor
- Indeksiranje videozapisa

- Prepoznavanje gesti i druge interakcije između čovjeka i računala
- Praćenje prometa
- Navigacija vozila

## 1.2. Izazovi u praćenju objekata

Praćenje objekata otežava šum u slikama, kompleksno kretanje objekta, zaklanjanje i nestajanje objekta na slici, promjene osvjetljenja i drugih komponenti slike te potreba za obradom u stvarnom vremenu. Zbog toga se uvode dodatne pretpostavke prilikom praćenja objekata. Neke od tih pretpostavki su da se objekt neće kretati velikom brzinom te da neće naglo mijenjati smjer i brzinu kretanja. Posljedica tih pretpostavki je da je potreban manji prostor pretraživanja za određivanje pozicije objekata. Alternativno, problem se može pojednostavniti uvođenjem nekih apriornih pretpostavki o kretanju objekata[14].

Dodatni problemi koji se javljaju prilikom praćenja objekata i otežavaju praćenje su zamjena identiteta, varijacije u prostornim skalama i veliki broj objekata na slikama. Objekti u okvirima videozapisa mogu zauzimati različite udjele okvira u određenim trenucima, odnosno kada je objekt blizu, zauzimat će veći dio okvira, a dok je udaljeniji zauzimat će manji dio okvira. Algoritam praćenja mora biti invarijantan na promjene veličine objekata. Veliki broj objekata u okviru otežava praćenje objekata jer je teže ekstrahirati značajke. Do zamjene identita često dolazi kada se objekti u slikama međusobno približe i zatim ponovo udalje ili kada se mimoilaze.

## 2. Razine praćenja objekata

Algoritmi praćenja objekata se u ovisnosti o broju objekata koji se prate dijele na algoritme praćenja jednog objekta (engl. single object tracking, SOT) i algoritme praćenja više objekata (engl. multiple object tracking, MOT).

### 2.1. Praćenje jednog objekta

Algoritmi praćenja jednog objekta (engl. single object tracking) prate jedan objekt u svakom okviru videozapisa. Praćenje jednog objekta ubraja se u kategoriju praćenja bez korištenja detekcije objekata. Cilj takvog praćenja je da korisnik ili prethodno trenirani model detekcije objekata može zadati inicijalnu poziciju nekog objekta, a algoritam dalje prati taj objekt isključivo na temelju karakteristika objekta i promjena unutar nadolazećih okvira. U okvirima se mogu nalaziti i drugi slični objekti, ali se algoritam mora fokusirati samo na jedan objekt te praćenje mora biti agnostično s obzirom na razred objekta.

Algoritmi praćenja jednog objekta se dijele na kratkotrajne i dugotrajne algoritme praćenja jednog objekta. Kratkotrajni algoritmi praćenja ne moraju pratiti objekt nakon zaklanjanja objekta na slici, dok se dugotrajni algoritmi praćenja moraju oporaviti od privremenog zaklanjanja objekta, odnosno moraju biti u stanju ponovo locirati objekt prilikom ponovnog pojavljivanja objekta na slici. Zaklanjanje objekta događa se ako objekt na slici prestane biti vidljiv jer interagira s nekim drugim objektom na slici, ali se objekt i dalje nalazi u području koje obuhvaća slika. Oporavak od zaklanjanja se primarno bazira na pretpostavki da zaklonjeni objekt neće raditi nagle promjene smjera i neće imati veliku brzinu. Neki od poznatih algoritama za kratko praćenje objekata baziraju se na sijamskim neuronskim mrežama ili na primjeni korelacijskog filtera. Postoje i algoritmi praćenja koji se baziraju na drugim metodama, no algoritmi praćenja bazirani na sijamskim neuronskim mrežama ili korelacijskom filteru najčešće daju najbolje rezultate (VOT 2018). Algoritmi praćenja koji koriste korelacijski filter daju bolje rezultate generalno, ali algoritmi praćenja bazirani na sijamskoj neuronskoj

mreži daju bolje rezultate kada je potrebno pratiti objekte u stvarnom vremenu. Razlika između algoritama praćenja koji koriste korelacijske filtere i algoritama baziranih na sijamskim neuronskim mrežama je činjenica se algoritmi praćenja koji koriste korelacijske filtere uče tijekom izvođenja (engl. on-line), dok su algoritmi praćenja bazirani na sijamskim neuronskim mrežama nepromjenjivi tijekom izvođenja (engl. off-line trackers).

### **2.1.1. Mjerenje performansi algoritama za praćenje jednog objekta**

Za evaluaciju performansi algoritama za praćenje jednog objekta prema VOT2015 koriste se preciznost, robusnost, očekivano prosječno preklapanje i broj obrađenih okvira po sekundi (engl. frames per second)[6]. Preciznost se određuje kao omjer presjeka i unije između dobivenih okvira koji omeđuju objekt i označenog okvira objekta. Mjerenje preciznosti se obavlja za svaki okvir i kao konačan rezultat se uzima prosjek preciznosti sve okvire. Robusnost mjeri koliko dobro algoritam prati objekte u situacijama zaklanjanja, mijenjanja osvjetljenja, veličine objekta, zamućenja uzrokovanog pokretom i sl. Robusnost se mjeri kao broj gubitaka pozicije objekta. Do gubitka pozicije dolazi kada omjer presjeka i unije između stvarne pozicije objekta i pozicije koju daje algoritam praćenja padne ispod određenog praga. Prosječno očekivano preklapanje je mjera koja kombinira preciznost i robusnost u jedinstvenu mjeru.

Za evaluaciju algoritama za dugotrajno praćenje objekata koristi se prilagođena F-vrijednost. Prilagođena F-vrijednost ovisi o preciznost i odzivu. Preciznost za ove algoritme praćenja je definirana kao srednja vrijednost omjera presjeka i unije između stvarne pozicije objekta i pozicije dobivene algoritmom praćenja. Pri tome se uzimaju u obzir samo okviri u kojima je algoritam praćenja vratio rezultat, odnosno prati neki objekt. Odziv također računa srednju vrijednost svih presjeka nad unijom, ali u obzir uzima okvire u kojima je označen objekt[7].

## **2.2. Praćenje više objekata**

Algoritmi praćenja više objekata (engl. Multiple object tracking) za cilj imaju odrediti i jedinstveno identificirati više objekata u videozapisu te pratiti i čuvati identitete svih objekata. Takvi algoritmi nemaju apriorno znanje o broju i izgledu objekata. Za razliku od algoritama praćenja jednog objekta, ovi algoritmi se izrazito oslanjaju na algoritme detekcije objekata[4].

Za praćenje objekata u ovom pristupu može se koristiti batch metoda ili online

metoda. Batch metoda koristi i buduće okvire za određivanje identiteta objekta te daje kvalitetnije rezultate praćenja od online metode. Druga metoda je online metoda praćenja i ona koristi samo trenutne i prijašnje informacije za donošenje odluka u trenutnom okviru. Ta metoda, iako lošija, je nekada jedina opcija jer u nekim primjenama nije moguć pristup budućim okvirima[4].

Većina algoritama praćenja više objekata dijeli određene faze u izvođenju. Te faze su: faza detekcije, ekstrakciranje značajki/predviđanje kretanja, faza afiniteta i faza asocijacije. U fazi detekcije algoritam detekcije objekata vraća podatke o detekcijama objekata u nekom okviru. U fazi ekstrakcije značajki/predviđanja kretanja se analiziraju izgledi, kretanje i interakcije objekata te se po potrebi predviđa sljedeća pozicija objekta. U fazi afiniteta se koriste značajke objekata i predviđanje kretanja za izračun sličnosti između parova detekcija. U fazi asocijacije se objektima dodjeljuju pripadni identifikatori[4].

### 2.2.1. Mjerenje performansi algoritama za praćenje više objekata

Neke mjere za evaluiranje performansi algoritama za praćenje više objekata su većinski praćene putanje, većinski izgubljene putanje i djelomično praćene putanje. Većinski praćene trajektorije se definiraju za svaki objekt zasebno te se trajektorija tog objekta smatra većinski praćenom ako je algoritam praćenja uspješno pratio objekt u više od 80% okvira. Većinski izgubljene trajektorije su one trajektorije za koje je algoritam praćenja uspješno pratio objekt u manje od 20% okvira. Djelomično praćene trajektorije su one trajektorije za koje je algoritam praćenja uspješno pratio objekt u između 20% i 80% okvira. Dodatna mjera koja se koristi u višestrukom praćenju objekata je preciznost višestrukog praćenja objekata (engl. multiple object tracking accuracy, MOTA). Ta preciznost se definira sljedećom formulom:

$$MOTA = 1 - \frac{\sum_t FN_t + FP_t + IDS_t}{\sum_t GT_t}$$

gdje  $FN_t$  predstavlja broj propuštenih praćenja objekata,  $FP_t$  je definiran kao broj praćenja nepostojećih objekata i  $IDS_t$  je broj zamjena identiteta u okviru  $t$ . Za svaki okvir se ove mjere izračunavaju zasebno te se konačno sumiraju. U nazivniku se zbrajaju stvarne pozicije objekata za sve okvire. Ispravnost praćenja u svakom okviru se definira omjerom presjeka i unije između pozicije objekta koju je dao algoritam praćenja i stvarne pozicije. Sve pozicije koje imaju omjer presjeka i unije veći od određenog praga se smatraju ispravno praćenima.

## 3. Metode praćenja objekata

Metode praćenja objekata se prema Verma[12] dijele se u 5 kategorija:

- metode bazirane na značajkama
- metode bazirane na segmentaciji
- metode bazirane na procjeni funkcije gustoće vjerojatnosti
- metode bazirane na izgledu
- metode bazirane na učenju

### 3.1. Metode bazirane na značajkama

Praćenje objekata u ovom pristupu bazira se na temelju značajki poput boje, teksture, gradijenata i sl. Značajke objekta koji se prati se koriste u daljnjim okvirima za ponovno prepoznavanje objekata. U idućem okviru iz dijelova slike se ekstrahiraju značajke i uspoređuju sa značajkama objekta po nekom kriteriju sličnosti te dio slike koji je najbliži značajkama objekta se određuje kao praćeni objekt u novom okviru[12][11]. Pretpostavka ovakvog pristupa je da će značajke objekta biti jedinstvene, odnosno da se u slici neće nalaziti ni jedan segment kojemu su značajke slične značajkama objekta. Metode bazirane na značajkama se dalje dijele na metode koje koriste boje objekta, teksture, točke od interesa i optički tok.

#### 3.1.1. Optički tok

Optički tok (engl. optical flow) je uzorak kretanja objekata između slijednih okvira izazvanih kretanjem objekta ili kamere. Optički tok se može predstaviti kao polje vektora u kojem svaki vektor pokazuje kretanje jedne točke između prvog i drugog okvira. Optički tok se bazira na pretpostavkama da se intenziteti piksela objekta na slici neće naglo mijenjati kroz vrijeme te da susjedni pikseli imaju slično kretanje. Ako pretpostavimo da je intenzitet piksela u nekom trenutku  $t$  zadan funkcijom



$I(x, y, t)$ , onda na temelju prethodnih pretpostavki za sljedeći okvir možemo tvrditi da je  $I(x, y, t) = I(x + dx, y + dy, t + dt)$ . Iz Taylorovog reda se dalje može dobiti jednačba optičkog toka za svaku točku u slici:  $f_x u + f_y v + f_t = 0$ .  $f_x$  i  $f_y$  su gradijenti slike, a  $f_t$  je gradijent po vremenu, a  $u$  i  $v$  su nepoznanice koje predstavljaju brzinu promatrane točke u  $x$  i  $y$  smjeru i cilj je odrediti nepoznanice za točke od interesa[10]. Kako je zadana samo jedna jednačba za dvije nepoznanice, sustav jednačbi nije rješiv klasičnim metodama te se stoga primjenjuju druge metode rješavanja poput metode Lucasa i Kanadea.

Postupke za procjenu optičkog toka dijelimo na postupke za dobivanje rijetkih(engl. sparse) i gustih(engl. dense) optičkih tokova. Rijetki optički tokovi se fokusiraju na konačan broj piksela za koje se onda određuju vektori toka, dok gusti optički tok funkcionira na globalnoj razini i daje vektore toka za sve piksele unutar okvira. Metoda koja se često koristi za određivanje rijetkog optičkog toka je metoda Lucasa i Kanadea. Za rješavanje gustog optičkog toka se često koristi Farnebackova metoda.

Metoda Lucasa Kanadea za rješavanje sustava koristi susjedne piksele, pri čemu se pretpostavlja da je tok lokalno gladak, odnosno da nema velikih razlika u toku susjednih piksela te da susjedni pikseli imaju jednak pomak. Pod tim pretpostavkama se mogu dobiti dodatne jednačbe koje omogućavaju određivanje nepoznanica  $u$  i  $v$ . Metoda Lucasa i Kanadea se ne snalazi dobro s brzim kretanjima točaka te se stoga za detektiranje svih kretanja često primjenjuje struktura piramide. U toj strukturi na najnižoj razini se nalaze slike velikih dimenzija u kojima se detektiraju manja kretanja. Na višim razinama se koriste slike malih dimenzija u kojima velika kretanja postaju manja kao posljedica skaliranja. Takva kretanja se u manjim slikama mogu odrediti klasičnim algoritmom Lucasa i Kanadea . Rezultat korištenja te piramide je mogućnost određivanja i velikih i malih kretanja piksela u videozapisu.

Farnebackova metoda [5] određuje optički tok za sve piksele u slici na temelju trenutnog i prethodnog okvira. Metoda aproksimira susjedstvo svakog piksela u trenutnom i prethodnom okviru kvadratnim polinomom. Promatranjem transformacija u polinomu izazvanih translacijom piksela izvodi se metoda za procjenu pomaka piksela. Procjene pomaka se dodatno obrađuju te je rezultat obrade robustan algoritam određivanja optičkog toka. [5] Za praćenje objekata pomoću optičkog toka se odabiru točke od interesa koje se onda prate optičkim tokom.

### 3.1.2. Mean-shift

Mean-shift je neparametarska metoda za nalaženje maksimuma funkcije gustoće na temelju uzorkovanih podataka koji predstavljaju tu funkciju. Mean-shift je iterativna metoda u kojoj se kreće od neke početne procjene  $\vec{x}$ . Nadalje je potrebno izračunati težinski prosjek za ponovnu procjenu sredine. Za to se koriste jezgrene funkcije koje ovise o nekoj točki iz susjedstva i točki  $\vec{x}$ . Težinski prosjek se onda dobiva formulom  $\vec{m}(\vec{x}) = \frac{\sum_{\vec{x}_i \in N(\vec{x})} K(\vec{x}_i - \vec{x}) \vec{x}_i}{\sum_{\vec{x}_i \in N(\vec{x})} K(\vec{x}_i - \vec{x})}$ . Iz toga se može odrediti  $\vec{m}(\vec{x}) - \vec{x}$  što predstavlja pomak sredine. Algoritam mijenja trenutnu procjenu postavljajući vrijednost  $\vec{m}(\vec{x})$  kao novu  $\vec{x}$  vrijednost. Postupak se iterativno ponavlja sve dok postupak ne konvergira.

Mean-shift se ubraja u algoritme praćenja za koje nije potrebno strojno učenje. Za korištenje mean-shift algoritma za praćenje objekta, prvo je potrebno zadati područje koje se prati. Ako se objekt u idućim okvirima pomakne, mean-shift iterativno pomiče omeđujući pravokutnik objekta iz prethodnog okvira kako bi on ponovno obuhvatio objekt. Problem mean-shift algoritma je što se jednom zadan omeđujući pravokutnik objekta ne može mijenjati po veličini i orijentaciji. U slučaju da se objekt u videozapisu udalji od kamere, takvo praćenje neće rezultirati najužom regijom koja obuhvaća objekt. Dodatno, mean-shift se vodi pretpostavkom da se boje i tekstura objekta neće previše mijenjati. Zbog takve pretpostavke algoritam najbolje radi u kontroliranim uvjetima gdje nema velikih promjena u osvjetljenju, gdje je tekstura objekta konstanta i sl.

CAMshift je nadogradnja mean-shifta koja rješava problem konstantne veličine prozora koji sadrži objekt i nemogućnost rotacije prozora. Algoritam prvo primjenjuje mean-shift, a zatim mijenja veličinu prozora prema formuli  $s = 2 \times \sqrt{\frac{M_{00}}{256}}$  gdje je  $M_{00}$  nulti moment dobiven pomoću mean-shift algoritma. Dodatno iz prvog i drugog momenta dobivenog mean-shift algoritmom se računaju duljine osi i orijentacija elipse pomoću koje se prozor rotira.

## 3.2. Metode bazirane na segmentaciji

Praćenje objekata koristi razdvajanje pozadine od prednjeg dijela slike, odnosno razdvajaju se objekti u pokretu na slici i statična pozadina. Metode u ovom pristupu moraju biti robusne na promjene u osvjetljenju i u kretanju. [12]

### 3.3. Metode baziranje na procjeni funkcije gustoće vjerojatnosti

Praćenje objekata se formulira kao problem procjene stanja u kojem se jedan objekt prikazuje vektorom stanja. Navedeni vektor opisuje parametre kretanja objekta poput trenutne pozicije, brzine, ubrzanja i sl. Najpoznatije metode praćenja objekata bazirane na procjeni stanja su Kalmanov filtar i čestični filtar (engl. Particle filter)

#### 3.3.1. Kalmanov filtar

Kalmanov filtar se koristi za procjenu trenutnog stanja linearnog dinamičkog sustava na temelju prijašnjeg stanja i trenutnih mjerenja. Model procesa se opisuje formulom:

$$\vec{x}_k = F\vec{x}_{k-1} + B\vec{u}_{k-1} + \vec{w}_{k-1}$$

gdje je  $F$  tranzicijska matrica,  $B$  je kontrolna ulazna matrica i  $\vec{w}$  je vektor šuma koji se ravna po Gaussovoj distribuciji. Model mjerenja opisuje vezu između stanja i mjerenja i opisan je formulom:

$$\vec{z}_k = H\vec{x}_k + \vec{v}_k$$

gdje je  $\vec{z}_k$  vektor mjerenja,  $H$  je matrica mjerenja i  $\vec{v}_k$  je vektor šuma u mjerenju. Prvi korak Kalmanovog filtra je izrada a priori procjene trenutnog stanja i izrada kovarijacijske matrice greške.

$$\vec{x}_k^- = A\vec{x}_{k-1} + B\vec{u}_{k-1}$$

$$P_k^- = AP_{k-1}A^T + Q$$

U drugom koraku se provodi korekcija procjene na temelju mjerenja. U tom koraku se izračunava Kalmanov faktor miješanja (engl. Kalman gain) i pomoću njega se ažuriraju procjene stanja i kovarijacijska matrica greške. U kontekstu praćenja objekta Kalmanov filtar ima različite primjene te se najčešće koristi za praćenje točaka u slikama koje sadrže šum, predviđa se pozicija i brzina objekta i sl.

#### 3.3.2. Čestični filtar

Čestični filtri ili sekvencijalne Monte Carlo metode su skup metoda koje se koriste za rješavanja problema filtriranja u obradi signala i Bayesovom statističkom zaključivanju. Za razliku od Kalmanovog filtra, filteri čestica ne pretpostavljaju linearnost niti da su varijable stanja normalno distribuirane.

### 3.4. Metode baziranje na strojnom učenju

Praćenje objekata nastoji se ostvariti primjenom strojnog učenja. Strojno učenje može se koristiti za ekstrakciju ključnih značajki, učenje izgleda objekta i predviđanje kretanja objekta. Soleimanitaleb et al.[11] navode 3 vrste metoda praćenja objekata baziranih na učenju: diskriminativne, generativne i metode podržanog učenja(engl. reinforced learning). Za praćenje se najčešće koriste diskriminativni modeli bazirani na dubokom učenju te ta vrsta modela daje jako dobre rezultate u praćenju objekata.

#### 3.4.1. Praćenje objekata pomoću sijamskih neuronskih mreža

Sijamske neuronske mreže su neuronske mreže koje imaju dvije ili više identičnih pod-mreža koje se koriste za generiranje vektora značajki za svaki od ulaza. Sve podmreže koriste jednake težine i parametre. Dobiveni vektori značajki se potom uspoređuju na temelju njihove sličnosti. Neki od poznatijih algoritama praćenja objekata baziranih na sijamskim neuronskim mrežama su: SiamFC, SiamRPN, SiamRPN++.

Navedene mreže kao ulaz primaju cijelu sliku na jednom ulazu i sliku objekta koji se nastoji naći na slici na drugom ulazu. Tako se smanjuje potreba ispitivanjem svake podslike u slici s ciljnim objektom. Kao izlaz se vraća dvodimenzijski vektor sličnosti u kojem veće vrijednosti označavaju visoku sličnost s objektom za korespondentnu regiju, dok niske vrijednosti označavaju slabu sličnost regije s ciljnim objektom[1].

#### 3.4.2. DeepSORT

SORT (engl. simple online and realtime tracking) je metoda praćenja više objekata u slici. Metoda koristi samo prethodni i trenutni okvir za praćenje objekata. Za procjenu kretanja i asocijaciju podataka koristi se samo trenutna pozicija i veličina objekta te značajke izgleda dobivene detektorom objekata. Metoda se sastoji od detekcije, asocijacije trenutnih detekcija sa postojećim objektima i upravljanja životnim ciklusom praćenjih objekata. Za detekciju objekata koristi se FrRCNN(VGG16) konvolucijska neuronska mreža. Za predviđanje procjenu kretanja objekta i asociiranje podataka koriste se Kalmanov filter i Mađarska metoda. SORT zanemaruje probleme privremenog i potpunog zaklanjanja objekata s ciljem smanjivanja kompleksnosti mreže[2].

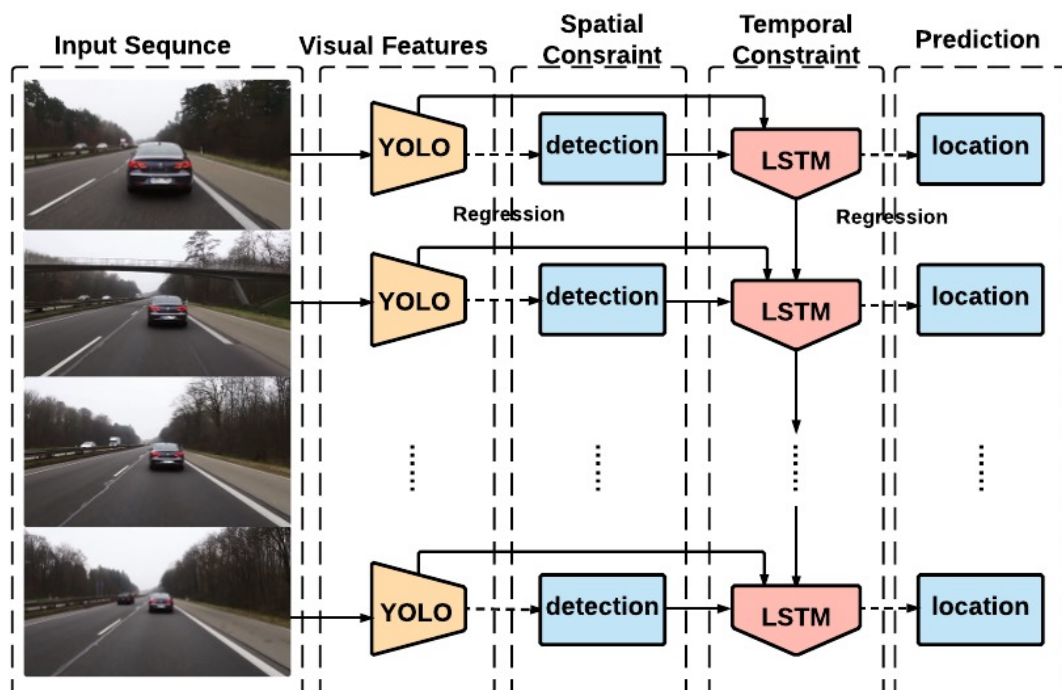
DeepSORT je proširenje SORT (engl. Simple online and Realtime Tracking) metode za praćenje objekata. Metoda SORT se proširuje tako da koristi značajke koje opisuju izgled objekta za praćenje objekta. Proširenje se postiže primjenom konvolucijske neuronske mreže koja je trenirana da razlikuje pješake na slikama. Prednost

takvog proširenja je da se ublažava problem zamjene identiteta objekta te se objekti uspijevaju duže pratiti.[13]

### 3.4.3. ROLO

ROLO je konvolucijska neuronska mreža koja se nadograđuje na YOLO detektor objekata. Arhitektura ROLO mreže se sastoji od YOLO detektora i LSTM(engl. long short-term memory) povratne mreže. Detektor objekata YOLO se koristi za dobivanje preliminarne prostorne informacije o objektima i za dobivanje robusnih značajki visoke razine koje opisuju objekt te za detekciju novih i postojećih objekata. LSTM se koristi kao modul za praćenje objekta koji koristi prethodno naučene kontekstualne informacije i novodobivene prostorne informacije o objektima za precizno praćenje objekata. LSTM kao kontekstualne informacije pamti prostorne i vizualne značajke visoke razine objekta, odnosno značajke dobivene prolaskom ulaznog niza kroz konvolucijski dio mreže[9]. Princip rada ROLO mreže koji opisuje ovakav tok praćenja objekata ilustriran je slikom 3.1.

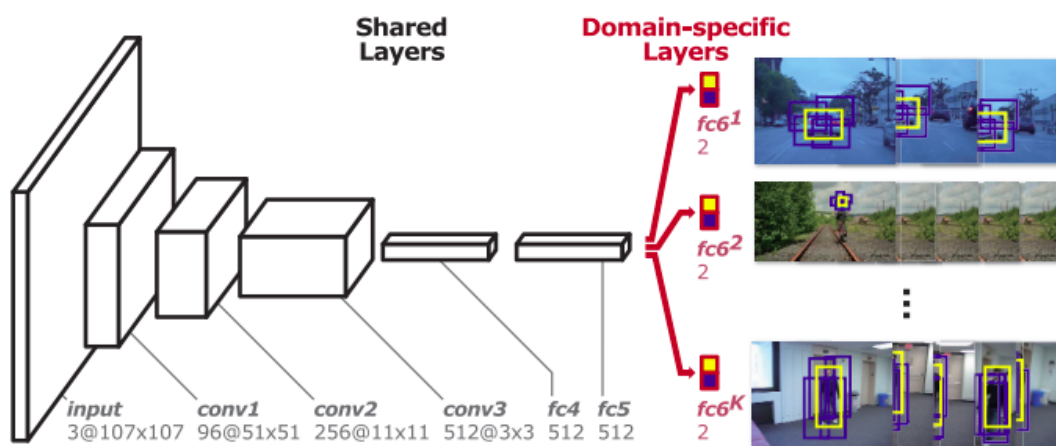
Mreža za praćenje objekata ROLO pokazuje vrlo dobre rezultate u situacijama kada dolazi do zaklanjanja objekta gdje uspijeva pratiti objekt i kada detekcija objekta od strane YOLO modula zakaže. Dodatno mreža ROLO uspijeva preciznije odrediti lokaciju objekta prilikom efekta zamućivanja objekta uzrokovanog kretanjem(engl. motion blur) u odnosu na lokaciju koju određuje isključivo YOLO detektor. Razlog tome je što LSTM pri zaklanjanju iskorištava prethodno znanje o lokaciji i vizualnim karakteristikama objekta za određivanje trenutne lokacije djelomično ili potpuno zaklonjenog objekta[9].



Slika 3.1: Princip rada ROLO mreže[9]

#### 3.4.4. MDNet

MDNet je konvolucijska neuronska mreža koja se koristi za praćenje jednog objekta (engl. single object tracker). Arhitektura MDNet mreže prikazana na slici 3.2 sastoji se od dijela za učenje domenski neovisnih značajki te se taj dio koristi za razlikovanje pozadine od objekata, dok je drugi dio domenski specifičan, odnosno za svaku domenu postoji zasebna grana neuronske mreže koja sadrži domenski specifične slojeve mreže. Tijekom učenja svaki od tih slojeva prilagođava težine u svom i u dijeljenom dijelu arhitekture te tako domenski neovisan dio mreže uči ekstrahirati dobre karakteristike bilo kakvih objekata. Jednom kada je treniranje gotovo, domenski specifični slojevi se uklanjaju i dodaje se sloj neuronske mreže koji se uči pratiti neki objekt tijekom izvođenja, odnosno težine domenski specifičnog sloja se prilagođavaju tijekom izvođenja (engl. on-line)[8].



**Slika 3.2:** Arhitektura MDNet mreže za praćenje objekata[8]

## 4. Zaključak

Praćenje objekata je široko područje u konstantnom razvoju te uživa široku primjenu u područjima poput sigurnosnog nadzora, navigacije vozila, poljoprivrede, marikulturi, robotici i sl. Tijekom relativno kratke povijesti razvijene su brojne metode praćenja objekata koje iskorištavaju značajke koje opisuju objekte poput boje, teksture, oblika, ključnih točaka i sl.

Prednost praćenja objekata u odnosu na isključivu primjenu detekcije objekata je ta da se može analizirati i predviđati kretanje objekata te je lokalizacija objekta robusnija na zaklanjanja i smetnje u slici.

Sustavi za praćenje objekata su rijetko kada jedina komponenta sustava te se često uz praćenje objekata koriste sustavi za detekciju objekata o kojima veliki broj algoritama za praćenje objekata uvelike ovisi i sustavi za modeliranje reprezentacije objekata. Praćenje objekata otežavaju razne prirodne pojave poput promjene osvjetljenja, pozadine, interakcije objekata i neočekivanih promjena u kretanju objekata.

Za praćenje objekata razvijene su metode koje ne koriste strojno učenje za praćenje objekata već isključivo baziraju svoja predviđanja putanja na temelju karakteristika objekta i slike. Neke od tih metoda su metode bazirane na optičkom toku, Kalmanovim filterima i filterima čestica te na pomaku srednje vrijednosti (engl. mean-shift).

S porastom popularnosti dubokog učenja, sve se češće koriste metode za praćenje jednog ili više objekata bazirane na dubokom učenju pri čemu je praćenje više objekata u jednom okviru puno izazovniji zadatak. Duboko učenje u praćenju objekata može imati razne primjene poput detekcije objekata koji se prate, ekstrakiranje značajki objekta, predviđanje kretanja, određivanje sličnosti i sl.[4] Mreže DeepSORT, ROLO, MDNet i GOTURN samo su neke od brojnih dubokih konvolucijskih neuronskih mreža koje se primjenjuju za praćenje objekata.



## 5. Literatura

- [1] Luca Bertinetto, Jack Valmadre, João F. Henriques, Andrea Vedaldi, i Philip H. S. Torr. Fully-convolutional siamese networks for object tracking, 2021.
- [2] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, i Ben Upcroft. Simple online and realtime tracking. *2016 IEEE International Conference on Image Processing (ICIP)*, Sep 2016. doi: 10.1109/icip.2016.7533003. URL <http://dx.doi.org/10.1109/ICIP.2016.7533003>.
- [3] Erik Bochinski, Volker Eiselein, i Thomas Sikora. High-speed tracking-by-detection without using image information. U *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, stranice 1–6, 2017. doi: 10.1109/AVSS.2017.8078516.
- [4] Gioele Ciaparrone, Francisco Luque Sánchez, Siham Tabik, Luigi Troiano, Roberto Tagliaferri, i Francisco Herrera. Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 381:61–88, Mar 2020. ISSN 0925-2312. doi: 10.1016/j.neucom.2019.11.023. URL <http://dx.doi.org/10.1016/j.neucom.2019.11.023>.
- [5] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. svezak 2749, stranice 363–370, 06 2003. ISBN 978-3-540-40601-3. doi: 10.1007/3-540-45103-X\_50.
- [6] Matej Kristan, Jiri Matas, Ale Leonardis, Michael Felsberg, Luka Cehovin, Gustavo Fernandez, Toma Vojir, Gustav Hager, Georg Nebehay, Roman Pflugfelder, Abhinav Gupta, Adel Bibi, Alan Lukezic, Alvaro Garcia-Martin, Amir Saffari, Alfredo Petrosino, i Andrés Solis Montero. The visual object tracking vot2015 challenge results. U *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, stranice 564–586, 2015. doi: 10.1109/ICCVW.2015.79.

- [7] Matej Kristan, Ales Leonardis, Jiri Matas, Michael Felsberg, Roman Pfugfelder, Luka Čehovin Zajc, Tomas Vojir, Goutam Bhat, Alan Lukezic, Abdelrahman Eldesokey, Gustavo Fernandez, i et al. The sixth visual object tracking vot2018 challenge results, 2018.
- [8] Hyeonseob Nam i Bohyung Han. Learning multi-domain convolutional neural networks for visual tracking, 2016.
- [9] Guanghan Ning, Zhi Zhang, Chen Huang, Zhihai He, Xiaobo Ren, i Haohong Wang. Spatially supervised recurrent convolutional neural networks for visual object tracking, 2016.
- [10] Raul Rojas. Lucas-kanade in a nutshell, 2021.
- [11] Zahra Soleimanitaleb, Mohammad Ali Keyvanrad, i Ali Jafari. Object tracking methods:a review. U *2019 9th International Conference on Computer and Knowledge Engineering (ICCCKE)*, stranice 282–288, 2019. doi: 10.1109/ICCCKE48569.2019.8964761.
- [12] Rachna Verma. A review of object detection and tracking methods. *International Journal of Advance Engineering and Research Development*, 4:569–578, 10 2017.
- [13] Nicolai Wojke, Alex Bewley, i Dietrich Paulus. Simple online and realtime tracking with a deep association metric, 2017.
- [14] Alper Yilmaz, Omar Javed, i Mubarak Shah. Object tracking: A survey. *ACM Comput. Surv.*, 38(4):13–es, dec 2006. ISSN 0360-0300. doi: 10.1145/1177352.1177355. URL <https://doi.org/10.1145/1177352.1177355>.