

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

PROJEKT IZ RASPOZNAVANJA UZORAKA
kontrolna točka

**Detekcija pješaka u urbanim
okruženjima korištenjem značajki
temeljenih na teksturi i boji**

Iva Miholić, Gustav Matula, Kristijan Franković, Tomislav Kiš

Zagreb, siječanj 2015.

SADRŽAJ

| | |
|---|----------|
| 1. Detekcija pješaka u urbanim okruženjima | 1 |
| 1.1. Opis projektnog zadatka | 1 |
| 1.2. Pregled i opis postojećih rješenja | 1 |
| 1.3. Konceptualno rješenje zadatka | 3 |
| 1.3.1. Pregled značajki temeljenih na teksturi i boji | 3 |
| 1.3.2. Redukcija dimenzije | 4 |
| 1.3.3. Plan arhitekture sustava računalnog vida | 5 |
| 2. Postupak rješavanja zadatka | 6 |
| 2.1. Ekstrakcija značajki | 6 |
| 2.2. Treniranje prve faze klasifikatora | 6 |
| 2.3. Treniranje druge faze klasifikatora | 6 |
| 3. Ispitivanje rješenja | 7 |
| 3.1. Ispitna baza | 7 |
| 3.2. Rezultati učenja i ispitivanja | 8 |
| 3.3. Analiza rezultata | 8 |
| 4. Literatura | 9 |

1. Detekcija pješaka u urbanim okruženjima

1.1. Opis projektnog zadatka

Detekcija pješaka u urbanim okruženjima problem je povezan sa automobilskom industrijom, sigurnošću u prometu, sigurnosnim nadziranjem i robotikom. Rješava se kao detekcija objekta u okviru područja računalnog vida. Ovaj projektni zadatak obuhvaća izgradnju detektora pješaka na fotografijama iz urbanih okruženja korištenjem značajki temeljenih na bridovima, teksturi i boji.

1.2. Pregled i opis postojećih rješenja

Pregled najznačajnijih rješenja ovog problema dan je u [1]. Trenutačno postoje dva glavna pristupa tom problemu. Prva klasa metoda obuhvaća detektore za pojedine dijelove tijela koji se zatim kombiniraju u detektor čvojeke - pješaka dok druga klasa prihvaća problemu statističkom analizom kombinirajući mnogobrojne značajke unutar detekcijskog prozora u klasifikator.

Prvi značajniji napredak u detekciji pješaka bila je primjena *VJ* detektora objekata [7] na ovaj problem. Detektori temeljeni na histogramu usmjerenih gradijenata (engl. *Histogram of Oriented Gradients, HOG*) [2] uz linearni ili nelinearni skup potpornih vektora, postigli su značajne rezultate. No, primjena samo HOG značajki vodi do broja pogrešno pozitivno klasificiranih primjera čime se može stati na kraj uvodeći u sustav značajke temeljene na svojstvima boje, tekstura i oblika.

Više o HOG pristupu bit će riječi u sljedećim poglavljima jer ćemo ga koristiti kao kostur našeg prostora značajki.

Od ostalih rezultata, potrebno je izdvojiti postupke temeljene na modelu rastavljivih dijelova (engl. *Deformable Part Models, DPM*) u kojem se detekcija



Slika 1.1: Fotografija koja je klasificirana detektorom pješaka metodom kliznog prozora. Žuti okviri prikazuju okvire onih prozora koji su klasificirani kao prikaz pješaka.

dijelova tijela sažima u detekciju cijelog pješaka te nelinearne postupke učenja temeljenih na neuronskim mrežama i stabilima odluke [1]. Takvi složeniji postupci uspoređeni su sa linearnim SVM-om uz HOG i druge značajke nisu dali značajno bolji rezultat upućujući na korektnost naše odluke o arhitekturi detektora.

Pristup na koji ćemo se fokusirati koristi metodu skalabilnog kliznog prozora. Unutar prozora na fotografiji određuju se značajke te se skup piksela unutar prozora binarno klasificira kao fotografija pješaka. Prozor zatim "putuje" po fotografiji testirajući druge skupove piksela. Slika se onda može dodatno skalirati (efektivno skalirajući prozor) nakon čega se ponavlja isti postupak. Primjer klasificirane fotografije ovakvim postupkom vidljiv je na slici 1.1. Problem takvog pristupa je ignoriranje konteksta oko okvira koji se promatra što se uspješno rješava uvođenjem novih značajki, odnosno novih informacija o prozoru u sustav.

Do sada su najviše korištene značajke temeljene na informaciji o bridovima, boji, teksturi lokalnim oblicima te svojstvima gradijenta i kovarijance. Dodavanje novih značajki pokazao se kao vrlo uspješan način poboljšanja rada detektora pješaka, no proširenje prostora značajki može biti problematično za klasične algoritme učenja što se rješava redukcijom prostora značajki metodom Fisherove diskriminantne analize, analizom glavnih komponenti ((engl. *Principal Component Analysis*, *PCA*), ili postupkom parcijalnih najmanjih kvadrata (engl. *Partial Least Squares*, *PLS*) [6].

1.3. Konceptualno rješenje zadatka

1.3.1. Pregled značajki temeljenih na teksturi i boji

Promatrajući pješake na fotografijama, možemo uočiti karakteristike koje ih razlikuju od, na primjer, pozadine: vertikalni bridovi uz rub siluete, često uniformna boja odjeće, razlika u teksturi iste naspram teksturi pozadine, boja kože u području glave, ruku i nogu... Motivacija za komplementiranjem značajki boje i teksture značajkama vezanim uz bridove (HOG deskriptori) tako se prirodno nameće kao bolji opis sustava.

Značajke temeljene na teksturi i boji obično se koriste kao nadopuna značajkama fokusiranim na bridove, za koju ćemo koristiti popularnu HOG metodu Dalal i Triggsa [2]. Njihova se metoda pokazala dobrom na više baza podataka, ali promatranjem isključivo gradijenata potencijalno odbacujemo korisne izvore informacija, što vodi lažnoj pozitivnoj klasifikaciji. Primjerice tekstura nam dosta pomaže kod prepoznavanja odjeće, ali i pozadine, a boja kod prepoznavanja boje kože.

Histogram usmjerenih gradijenata ili HOG značajke

Značajke temeljene na teksturi

Vjerojatno najpoznatija metoda ekstrakcije značajki koje opisuju teksturu potječe iz Haralickovog rada [5] još iz 1979. Za opis teksture koristi takozvanu *co-occurrence* matricu, iz koje se zatim računaju same značajke, kao što su npr. korelacija, srednje vrijednosti i razne mjere entropije.

Osnovna ideja iza matrice jest određivanje vjerojatnosti susjedstva svih parova intenziteta boje. Tako primjerice horizontalna matrica H kao element $h_{i,j}$ sadrži vjerojatnost da je nasumični par horizontalno susjednih piksela ima intenzitete redom i i j . Matrica se tipično računa za horizontalni, vertikalni, te oba dijagonalna smjera (smjerovi 0° , 45° , 90° , 135°). Još jedan parametar koji se može koristiti jest udaljenost d , pa tako matrica $V^{(d)}$ uzima u obzir parove koji su vertikalno susjedni na udaljenosti d .

Svaka tako dobivena matrica promatra se kao matrica zajedničke distribucije vjerojatnosti $p(i, j)$, te iz nje računamo značajke. Tipični primjeri značajki koje Haralick navodi u svome radu su primjerice:

– varijanca:

$$\sum_i \sum_j (i - \mu)^2 p(i, j)$$

– korelacija:

$$\frac{\sum_i \sum_j (ij) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

– entropija:

$$-\sum_i \sum_j p(i, j) \log(p(i, j))$$

Značajke temeljene na boji

U [6] se za iskorištavanje informacije sadržane u boji koristi jednostavno proširenje HOG histograma. Tijekom izgrade HOG histograma, promatramo koji boji pripada gradijent s najvećom normom te se formira trostupčani histogram frekvencija svake od tri boje za trenutni prozor detekcije. Tako uz histogram gradijenata promatramo i histogram boja.

1.3.2. Redukcija dimenzije

Dovođenje mnogo značajki u sustav, odnosno visoka dimenzionalnost vektora značajki, može predstavljati problem za klasične tehnike učenje poput SVM-a, posebice uz često mali skup primjera za učenje.

Pretpostavimo (slično kao u [6]) da je detekcijski prozor podijeljen na preklapajuće blokove u iz kojih se zatim ekstrahiraju opisane značajke. Vektori značajki za sve blokove istog prozora konkatenuiraju se u vektor značajki prozora. Velika dimenzija takvog prostora značajki koju dobivamo promatrajući, uz gradijente, teksturu i boju, predstavlja problem za klasične metode strojnog učenja. No, zbog podijele prozora detekcije u blokove, značajke se ekstrahiraju iz susjednih blokova, što neizbježno vodi do sličnih značajki u bliskim blokovima te samim time i do kolinearnosti. Tako ima smisla iskoristiti nekakvu tehniku redukcije dimenzije, kao što su PCA (*Principal Component Analysis*) ili FDA (*Fisher Discriminant Analysis*). [6] koristi PLS (*Partial Least Squares*), inače regresijsku metodu, praktičnu i za redukciju dimenzije skupa značajki zbog njene brzine i primjenjivosti na ovaj specifičan model.

Nakon redukcije dimenzije za treniranje klasifikatora možemo koristiti metode učenja kao što je SVM (*Support Vector Machine*) koje bi na prostoru s previše dimenzija bile neupotrebljive.

1.3.3. Plan arhitekture sustava računalnog vida

Sustav će se kao i obično sastojati od dvije osnovne komponente: treniranja i testiranja te primjene.

Treniranje klasifikatora



Slika 1.2: Dijagram treniranja SVM klasifikatora

Treniranje se sastoji od nekoliko koraka koji su okvirno prikazani na slici 1.2. Za svaku sliku iz baze računaju se gradijenti (horizontalni i vertikalni smjer). Zatim po slici pomičemo klizni prozor, dijelimo ga na preklapajuće blokove i za svaki blok računamo histogram gradijenata i histogram boja. Prozor potom ponovno dijelimo na blokove (ne nužno istih dimenzija kao u prethodnom koraku), te računamo *co-occurrence* matrice za sva četiri smjera i udaljenost $d = 1$ matricu, iz kojih dobivamo značajke koje opisuju teksturu. U sljedećem koraku reduciramo dimenziju prostora značajki. Konačno, na tako pojednostavljenim primjerima treniramo SVM. U tijeku postupka parametri metode redukcije i SVM-a dobivaju se metodom 10-koračne križne validacije (*10-fold-cross-validation*).

Testiranje klasifikatora

Novotreniranom klasifikatoru ćemo zatim testirati sposobnost generalizacije i pre-treniranost.

Primjena klasifikatora

Slično kao kod treniranja promatramo klizni prozor i za njega računamo značajke, kojima zatim reduciramo dimenziju te ih dajemo kao ulazne podatke klasifikatoru. Ukoliko klasifikator procijeni da se radi o pješaku, dojavljujemo poziciju trenutnog kliznog prozora.

2. Postupak rješavanja zadatka

2.1. Ekstrakcija značajki

2.2. Treniranje prve faze klasifikatora

2.3. Treniranje druge faze klasifikatora

3. Ispitivanje rješenja

3.1. Ispitna baza

Za treniranje i verifikaciju rješenja koristit ćemo skup podataka INRIA [3]. To je skup fotografija urbanog okruženja u boji i pripadnih anotacija. Za svaku fotografiju zabilježen je skup graničnih prozora (engl. *bounding window*) unutar kojih se nalaze prikazi uspravnih osoba.

INRIA se do sada često koristio za trening detektora zbog raznolikosti pozadinskih okruženja osoba na slikama i točnosti anotacija [1]. Za razliku od ostalih javno dostupnih baza podataka za detekciju pješaka, ove fotografije nisu dobivene iz videa te su relativno visoke kvalitete i fotografirane su iz različitih točaka gledišta. U drugim bazama podataka, pješaci su većinom konecentrirani u centralnoj horizontali fotografije jer je ista dobivena iz vozačeve perspektive.

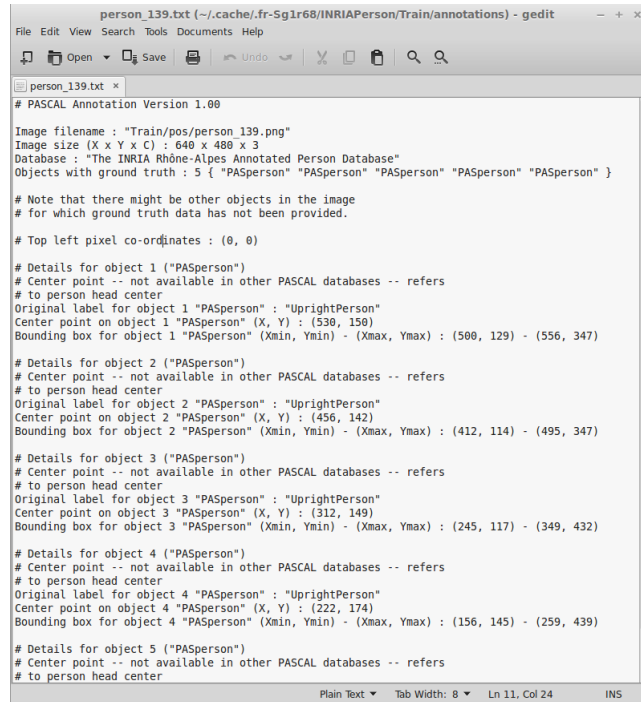


Slika 3.1: Primjeri fotografija iz baze podataka INRIA.

Baza slika INRIA sadrži slike iz nekoliko različitih izvora:

- *Graz 01* skup podataka sa pridodanim bilješkama
- slike iz osobne kolekcije samog autora koje su zbog svoje veličine u originalu izrezane tako da prikazuju samo osobu
- slike sa servisa Google Images

Sama se baza sastoji od podskupa podataka za trening i podskupa za evaluaciju. U podskupu za trening označena su 1208 pješaka u 614 od 1832 fotografije. U podskupu za testiranje označeno je 566 pješaka u 453 od 741 fotografije [4]. Primjeri fotografija sa i bez pješaka dani su na slici 3.1. Primjer anotacije za lijevu fotografiju sa slike 3.1 dan je na slici 3.2.



```

person_139.txt (-/cache/fr-5glr68/INRIAPerson/Train/annotations) - gedit
File Edit View Search Tools Documents Help
person_139.txt x
# PASCAL Annotation Version 1.00

Image filename : "Train/pos/person_139.png"
Image size (X x Y x C) : 640 x 480 x 3
Database : "The INRIA Rhône-Alpes Annotated Person Database"
Objects with ground truth : 5 { "PASperson" "PASperson" "PASperson" "PASperson" "PASperson" }

# Note that there might be other objects in the image
# for which ground truth data has not been provided.

# Top left pixel co-ordinates : (0, 0)

# Details for object 1 ("PASperson")
# Center point -- not available in other PASCAL databases -- refers
# to person head center
Original label for object 1 "PASperson" : "UprightPerson"
Center point on object 1 "PASperson" (X, Y) : (530, 150)
Bounding box for object 1 "PASperson" (Xmin, Ymin) - (Xmax, Ymax) : (500, 129) - (556, 347)

# Details for object 2 ("PASperson")
# Center point -- not available in other PASCAL databases -- refers
# to person head center
Original label for object 2 "PASperson" : "UprightPerson"
Center point on object 2 "PASperson" (X, Y) : (456, 142)
Bounding box for object 2 "PASperson" (Xmin, Ymin) - (Xmax, Ymax) : (412, 114) - (495, 347)

# Details for object 3 ("PASperson")
# Center point -- not available in other PASCAL databases -- refers
# to person head center
Original label for object 3 "PASperson" : "UprightPerson"
Center point on object 3 "PASperson" (X, Y) : (312, 149)
Bounding box for object 3 "PASperson" (Xmin, Ymin) - (Xmax, Ymax) : (245, 117) - (349, 432)

# Details for object 4 ("PASperson")
# Center point -- not available in other PASCAL databases -- refers
# to person head center
Original label for object 4 "PASperson" : "UprightPerson"
Center point on object 4 "PASperson" (X, Y) : (222, 174)
Bounding box for object 4 "PASperson" (Xmin, Ymin) - (Xmax, Ymax) : (156, 145) - (259, 439)

# Details for object 5 ("PASperson")
# Center point -- not available in other PASCAL databases -- refers
# to person head center

```

Slika 3.2: Primjer anotacije pozitivno klasificiranih kliznih prozora za lijevu fotografiju na slici 3.1.

3.2. Rezultati učenja i ispitivanja

3.3. Analiza rezultata

4. Literatura

- [1] Rodrigo Benenson, Mohamed Omran, Jan Hendrik Hosang, and Bernt Schiele. Ten years of pedestrian detection, what have we learned? [abs/1411.4304](#), 2014.
- [2] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *In CVPR*, pages 886–893, 2005.
- [3] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *International Conference on Computer Vision & Pattern Recognition*, volume 2, pages 886–893, INRIA Rhône-Alpes, ZIRST-655, av. de l’Europe, Montbonnot-38334, June 2005.
- [4] Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona. Pedestrian detection: An evaluation of the state of the art. 34(4):743–761, April 2012.
- [5] Robert M. Haralick, K. Sam Shanmugam, and Its’hak Dinstein. Textural features for image classification. (6), 1973.
- [6] William Robson Schwartz, Aniruddha Kembhavi, David Harwood, and Larry S. Davis. Human detection using partial least squares analysis.
- [7] Paul Viola and Michael J. Jones. Robust real-time face detection. 57(2):137–154, May 2004.