

# Modeliranje kontekstualno zavisne vizuelne reprezentacije nepoznatog objekta u realnom vremenu

Inspirisano Monte-Carlo metodama

Igor Perić

Student prvog ciklusa studija  
Slobomir P Univerzitet  
Bijeljina, Bosna I Hercegovina  
igorperic@live.com

Igor Tomić

Student prvog ciklusa studija  
Fakultet tehničkih nauka  
Novi Sad, Srbija  
teslabn@gmail.com

*Sadržaj*—Problem određivanja ključnih karakteristika prilikom učenja parametara vizuelnog modela nekog nepoznatog objekta u nekoj video sekvenci se često susreće u oblasti procesiranja slike. Postoji veliki broj algoritama iz oblasti Mašinskog učenja koji se oslanjaju na numeričke i statističke metode modelovanja ključnih obilježja slike (features) nekog objekta. U ovom radu predstavljen je algoritam za efikasno određivanje dovoljnog broja i vrijednosti parametara numeričkog modela koji nastoji da dovoljno dobro opiše najskorije pojavljivanje objekta u video sekvenci. Algoritam je tolerantan na relativno male promjene u izgledu objekta i, kao takav, omogućava postepenu tranziciju trenutno poznatog modela objekta ka domenu idealnog modela. Akcenat rada je na principima učenja i postepenog korigovanja modela u realnom vremenu.

*Ključne riječi* – mašinsko učenje; numerički model; procesiranje slike;

## I. UVOD

Problem prepoznavanja, detekcije i praćenja objekata predstavlja jedan od osnovnih problema u oblasti procesiranja slike. Elementarni principi velikog broja predloženih rješenja nalažu postojanje dvije faze:

1. Faza učenja izgleda objekta - „treniranje“
2. Faza detekcije objekta na osnovu modela

Prva faza podrazumijeva odabir numeričkog ili statističkog modela, koji je zapravo funkcija sa određenim brojem parametara te, u zavisnosti od prirode modela, vrijednost ove funkcije predstavlja ili tačne vrijednosti ili raspodjele vrijednosti zavisne od skupa ulaznih veličina. Odabirom nekog skupa vrijednosti za parametre vrši se odabir jedne funkcije iz skupa svih funkcija određenih parametrima. Za odabranu funkciju se može reći da ona opisuje skup ulaznih veličina na dovoljno dobar način a proces pronalazjenja odgovarajućih vrijednosti parametara modela za zadane vrijednosti ulaznih veličina naziva se **treniranje modela** ili **učenje**[1], a skup ulaznih veličina **skup podataka za treniranje**.

Kao rezultat učenja modela za detekciju objekata nad bilo kojim skupom slika koje su različite vizuelne reprezentacije istog objekta dobijamo model tog objekata u vidu vrijednosti za skup parametara odabranog numeričkog modela, koji je

upotrebljiv za procjenu detekcije postojanja objekta za proizvoljnu sliku datu kao (do sad nepoznat) ulaz.

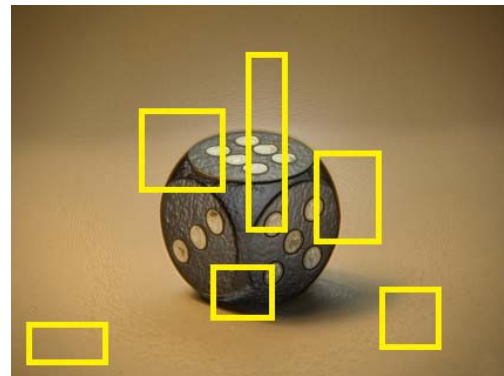
Učenje parametara modela ne mora da bude proces nezavistan od detekcije i prepoznavanja. U ovom radu ispitana je mogućnost paralelnog učenja izgleda objekta u toku njegove detekcije i upotreba rezultata prepoznavanja za poboljšanje procesa učenja.

## II. KLJUČNE KARAKTERISTIKE SLIKE I NJIHOVE REPREZENTACIJE

Da bi se vizuelna reprezentacija objekta (slika) mogla uvrstiti u proračune neophodno je posmatrati je kao dvodimenzionalnu matricu  $X^{M \times N}$  uređenih parova  $(R,G,B)_{i,j}$  koji predstavljaju udio pojedinačnih kanala boje u rezultujućoj boji svakog piksela, pri čemu su  $M$  i  $N$  širina i visina slike u pikselima, respektivno, a vrijednosti  $i$  i  $j$  redni broj reda i kolone posmatranog piksela. Skup slika na osnovu kojih se vrši treniranje modela označen je kao:

$$X = \{ X_i | i = 0, 1, \dots, K \}$$

pri čemu je  $K$  ukupan broj slika u skupu podataka za treniranje.



Slika 1. Primjer vizuelne reprezentacije objekta sa 6 proizvoljnih markera

Za svaku sliku  $X_i$  od  $K$  slika iz skupa podataka za treniranje poznat je skup  $M_i$  od  $\alpha$  pravougaonih oblasti dimenzija  $(W_i, H_i)$  na lokacijama  $(R_i, C_i)$ , za koje važi:

$$0 \leq W_i \leq N, i = 0, 1, \dots, \alpha$$

$$0 \leq H_i \leq M, i = 0, 1, \dots, \alpha$$

$$0 \leq R_i + H_i \leq M, i = 0, 1, \dots, \alpha$$

$$0 \leq C_i + W_i \leq M, i = 0, 1, \dots, \alpha$$

Ove pravougaone oblasti koje se, kao što se iz nejednakosti iznad može zaključiti, nalaze cijelom svojom površinom unutar slike objekta zvaćemo **markeri**. Svrha postojanja markera je da opišu region vizuelne reprezentacije objekta koji je od značaja za njegovo prepoznavanje[2].

Svaki  $M_i$  od ukupno  $\alpha$  markera za proizvoljnu sliku  $Z$  dimenzija  $M \times N$  može da se posmatra kao funkcija te slike koja preslikava dvodimenzionalnu matricu uređenih parova  $Z_{i,j}(R,G,B)$  u jedan uređen par  $(R,G,B)$ . Relacija  $f$  koja opisuje prirodu ove zavisnosti zapisuje se na sledeći način:

$$f(M_i, Z): R^{M \times N \times 3} \rightarrow R^3$$

Ova funkcija, koja se može smatrati **funkcijom vrijednosti određenog markera**, data je izrazom:

$$f(M_i, Z) = \sum_{p=R_i}^{R_i+H_i} \sum_{q=C_i}^{C_i+W_i} Z_{p,q}$$

Može se zaključiti da za svaku sliku  $Z$  svaki marker  $M_i$  karakterišu tri numeričke vrijednosti (za svaki kanal boje pojedinačno) koje predstavljaju sumu vrijednosti svih piksela slike  $Z$  koji se nalaze unutar pravougane oblasti koju ograničava marker  $M_i$ .

### III. STRUKTURA PODATAKA ZA REPREZENTACIJU SLIKE

S obzirom da se algoritam oslanja samo na informacije o boji piksela mali broj markera nije u mogućnosti da sa dovoljnom preciznošću obuhvati sve bitne vizualne karakteristike objekta. Iz tog razloga pribjegava se povećanju vrijednosti  $\alpha$ , odnosno za objekat koji se želi modelovati određuje se relativno veliki broj markera različitih dimenzija  $(W_i, H_i)$  i pozicija  $(R_i, C_i)$ . Složenost funkcije vrijednosti markera je  $O(N^2)$  što bi, za veliko  $\alpha$ , moglo predstavljati problem kada su ukupne performanse u pitanju. U svrhu održanja ukunih performansi algoritma na dozvoljenom nivou uprkos ovom povećanju broja markera utićemo na složenost izračunavanja funkcije vrijednosti markera odabirom adekvatne strukture podataka: kumulativne tabele[3].

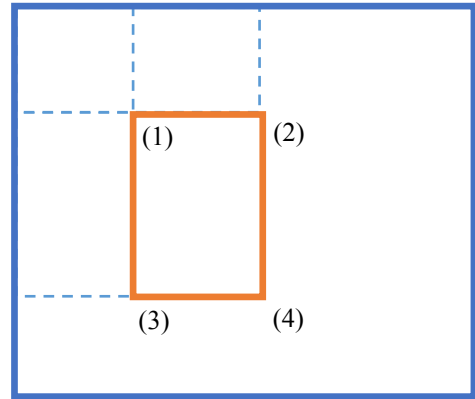
Definisaćemo matricu uređenih trojki  $Q^{M \times N}$  za proizvoljnu sliku  $Z$  na sljedeći način:

$$Q_{i,j} = \sum_{p=0}^i \sum_{q=0}^j Z_{p,q}$$

Nad ovako definisanom matricom, koja je funkcija slike  $Z$ , računanje vrijednosti proizvoljnog markera  $M_i$  je složenosti  $O(1)$  ukoliko se koristi sledeća metoda za njeno izračunavanje:

$$f(M_i, Q) = Q_{R_i+H_i, C_i+W_i} - Q_{R_i, C_i+W_i} - Q_{R_i+H_i, C_i} + Q_{R_i, C_i}$$

Umjesto  $H_i \times W_i$  operacija ova metoda pruža isti rezultat u samo 4 operacije. Vrijedi pomenuti da je, za razliku od rješenja sa kvadratnom složenošću, potrebno utrošiti dodatno vrijeme za pripremu matrice  $Q$  ali i da je to vrijeme zanemarivo u odnosu na porast u performansama prilikom znatno većeg broja izračunavanja vrijednosti markera.



Slika 2. Ilustracija izračunavanja vrijednosti markera upotrebom kumulativne tabele  $Q$

Slika 2 ilustruje princip izračunavanja vrijednosti markera. Svaki element  $Q_{i,j}$  u matrici predstavlja sumu svih elemenata od gornjeg lijevog elementa  $(Q_{0,0})$  do posmatranog elementa  $Q_{i,j}$ . S obzirom da je vrijednost markera ništa drugo nego suma svih piksela obuhvaćenih markerom, sa slike se može zaključiti da se izraz za  $f(M_i, Q)$  može napisati kao:

$$f(M_i, Q) = (4) - (2) - (3) + (1)$$

Ova metoda inspirisana principima prvi put predstavljenim 2001. godine u radu autora Viola-Jones[4] koji su se fokusirali na primjenu ove tehnike na detekciju lica.

### IV. ELEMENTARNI PRINCIPI ALGORITMA UČENJA

U nastavku rada će biti elaborisane metode učenja izgleda samo jednog objekta uz napomenu da se isti principi mogu paralelno primjenjivati i na proizvoljan broj objekata čiji se izgled istovremeno modeluje.

Svijest o objektu čiji je izgled neophodno naučiti u svakom trenutku određen je neuređenim parom  $(M, X)$ , gdje je:

- $M$  -  $\alpha$ -dimenzioni vektor markera
- $X$  -  $K$ -dimenzioni vektor fotografija modela

Numerička vrijednost  $\alpha$  je, zapravo, **parametar algoritma** i kvantitativno određuje odnos otpornosti modela na greške i tačnosti modela. Ovaj zaključak će se potkrijepiti u nastavku rada detaljnim obrazloženjem algoritma.

**K je**, takođe, **parametar algoritma** i zapravo predstavlja količinu privremene memorije (izraženu u broju vizuelnih

reprezentacija objekta) koji će algoritam uzimati u obzir prilikom uticaja na parametre modela. Veće vrijednosti parametra K omogućavaju svrstavanje u model onih pojavljivanja objekta koji su se dogodili ranije u prošlosti (u daljoj tački na vremenskoj osi). S druge strane, male vrijednosti parametra K omogućavaju kratkoročniju memoriju i, efektivno, veću toleranciju na brže promjene izgleda objekta.

Postupak inicijalizacije se sastoji iz dva jako jednostavna koraka:

$$X = \emptyset$$

$$M = \emptyset$$

što simbolizuje činjenicu da objekat čiji je izgled potrebno modelovati nije odabran.

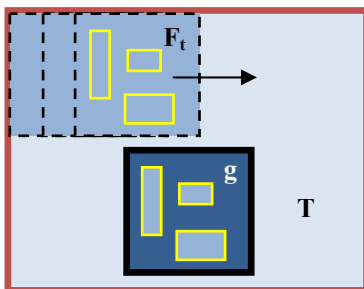
U radu je diskutovan samo algoritam učenja a ne detekcije objekta te se, imajući to u vidu, može reći da se prilikom gubljenja objekta iz vida svijest o izgledu objekta smatra uništenom, odnosno poziva se postupak inicijalizacije opisan iznad.

Odabir objekta za praćenje podrazumijeva određivanje (selekciju) pravougaonog regiona dimenzija MxN direktno na prikazu snimka sa kamere. Naznačeni region postaje X<sub>0</sub>, prvi u nizu grafičkih reprezentacija objekata.

Istovremeno se generiše α nasumičnih markera unutar selektovanog regiona i izračunavaju se vrijednosti tih markera. U trenutku t = 0 ove vrijednosti, zajedno sa grafičkom reprezentacijom X<sub>0</sub>, predstavljaju kompletnu svijest sistema o izgledu objekta. Daljim radom algoritma kardinalnost skupa X će postepeno da raste a model g(X, M) će u svakom trenutku biti definisan skupom srednjih vrijednosti pojedinačnih markera nad svim slikama iz X.

$$g(X, M) = \left\{ \frac{1}{K} \sum_{j=0}^K f(M_i, X_j) \mid i = 0, 1, \dots, \alpha \right\}$$

Svaka naredna slika u video sekvenci se analizira takozvanom „metodom ključućeg prozora“, regionom dimenzija MxN koji se postepeno kreće po slici u potrazi za regionom karakteristika sličnih modelu.



Slika 3. Prikaz analize frejma ključućim prozorom

Uzmimo da je prilikom jednog pomjeraja prozor opisao mali dio trenutnog frejma T koji se označava sa F<sub>t</sub>. Ako se

svakom markeru M<sub>i</sub> pridruži realan broj P<sub>i</sub> takav da je P<sub>i</sub> = W<sub>i</sub> × H<sub>i</sub>, tada možemo izraziti **težinski koeficijent** E<sub>i</sub> svakog markera kao:

$$E_i = \frac{P_i}{W \times H}$$

Uloga težinskog koeficijenta je da kvantitavno odredi uticaj određenog markera u zavisnosti od njegove relativne površine. Uz ovako definisane veličine za frejm F<sub>t</sub> možemo izračunati prosječnu mjeru sličnosti δ<sub>t</sub>(F<sub>t</sub>) sa dosadašnjim modelom g:

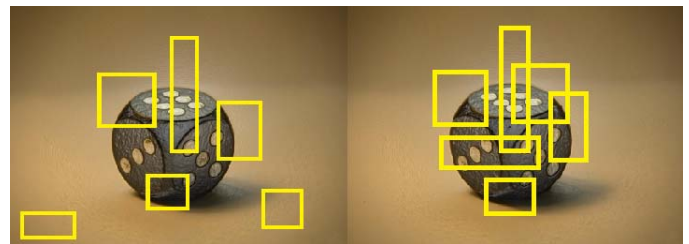
$$\delta_t(F_t) = \frac{f(M, F_t)}{g(X, M)} = \frac{1}{\alpha} \sum_{i=0}^{\alpha} \left( \frac{f(M_i, F_t)}{g(X, M_i)} \times E_i \right)$$

Izračunata mjera sličnosti je realan broj iz intervala [0,1]. Binarnim klasifikatorom:

$$\Delta(F_t | \delta) = \begin{cases} 1, & \delta_t(F_t) \geq \delta \\ 0, & \delta_t(F_t) < \delta \end{cases}$$

se donosi odluka o tome da li frejm F<sub>t</sub> vizuelna reprezentacija objekta koja odgovara dosadašnjem modelu ili nije. **δ je treći parametar algoritma** čijim se izborom direktno utiče na nivo detalja na koje će algoritam obraćati pažnju. Izborom manjih vrijednosti za δ model će dozvoljavati značajnije varijacije u izgledu objekata koje će smatrati istim.

Za svaki F<sub>t</sub> za koji važi Δ(F<sub>t</sub> | δ) = 1 računa se Euklidska udaljenost od poslednje poznate lokacije modela g u trenutku t - 1. Ukoliko je ta udaljenost veća od **realnog broja D**, čija se vrijednost zadaje u vidu **četvrtog parametra algoritma**, potrebno je najlošije odabran marker u modelu odbaciti i izabrati novi nasumični marker. Ovaj postupak osigurava jako dobar izbor markera sa porastom vremena izvršavanja



Slika 4. Postepena automatska konvergencija početnog skupa (lijevo) ka skupu jako dobro odabranih markera (desno)

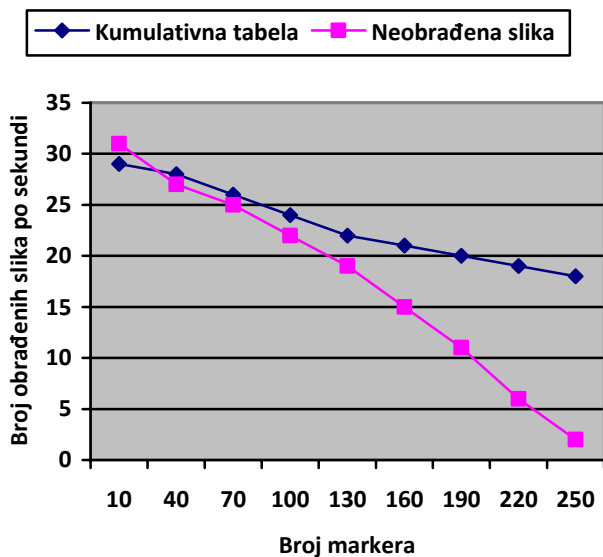
algoritma jer će skup nasumično odabranih markera postepeno konvergirati ka skupu kvalitetno odabranih markera, koji će doprinijeti smanjenju takvih F<sub>t</sub> koji će biti prepoznati kao objekti za koje se smatra da su reprezentacije objekta čiji se izgled uči a da su, pri tome, te reprezentacije previše udaljene od prošle poznate tačne lokacije objekta.

U slučaju da je Euklidska udaljenost  $F_t$  manja od parametra  $D$ ,  $F_t$  se upisuje u  $X$  u FIFO maniru (First In First Out) a najstariji (najduže prisutan u skupu)  $X_{old}$  se iz skupa  $X$  izbacuje ukoliko je dodavanjem poslednjeg  $F_t$  kardinalnost skupa  $X$  postala veća od  $K$ . Preciznije, prepoznavanje nove reprezentacije objekta u neposrednoj blizini prethodno poznate može se opisati sledećom jednakošću:

$$X = \begin{cases} X \cup F_t & |X| = K - 1 \\ (X \cup F_t) / X_{old} & |X| = K \end{cases}$$

Parametar  $D$  direktno utiče na dozvoljenu udaljenost lokacija objekta između dva uzastopna frejma te se, kao takav, koristi da ograniči dozvoljenu brzinu kretanja objekta nakon koje se smatra da je objekat izgubljen iz vidokruga.

Zavisnost broja obrađenih slika po jedinici vremena (sekundi) od broja markera data je na Slici 5. Sa grafikona se jasno može vidjeti da su na tek na velikom broju markera kumulativne tabele pokazale poboljšanje u performansama, odnosno omogućile izbor relativno velike vrijednosti za parametar  $\alpha$  bez toliko drastičnog pada u performansama.



Slika 5. Grafički prikaz zavisnosti performansi od broja upotrebljenih markera

## V. ZAKLJUČAK

Opisan algoritam na dovoljno dobar način i sa zadovoljavajućim performansama omogućava određivanje onih regiona slike koji dovoljno dobro opisuju najskorije pojavljivanje željenog objekta s akcentom na uslove u kojima se on pojavljuje (okruženje). Predloženi algoritam je tolerantan na promjene u izgledu objekta i, u skladu sa odabranim parametrima, konstantno bira i uči nova karakteristična obilježja izmjenjenog objekta ili izmjenjene okoline. Upotrebljiv je u situacijama u kojima je potrebno za relativno kratko vrijeme formirati model objekta čiji se izgled ne može naslutiti. Pogodan je za privremeno i kratkoročno čuvanje znanja o izgledu i, kao takav, primjenu može naći u sistemima sa visokim nivoom interaktivnosti.

## LITERATURA

- [1] U. V. Luxburg, B. Scholkopf, "Statistical Learning Theory: Models, Concepts, and Results" (neobjavljen)
- [2] [http://en.wikipedia.org/wiki/Feature\\_detection\\_\(computer\\_vision\)](http://en.wikipedia.org/wiki/Feature_detection_(computer_vision))
- [3] [http://en.wikipedia.org/wiki/Summed\\_area\\_table](http://en.wikipedia.org/wiki/Summed_area_table)
- [4] P. Viola, M. Jones, "Robust Real-time Object Detection", International Journal of Computer Vision

## ABSTRACT

Problem of determination of key features for learning visual appearance model of a unknown object in certain video sequence is very common in image processing. There is a number of algorithms in field of Machine learning which rely on numerical and statistical methods of modeling key image features of an object. In this paper we presented an algorithm for efficient determination of large number and values of numerical model parameters which tend to sufficiently well describe the most recent appearance of object in a video sequence. Algorithm is resistant to relatively small changes in appearance and, as such, enables gradual transition of currently known model of object to domain of ideal model. Work in this paper is focused on principles of learning and gradual corrections of model in real time.

## Modeling of contextually dependent visual representation of unknown object in real time Inspired by Monte-Carlo methods

Igor Peric, Igor Tomic