

Robusna detekcija lica

Martin Soldić

Sveučilište u Zagrebu

Fakultet elektrotehnike i računarstva

martin.soldic@fer.hr

Sažetak— Robusna detekcija lica je proces utvrđivanja prisutnosti lica na statičnoj slici ili u videu, te određivanje lica pozicija ukoliko je utvrđena njihova prisutnost, pri čemu slika ili video mogu sadržavati veći broj lica, snimanih u raznim pozama s mogućim prekrivanjem i nekontroliranim uvjetima osvjetljenja. U ovom radu se ukratko opisuju metode za detekciju lica, koje su podijeljene u 4 kategorije prema načinu rada: metode temeljene na znanju, metode sa značajkama invarijantnim na poze i uvjete osvjetljenja, metode temeljene na predlošcima i metode bazirane na izgledu. Robusna detekcija lica ključ je za mnoge primjene u interakciji čovjeka i računala, te u primjeni deidentifikacije ljudi u sustavima video nadzora sa svrhom zaštite privatnosti.

Ključne riječi - lokalizacija lica, detekcija lica, deidentifikacija lica, video nadzor, strojno učenje

Uvod

S povećanjem broja terorističkih napada po ulicama svjetskih gradova, intenzivnim razvojem prometa i izgradnjom velikih stambenih naselja, javlja se sve veća potreba za automatiziranim video nadzorom i praćenjem ljudi kako bi se povećala razina sigurnosti. Nasuprot povećanju sigurnosti, kontinuiranim snimanjem i praćenjem smanjuna se prava na privatnost. Još veći problem ugrožavanja privatnosti predstavljaju društvene mreže, na kojima se često mogu naći slike djece i ljudi, koje su postavljene da budu javno dostupne. Razvoj današnjeg društva i istovremena potreba za visokom sigurnosti često djeluju kontradiktorno. Zbog toga su napravljena mnoga istraživanja kako zadovoljiti obje potrebe uz pomoć automatizirane deidentifikacije ljudi na slikama i videu. Kao početni korak u deidentifikaciji pojedinca, ključna je robusna detekcija lica, jer lica sadrže jaka identifikacijska obilježja. Detekcija lica je dugogodišnje istraživano područje u domeni računalnog vida, s više tisuća radova objavljenih tokom zadnjih 40 godina. Iako su mnoge metode testirane i svake se godine baze nadopunjuju, još uvijek postoji veliki raskorak između uspješnosti detekcije od strane čovjeka u odnosu na računalnu detekciju [1]. Kao najveći problem u detekciji ističe se pronalaženje dovoljno robusnih i diskriminatornih značajki s obzirom na različite varijacije u osvjetljenju, kontrastu, pozama lica i izrazima lica. Također, veliki problem je optimizirati proces strojnog učenja kako bi se postigla visoka detekcija uz jako dobru generalizaciju pri

detektiranju raznolikih poza i izraza lica, a koje nisu evidentirane u bazi nad kojom se provodi učenje ili opisuju pravila detekcije. Potrebno je objasniti što je detekcija i u kakvom je odnosu s terminom lokalizacija. Detekcija lica (objekta) u slici je prema radu Yang et al. [2] proces utvrđivanja prisutnosti lica na slici, te određivanje njihovih pozicija ukoliko ona postoje. Lokalizacija je proces nalaženja pozicija lica (objekta) u slici pod uvjetom da je utvrđeno ili pretpostavljeno njihovo postojanje u slici. S obzirom da određene metode, koje su opisane u prva tri poglavlja ovog rada, mogu samo lokalizirati lica, u tim poglavljima je strogo rečeno kojoj kategoriji pripada pojedina metoda. Iako nije uvijek jednostavno napraviti podjelu po načinu rada algoritama za detekciju lica, ovdje se koristi podjela prema preglednom radu Yang et al. [2]. Prema njemu, postoje 4 tipa pristupa u detekciji lica: metode temeljene na znanju, metode sa značajkama invarijantnim na uvjete osvjetljenja i promjene poze, metode temeljene na definiranim predlošcima oblika i metode temeljene na izgledu (učene na velikom broju primjera). Sukladno toj podjeli, ovdje se u poglavljima od 1 do 4 opisuju određeni reprezentativni radovi, a u poglavlju 5 opisuju se baze za treniranje i testiranje algoritama opisanih u prethodnim poglavljima. Na kraju je dan zaključak i popis literature.

I. METODE TEMELJENE NA ZNANJU

Ove metode su među najstarijim pristupima za detekciju lica na slici, a koriste skup pravila napisanih od strane ljudskih eksperata, kojima se određuju vrste značajki, te njihovi međusobni geometrijski i metrički odnosi uz definirane pragove. Iako neke od ovih metode mogu utvrditi nalazi li se lice na slici, one su većinom dizajnirane za lokalizaciju lica, pri čemu je najčešće samo jedno uspravno lice prisutno na slici u relativno visokoj rezoluciji. Jedan od najvećih problema u izradi ovih metoda je kvaliteta prijenosa ljudskog znanja u dobro definirana pravila. Pravila koja su detaljno opisana često nemaju dobru generalizacijsku moć i takvi algoritmi imaju veliki broj promašaja u detekciji (engl. *false negative*), pa su često beskorisni u praksi. S druge strane, metode s relativno malo pravila i robusnim pragovima kod izvođenja pravila, često daju veliki broj lažno pozitivnih (engl. *false positive*) detekcija, čime također gube praktični smisao. Ovdje se može

istaknuti algoritam Yang i Huang [3], koji pripada hijerarhijskoj detekciji lica po principu *s vrha prema dnu*, a sastoji se od 3 razine. Ulazna slika se pretvori u torazinsku rezolucijsku piramidu, kako bi se na najgrubljoj rezoluciji mogle pronaći potencijalne lokacije lica koristeći jednostavan skup pravila, a analizom dotičnih kandidata na višim rezolucijama uz pomoć pravila razina 2 i 3, može se donijeti konačna odluka nalazi li se na toj lokaciji lice ili ne. Potencijalna lica se nalaze metodom klizećeg prozora, koji tijekom svog pomicanja po ulaznoj slici uzima dijelove slike nad kojom se evaluiraju pravila za detekciju lica. Na najgrubljoj rezoluciji (razina 1) pravila uzimaju u obzir odnose srednje osvijetljenosti centralnog dijela klizećeg prozora (lica) u odnosu na gornji ili rubni dio lica. Ova velika razlika intenziteta je više istaknuta za lica, nego za druge objekte ili pozadinu. Ako trenutna pozicija prozora zadovoljava spomenute kriterije, detaljnija analiza se evaluira pravilima razina 2 i 3 pri višim rezolucijama slike, kojima se utvrđuju značajke očiju, nosa ili usta. Kao mnogi drugi algoritmi ove kategorije [4], postoji relativno visoki broj lažno pozitivnih detekcija, te niža kvaliteta detekcije u odnosu na druge pristupe. Zadnjih 20 godina, ove metode se praktički više ne koriste.

II. METODE TEMELJENE NA ROBUSNIM ZNAČAJKAMA (RUBOVI, TEKSTURE, BOJE, OBLICI)

Ove metode koriste strukturalne značajke lica, koje ostaju relativno postojane promjenom poze ili osvijetljenja lica. Najčešće se koriste rubovi i oblici (elipse), te definirani rasponi boja tipičnih za boju kože u prostorima HSV, RGB, YCrCb, CIE LUV ili drugima. Za izradu uzoraka tekstura lica, koriste se značajke poput kože ili kose, nad kojima se provode statističke analize (engl. *Spatial Grey Level Dependence Matrix - SGLD*) [5]. Dodatkom informacije o boji i ispravnom klasifikacijom teksture lica, može se dobiti relativno robustan lokalizator lica, neovisan o promjeni poze i rotaciji lica [6]. Metode temeljene na oblicima baziraju se na detekciji rubova, analizi njihove lokalne orijentacije i usporedbi s predlošcima oblika za neka istaknuta mjesta na licu (obrve, usta, nos). Morfološkim povezivanjem komponenti utvrđuje se lokacija lica. Han et al. [7] intenzivno koriste morfologiju područja oko očiju s obzirom da autori smatraju oči ključnim dijelom lica, koji je robustniji na promjene oblika u odnosu na usta. Metode temeljene na boji su najrobustnije ako se modeliraju mješavinom Gaussovih razdioba [8,9], pri čemu se parametri pronalaze EM (maksimizacija očekivanja) algoritmom. Ideja je da ljudi s obzirom na različitu rasnu pripadnost i pri različitim uvjetima osvijetljenja nemaju unimodalnu distribuciju boja na licu, neovisno o izboru prostora boja, pa se mješavinom unimodalnih razdioba omogućuje robustnija detekcija. Često se značajke iz ove kategorije međusobno kombiniraju [10].

III. METODE TEMELJENE NA PREDDEFINIRANIM PREDLOŠCIMA

Prije pristupanja detekciji ili lokalizaciji lica potrebno je imati skup prethodno definiranih vizualnih predložaka, koji predstavljaju pojedinu značajku lica. Za svaku značajku obično se definira više predložaka kako bi se pokrile brojne varijacije oblika kod značajki, te stvorila sposobnost prilagodbe različitim veličinama (skalama) lica. S povećanjem broja predložaka, omogućuje se veća robusnost na uspješnu detekciju/lokalizaciju. Kod procesa detekcije lica na slici, moraju se dobiti rubovi slike pomoću rubnog operatora (Canny, Sobel), te se korelacijom rubnih značajki pojedinog dijela lica s preddefiniranim predlošcima utvrđuje lokacija značajki lica, a time i cijelog lica. Neki radovi koji koriste ovu tehniku su [11, 12, 13]. Jedan od prvih radova uopće na detekciji lica bio je Sinha [14]. Postoje radovi s deformirajućim predlošcima – aktivne konture, zmije (engl. *Snakes*) [15].

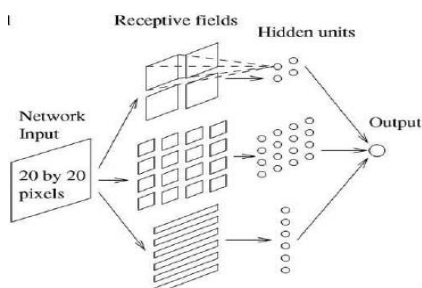
IV. METODE TEMELJENE NA IZGLEDU I UČENJU NA PRIMJERIMA

Za razliku od prethodnih metoda, koje koriste različita ekspertna znanja o oblicima i povezanosti značajki, ovdje se dinamički izgrađuje model lica procesom strojnog učenja nad skupom vektora vizualnih značajki dobivenih iz velikog skupa pozitivnih i negativnih primjera slika lica. Kako bi izgrađeni model davao kvalitetne rezultate detekcije, vektore koji sudjeluju u strojnom učenju treba dobiti iz značajki koje su diskriminativne s obzirom na klasnu pripadnost (lice), a dovoljno robusne s obzirom na promjenu intraset značajki (oblici lica, osvijetljenje, poze, izrazi). Ove metode se razlikuju u izboru vizualnih značajki, te metodama strojnog učenja. Među prvim metoda strojnog učenja bile su neuronske mreže [16, 17]. U nastavku se opisuju neke najuspješnije metode temeljene na neuronskim mrežama, dubokom učenju (engl. *Deep Learning*), naivnom Bayesu, metodama jačanja klasifikatora (engl. *Boosting*), potpornim vektorima, te jačanju uz stabla odluke.

A. Metode temeljene na neuronskim mrežama kao tehnicu učenja

S razvojem ideje višeslojnog perceptrona i sklopovlja, tijekom osamdesetih godina razvila se zamisao o korištenju neuronske mreže u domeni računalnog vida. Neuronska mreža sadrži višeslojnu neuronsku arhitekturu, obično podijeljenu na tri sloja – ulazni sloj, unutarnji, te vanjski (izlazni), koji može dati binarne vrijednosti 0 ili 1, ali i raspone brojeva unutar granica [0,1]. Kao binarni klasifikator, koji može modelirati vrlo veliki broj klasno specifičnih (intraset) značajki, ova metoda je pogodna za detekciju lica na slici. Najveći problem je efikasno napraviti dobar raspored veza među neuronima i ugoditi parametre prijenosnih funkcija, kako bi se izbjegla pretreniranost, a omogućila dobra generalizacija. Jedan od prvih radova je Agui et al. [16]. U tom radu moguća je detekcija samo jednako velikih lica, a neuronska mreža sastoji se od dva dijela. Ulaz u neuronsku mrežu čine intenziteti sive

razine ulazne slike i intenziteti sive razine nakon provedene konvolucije Sobelovim operatorom. Kao najuspješniji predstavnik ove metode učenja ističe se Rowley et al. [17], koji implementira tehniku klizećeg prozora nad multirezolucijskom piramidom, čime je bila omogućena detekcija uspravnih lica različitih veličina. Mreža je bila troslojna, pri čemu su ulazni sloj činila 3 tipa receptivnih polja. Prvi tip je dijelio ulaznu sliku na 4 jednaka kvadratna dijela, drugi je dijelio na 16 kvadrata, dok je treći bio predstavljen sa šest pravokutnih blokova dimenzija 20x4 piksela. Skriveni sloj brojao je ukupno 26 neurona. Cijela arhitektura sustava predstavljena je slikom 1.



Slika 1. Rowley et al. [17] neuronska mreža. Preuzeto iz [2].

Ulaz u neuronsku mrežu čini slika lica s normaliziranim histogramom osvijetljenja. Izlaz je raspon vrijednosti između [-1 i 1], pri čemu je vrijednost blizu -1 oznaka za neprisutnost lica, a ona blizu 1 za prisutnost. Mreža je trenirana tehnikom propagacije unatrag, a rezultati su iznosili preko 91 % za CMU bazu (130 sivih slika s uspravnih 557 lica). Proširenje ove metode za rotirana lica dano je u Rowley et al. [18]. Od ostalih radova ističu se Sung i Poggio [19], te Lin et al. [20].

B. Metode temeljene na dubokim neuronskim mrežama (engl. Deep Learning)

Ovo je relativno novo područje učenja, koje je nastalo kao nadopuna klasične neuronske mreže, samo s mnogo više unutarnjih slojeva, čime je postalo moguće modelirati vrlo kompleksne odnose. Za probleme računalnog vida razmatraju se uglavnom konvolucijske mreže, kao podtip ovih mreža. Tip prijenosnih funkcija između neurona je praktički isti kao za obične neuronske mreže, ali se provodi nešto modificirani gradijentni spust za treniranje. Ovdje se daje kratki osvrt na rad Zhang i Zhang [21], u kojem se inicijalno koristi klasični detektor lica u više poza kao što je HeadHunter [1] ili Zhu i Ramanan [22], pa se nakon inicijalnog određivanja poze, slika lica procesuirala s konvolucijskom mrežom kao dodatnim klasifikatorom. Svaka slika koja izlazi iz prvog klasifikatora skalira se na 32 x 32 piksela, nakog čega se obavlja pretprocesiranje (izjednačavanje histograma svjetline, kontrasta). Mreža je sastavljena tako da se prva dva sloja sastoje od 32 jezgre, dok treći ima 24, a četvrti sloj 512 potpuno povezanih neurona. Izlaz mreže podijeljen je u tri kategorije: prisutnost lica (da ili ne), poza i određivanje ključnih značajki lica (obrve, nos, oči, usta). Treniranje je provedeno nad slikama iz CMU MultiPIE [23] baze, te nekih slika preuzetih s interneta. Ukupno je korišteno oko milijun slika. Testiranje je provedeno nad FDDB [24] bazom i

rezultati su pokazali osjetno povećanje kvalitete detekcije u odnosu na standardni detektor s više pogleda [22]. Ipak, ova metoda nije superiorna u odnosu na najnovije radove koji ne primjenjuju duboko učenje [1] i [25].

C. Metode temeljene na naivnom Bayesu

Jedan od najreprezentativnijih algoritama detekcije koji su vrlo uspješno koristili ovaj model učenja je detektor Schneiderman - Kanade [26], koji je podjednako uspješno detektirao slike uspravnih i profilnih lica. Algoritam izgrađuje cjeloviti model lica kombinirajući vizualne (engl. *appearance*) značajke i pozicije određenih regija lica putem naivnog Bayesa, pri čemu se značajke svake regije lica modeliraju zasebno. Iako ovo nije potpuno precizan model, dobra generalizirajuća svojstva naivnog Bayesa i kvalitetno izlučivanje vizualnih značajki regija lica, omogućilo je svojevremeno postizanje najviših rezultata u detekciji. Empirijski je utvrđena potreba za 17 prostornih regija, koje imaju dobra diskriminatorna svojstva, a vizualne značajke se uzimaju uzorkovanjem nisko i visokofrekventnih komponenti višerezolucijske piramide valića (engl. *Dyadic* dekompozicija). Ulazna siva slika s licima pretvori se u multirezolucijsku piramidu, te se detaljno skenira klizećim prozorom, a za svaki klizeći prozor računa se piramida valića i potom statistika pripadnosti regiji lica. U slučaju dobivanja posteriori vjerojatnosti veće od praga definiranog treniranjem, lice je detektirano. Da bi se napravila statistika svake regije lica, potrebno je kvantificirati koeficijente biortogonalnih valića i izraditi histogram pojave kvantificiranih koeficijenata valića, čime histogram odgovara funkciji uvjetne gustoće vjerojatnosti određene regije lica. Za uspješnu detekciju prednjeg uspravnog, te profilnog lica, potrebno je za svaki prozor istovremeno provesti detekciju prednje i profilne strane lica. Mehanizmom glasovanja (engl. *voting*) odlučuje se o točnoj pozi lica. Ovo glasovanje zapravo se provodi na kraju prolaska klizećeg prozora po svim pozicijama i skalama rezolucijske piramide, kako bi se omogućilo uspješno stapanje susjednih detekcija i dala prednost najvjerojatnijoj pozi u tom području. S detekcijom nešto boljom od Rowley et al. [17] na CMU+MIT bazama, algoritam je bio svojevremeno najuspješniji detektor lica, posebice profilnih. Iz ove kategorije još treba istaknuti rad Rikert et al. [27].

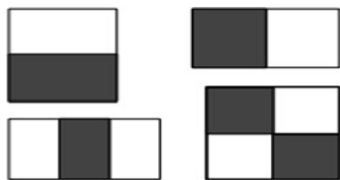
D. Metode temeljene na jačanju klasifikatora (engl. Boosting)

Metode temeljene na tehnikama jačanja klasifikatora većinom se temelje na AdaBoost algoritmu [28,29], te nešto boljem RealBoost algoritmu [30]. AdaBoost je algoritam strojnog učenja koji generira jaki klasifikator na temelju većeg broja slabih klasifikatora (decizijsko stablo, perceptron, step funkcija), a adaptivno se prilagođuje primjerima iz skupa za učenje, tako da konačni klasifikator ima kvalitetnu klasifikacijsku sposobnost uz što manji broj slabih klasifikatora, kako bi detektor bio brži. Formulom se to može predočiti s $F(\mathbf{x}) = \sum_{t=1}^T f^t(\mathbf{x})$, gdje $f^t(\mathbf{x})$ označuje slabi klasifikator (step funkcija nad Haarovom značajkom [31]), a \mathbf{x} je vektor slike klizećeg prozora. T je broj slabih klasifikatora, a za izlaz kod binarne klasifikacije, obično se stavlja $\text{sign}(F(\mathbf{x}))$, gdje je $\text{sign}(\cdot)$ signum funkcija. Jedan od najznačajnijih algoritama detekcije lica je iz 2001. godine Viola i Jones

algoritam [31], koji je na svojedobnim bazama bio dosta robustan u detekciji, a mogao je raditi u realnom vremenu, čime je otvorio vrata širokoj aplikativnoj primjeni detektora lica – mobilna platforma, fotoaparati. Izvorni algoritam radi samo na sivim slikama uspravnih lica. Idejom korištenja Haarovih značajki, za čije se računanje koristi integralna slika, te primjenom kaskadnih klasifikatora učenih AdaBoost tehnikom, pri čemu relativno mali broj klasifikatora početnih kaskada odbija veliki broj negativnih slika lica, omogućena je vrlo brza detekcija, neovisno o veličini slike uz relativno visoku točnost detekcije. U radu algoritma koristi se klizeći prozor kao i u brojnim prethodnim detektorima, samo se ovdje skalira prozor umjesto ulazne slike, kako bi se dobila detekcija različito velikih lica. Skaliranje prozora u ovom algoritmu je računski manje zahtjevna operacija u odnosu na skaliranje ulazne slike, jer se skaliranjem prozora trebaju samo izračunati nove Haarove značajke, dok bi se rezolucijskom piramidom nad svakom razinom piramide morale računati nove integralne slike iz kojih se računaju Haarove značajke. Kao što je rečeno, za vizualne značajke koriste se Haarove pravokutne regije, kod kojih razlika suma piksela crne i bijele regije predstavlja ulaz u neku funkciju – step funkciju, koja na izlazu daje 0 ili 1 ovisno o izboru praga. Osnovnih pravokutnih značajki ima 4 (slika 2), a ukupan broj dobije se kombiniranjem njihovih veličina i pozicija unutar osnovne dimenzije klizećeg prozora – 24 x 24 piksela. Tako je moguće dobiti preko 100 000 varijacija osnovnih značajki prije pokretanja učenja. Učenjem AdaBoost algoritmom taj se broj značajki smanji, tako da ih konačni detektor ima mnogo manje nego na početku učenja. Obično klasifikatori početnih kaskada broje samo nekoliko značajki, dok ih završne kaskade imaju po više tisuća. Da bi se izračunala pojedina značajka, koristi se ideja integralne slike. Na početku rada algoritma, za ulaznu sivu sliku s licima izračuna se integralna slika, tako što se za svaki piksel računa suma vrijednosti intenziteta slike svih piksela lijevo i iznad od trenutno promatranog piksela, uključivo trenutni piksel (4.1.).

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (4.1.)$$

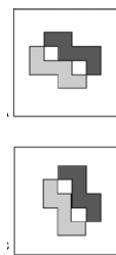
U (4.1.) vrijednosti piksela integralne slike dane su s $ii(x,y)$, a vrijednosti piksela izvorne slike označeni su s $i(x',y')$. Za izračun sume piksela pravokutne regije, koristi se formula dana slikom 4. Izračunom dviju ili više susjednih pravokutnih regija, moguće je vrlo brzo računati Haarove značajke.



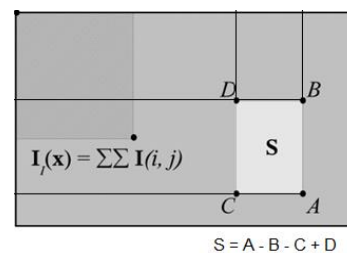
Slika 2. Haarove pravokutne značajke. (Preuzeto iz [31])

Integralna slika i kaskade omogućile su rad u stvarnom vremenu, tako da je za detekciju lica na slici od 320 x 240 piksela potrebno oko 50 ms na modernim računalima. Osim prednjih slika lica, algoritam se može nadopuniti proširenim Haarovim značajkama - Viola i Jones [32], tako da detektira rotirana uspravna ili profilna lica. Skup Haarovih pravokutnih značajki proširuje se dijagonalnim značajkama (slika 3). Ove proširene značajke traže nešto veći broj računskih operacija u

odnosu na izvorne značajke, a zbog dijagonalne strukture, omogućuje se robustno hvatanje relevantnih značajki na profilnim slikama lica. Za omogućavanje detekcije rotiranih lica, napravljena je podjela na zone od 30 stupnjeva, tako da postoji ukupno 12 različitih detektora, koji su trenirani na umjetno rotiranim slikama lica izvorne baze. Umjesto paralelnog korištenja svih detektora nad jednom slikom klizećeg prozora, ovdje je prvo napravljen detektor stupnja rotiranosti, pa se ovisno o najizglednijem kutu rotacije, upotrijebi detektor za taj izbor rotacije. Time je postignuto ubrzanje rada, ali je nešto smanjena točnost detekcije. Detektor kuta rotacije je stablo odluke, čiji čvorovi predstavljaju skupinu jednostavnih klasifikatora temeljenih na Haarovim dijagonalnim i pravokutnim značajkama. Nakon nekoliko grananja, može se odrediti koja rotacija je najvjerojatnija. Sličan postupak korištenja stabla odluke koristi se za određivanje vrste poze, koja se može nalaziti unutar raspona od -90 do 90 stupnjeva rotacije, a ukupno postoji 6 klasa poza, za koje je treniran poseban detektor. Detektor lica treniran za prethodno odabranu klasu poze donosi konačnu odluku o prisutnosti lica unutar klizećeg prozora.



Slika 3. Dijagonalni prošireni blokovi. (Preuzeto iz [32]).



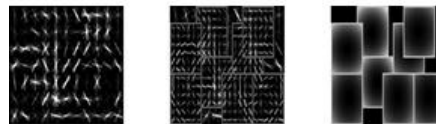
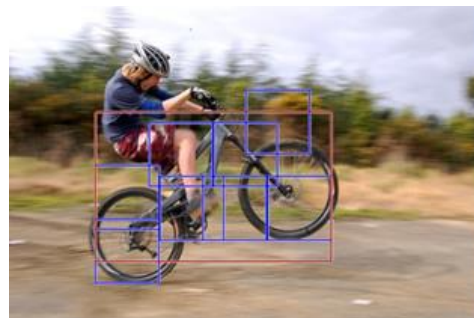
Slika 4. Suma intenziteta piksela nad površinom slike S kao aritmetička kombinacija integralnih suma točaka D,B,C i A. (Preuzeto iz [31]).

Dodatne varijacije stabala za višepoznu detekciju dani su u [33, 34], kako bi se otklonili nedostaci standardnih piramidalnih ili stablastih poznih klasifikatora. Neki su autori koristili združene Haarove značajke (engl. *joint Haar*) [35], koji kombiniraju Haarove značajke s raznih pozicija klizećeg prozora i onda tako enkodirani zapis ulazi u proces učenja. Baluja et al. [36] koriste umjesto Haarovih značajki parove piksela nad čijim se razlikama intenziteta računaju odzivi slabih klasifikatora pri AdaBoost algoritmu. Dobiveni su nešto slabiji rezultati u odnosu na Viola –Jones pristup, ali algoritam je bio brži. Jedan od velikih problema Haarovih značajki je

slaba prilagodljivost velikoj varijaciji osvjetljenja i geometriji lica, te se zbog toga javljaju radovi s anizotropnim Gaussovima filterima [37], te [38] s LAB značajkama (engl. *Locally assembled binary features*), koje su otporne na velike varijacije u osvjetljenju. Kvalitetan pregledni rad prethodno spomenutih značajki korištenjem tehnika jačanja klasifikatora za detekciju lica dan je u Zhang i Zhang [39].

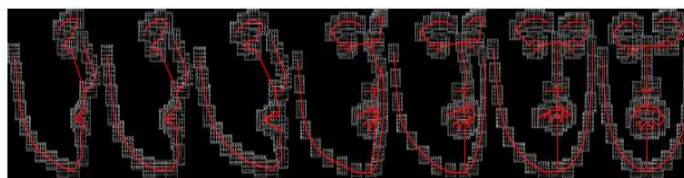
E. Metode temeljene na stroju s potpornim vektorima (SVM) i deformirajućim komponentama (Deformable Part-Based Model-DPM)

Kao metoda bazirana na kvadratnom programiranju kojom se omogućuje optimalno određivanje postavljanja decizijske hiperravnine između margine razdvajanja uzoraka dviju klasa, SVM je postao vrlo popularna tehnika strojnog učenja u binarnoj klasifikaciji, osobito u domeni računalnog vida. Prvi rad koji koristi SVM za detekciju lica bio je Osuna et al. [40], dok za pješake Oren et al. [41], čiji rad koristi valiče kao značajke za učenje klasifikatora. S razvojem značajki temeljenih na histogramima orijentiranih gradijenata (HOG) Dalat et al. [42], počeli su se razvijati napredniji detektori za proizvoljne objekte, a njihovo učenje najčešće se temeljilo na potpornim vektorima. Tako se razvija model s deformirajućim dijelovima (engl. *Deformable Part-Based Models-DPM*) [43] i [44], čiji su temelj posebno prostorno povezani predložci, koji imaju različite veličine i brojnost, ovisno o tipu objekta za kojeg se trenira detektor (slika 5). Temeljna vizualna značajka DPM detektora je HOG, a ubrzo se nadopunjuje s još nekoliko vizualnih značajki, tako da se ideja koristi ne samo za detekciju pješaka [43, 45], već i za lica [1, 22, 46]. U detekciji lica [1] koristi se izvorni DPM model s dodanim CIE LUV prostorom boja kao integralnim komponentama i 6 zasebnih modela ovisno o rotaciji, pri čemu prvi model pokriva kutove od -20° do $+20^\circ$, drugi od $+20^\circ$ do $+60^\circ$, treći od $+60^\circ$ do $+100^\circ$, dok su preostala tri reflektirana prethodna tri. Svaki model sadrži korijenski predložak, te 8 odvojenih (engl. *part*) predložaka. Ovako treniran detektor pokazuje superiorne rezultate u odnosu na sve detektore koji su bili poznati do 2014. godine. Cijeli detektor treniran je na LFW [47] bazi. Model ima dobra generalizirajuća svojstva, što je tipično za DPM, pa nije potrebno imati tisuće ulaznih slika, kao za druge algoritme [31, 32, 37, 43]. Za učenje deformirajućih modela koristi se tzv. latentni SVM - modifikacija metode potpornih vektora, koja kada se jednom namjesti latentne varijable (raspored predložaka za dijelove (engl. *parts*) za pozitivne slike, prevodi inicijalno polukonveksni u konveksni optimizacijski problem.



Slika 5. DPM model s pripadajućim HOG značajkama. (Preuzeto iz [43]).

Kao standardni primjer DPM algoritma za detekciju lica ističe se rad Zhu i Ramanan [22]. Glavna ideja algoritma je korištenje stabala za lokalizaciju značajki lica, na temelju kojih se zaključuje prisutnost i poza lica na nekoj poziciji unutar slike. Postoji 13 modela stabala, koja omogućuju uspješnu estimaciju poza u rasponu od -60 do $+60$ stupnjeva (slika 6). Čvorovi i listovi stabala predstavljaju pozicije točaka s kojih se uzimaju odzivi filtera fiksne veličine (5×5) nad prethodno izračunatim HOG i CSS [48] (engl. *Color self-similarity*) značajkama ulazne slike. Grane stabla nisu fiksne u duljini i poziciji, pa se omogućuje da sve karakteristične točke uspravnog lica mogu biti pronađene na proizvoljnim mjestima uz hijerarhijski i geometrijski raspored koji odgovara opisu ljudskog lica. Za svih 13 modela lica, koristi se kumulativno 146 filtera, dok je za prednja uspravna lica potrebno 68 filtera, čiji uspješan odziv odgovara detekciji pojedine karakteristične točke lica. Ovaj algoritam koristi pozicije na slici s maksimalnim odzivom filtera kao početnim točkama obilaska stabla, te u slučaju dovoljno dobrih odziva preostalih filtera pri obilasku stabla, može se utvrditi je li lice prisutno ili nije. Ovo je jedan od rijetkih primjera algoritama koji ne koriste ideju klizećeg prozora za detekciju. Jedan od sjajnih primjera nekorištenja klizećeg prozora za vrlo veliki broj detektora lica i pješaka dan je u radovima Gualdi et al. [49] i Lampert et al. [50].



Slika 6. Modeli stabala s lokalnim značajkama lica algoritma. (Preuzeto iz [22]).

Algoritam je po godini izlaska bio ravnopravan s najboljim komercijalnim rješenjima za detekciju lica, a model je treniran

nad svega nekoliko stotina pozitivnih i par tisuća negativnih slika. Algoritam najbolje radi na slikama u boji.

HeadHunter algoritam [1] kao jedan od najuspješnijih današnjice, polazi od integralnih značajki korištenih za detekciju pješaka - Dollar et al. [51], s tim da su uvedene određene promjene. Kao značajke se uzimaju magnitude gradijenata rubova, CIE LUV boje, te linearni filteri proizvoljno definirani, dok se za izradu jakog klasifikatora koristi AdaBoost, gdje slabije klasifikatore predstavljaju stabla odluke dubine 2. Detektor se sastoji od 5 modela treniranih u ovisnosti o rotaciji (yaw kut), tako da osnovni pokriva kutove između -20° i $+20^\circ$, druga dva modela pokrivaju kutove između 20° i 60° , te 60° i 100° , te simetrične njima pokrivaju druga dva modela. Učenjem je dobiveno oko 2000 slabijih klasifikatora, čime je postignut optimum za brzinu i kvalitetu detekcije, dok su korištene baze za treniranje i testiranje [23] i [24].

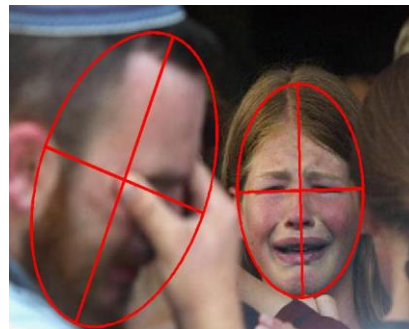
F. Metode temeljene na jačanju klasifikatora i stablima odluke

Umjesto tipičnog korištenja jačanja klasifikatora uz jednostavno stablo odluke (engl. *stump*) [31], detektor lica Liao et al. [25] koristi složeno stablo odluke (dubina 8) kao slabi klasifikator uz AdaBoost za ostvarenje jakog klasifikatora. Koristi se kaskadni model klasifikatora kao u [31], pri čemu za robusnu detekciju u više poza nije potrebno definirati zasebne detektore za svaku pozu kao u [32, 33]. Stablo odluke je izgrađeno na temelju kvantificiranja vrijednosti kvadratne funkcije (ax^2+bx+c) u 256 polja i definiranja pragova, gdje x označava vrijednost normalizirane razlike intenziteta piksela (engl. *Normalized Pixel Difference - NPD*), dok su a , b i c neke konstante. Normalizirana razlika intenziteta piksela (NPD) nad parovima piksela dana je izrazom $f(x, y) = \frac{x-y}{x+y}$, pri čemu x i y označavaju intenzitete piksela. NPD značajka ima svojstva invarijantnosti na promjenu skale slike, te mogućnost rekonstrukcije izvorne slike ako su poznate značajke nad parovima piksela cijele slike. Kombinacija NPD značajki i stabala odluke omogućuje mnogo veću otpornost na promjenu poze i osvjetljenja u odnosu na [31,32]. Prema navedenom, umjesto Haarovih značajki ili obične razlike intenziteta susjednih piksela, ovdje se koristi normalizirana razlika piksela, te se nad njom evaluira kvadratna funkcija, čijim se kvantificiranjem i definiranjem raspona pragova generira stablo odluke. Ovako treniran detektor je otporan na zamućenja lica, promjenu svjetline, poza, te prekrivanja. Brzina rada je neznatno manja od [31]. Algoritam je treniran i testiran na bazi [23], odnosno [24], a rezultati detekcije lica su dosad najbolji poznati u odnosu na sve razmatrane algoritme do 2016. godine.

V. BAZE ZA UČENJE I TESTIRANJE

CMU MultiPIE baza [23] sastoji se od 750 000 slika 337 različitih ljudi, snimanih pod različitim kutevima pogleda, osvjetljenjem, te izrazima lica. Baza sadrži slike prednjih lica, lica rotiranih ulijevo ili udesno za 45° i slike profila lica. Za prednja lica i ona rotirana do 45° postoji zapis od 68 točaka značajki lica (točke oko obrva, očiju, nosa, usta, rubovi lica). Za profilne slike lica definirano je 39 točaka. Korištena je pri

učanju algoritma [22]. LFW (Labeled Faces in the Wild) [47] baza sadrži označena lica s imenima ljudi, a služi za raspoznavanje ljudi u nekontroliranim uvjetima. S nešto više od 13 000 slika, te 1680 različitih ljudi, može se prilagoditi za određivanje poze lica [22], pa je baza kreirana u dotičnom radu nazvana AFW (Annotated Faces in the Wild), s tim da je nadopunjena dodatnim slikama sa <http://www.flickr.com>. Iako izvorni LFW nudi djelomična prekrivanja lica i različita svojstva osvjetljenja, na svakoj slici se nalazi samo jedno lice. S druge strane, AFW sadrži slike više lica osoba snimljenih iz različitih poza i s djelomičnim prekrivanjima. FDDB (Face Detection Data Set and Benchmark) baza [24] sadrži oko 5170 lica na slikama koje nisu napravljene u kontroliranim uvjetima (slika 7). Ova baza predstavlja noviji standard testiranja detektora lica.



Slika 7. Primjer slike iz FDDB baze s označavanjem lica. (Preuzeto iz [1]).

Veliki problem kod evaluacije algoritama je nekonzistentnost u označavanju bazičnih pozicija lica (engl. *ground truth*). Nije posve jasno zašto neka lica nisu označena iako su jasno vidljiva. Ove nadopune u označavanju izvornih pozicija i problemi oko nekonzistentnosti izneseni su u [1].

ZAKLJUČAK

Robusna detekcija lica vrlo je kompleksno područje iz domene računalnog vida i raspoznavanja uzoraka. Tijekom višegodišnje istraživanja razvijeni su brojni pristupi kako bi se uspješnost detekcije lica pokušalo dovesti do razine čovjeka. Sasvim sigurno se pokazuje da tehnike dubokog učenja predstavljaju praktički najbolji izbor među algoritmima strojnog učenja. Ipak, zbog nedostatka dovoljnog broja slika za učenje i relativne zatvorenosti arhitekture, često metode temeljene na jačanju klasifikatora (engl. *Boosting*) i regresiji sa stablima odluka uz korištenje integralnih značajki (boja, gradijenti, linearne komponente) pokazuju vrlo kompetitivne rezultate u ovom području [1, 25]. Nedavni radovi u domeni detekcije pješaka [52] i [53] pokazuju da spomenuti pristupi učenja postaju trenutno standard u široj domeni objektno detekcije. Ipak, još uvijek postoji veliki raskorak između učinka čovjeka i računala u ovoj domeni, zbog čega se trenutno automatska detekcija lica ne može smatrati dovoljno dobrom za primjenjenu deidentifikaciju ljudi na složenoj sceni. Ipak, razvojem saznanja o algoritmima učenja, radu mozga i izradom kompetitivnih baza, ovaj raskorak će biti sve manji.

LITERATURA

- [1] M. Matthias, R. Benenson, M. Pedersoli, L. Van Gool, "Face Detection without Bells and Whistles", *IEEE Proceedings of ECCV*, pp.720-735, 2014.
- [2] M.H. Yang, D.J.Kriegman, N. Ahuja, "Detecting Faces in Images. A Survey", *IEEE PAMI*, vol.24, no.1, 34-58, 2002.
- [3] G. Yang, T.S. Huang, "Human Face Detection in Complex Background", *Pattern Recognition*, vol. 27, no.1, pp.53-63, 1994.
- [4] C. Kotropoulos, I. Pitas, "Rule-Based Face Detection in Frontal Views", *IEEE Proceedings of International Conference on Acoustics, Speech and Signal Processing*, vol.4, pp.2537-2540, 1997.
- [5] R.M. Haralick, K. Shanmugam, I. Dinstein, "Texture Features for Image Classification", *IEEE Transactions on Systems, Man and Cybernetics*, vol.3, no.6, pp.610-621, 1973.
- [6] Y. Dai, Y. Nakano, "Face-Texture Model Based on SGLD and Its Application in Face Detection in a Color Space", *IEEE Proceedings of First International Workshop Automatic Face and Gesture Recognition*, pp.238-242, 1995.
- [7] C.C. Han, H.Y.M. Liao, K.C.Yu, L.H. Chen, "Fast Face Detection via Morphology-Based Pre-Processing", *IEEE Proceedings of Ninth International Conference on Image Analysis and Processing*, pp.469-476, 1998.
- [8] T.S. Jebara, A. Pentland, "Parameterized Structure from Motion 3D Adaptive Feedback Tracking of Faces", *IEEE Proceeding of CVPR*, pp.144-150, 1997.
- [9] T.S.Jebara, K. Russel, A. Pentland, "Mixtures of Eigenfeatures for Real-Time Structure from Texture", *IEEE Proceedings of ICCV*, pp. 128-135, 1998.
- [10] J.C.Terrillon, M. David, S.Akamatsu,"Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and Invariant Moments", *IEEE Proceedings of ICPR*, pp.1350-1355, 1998.
- [11] T. Sakai, M. Nagao, S. Fujibayashi, "Line Extraction and Pattern Detection in a Photograph", *IEEE Pattern Recognition*, vol.1, pp. 233-248, 1969.
- [12] I. Craw, H. Ellis, J. Lishman, "Automatic Extraction of Face Features", *IEEE Pattern Recognition Letters*, vol. 5, pp. 183-187, 1987.
- [13] I. Craw, D.Tock, A. Bennett, "Finding Faces Features", *IEEE Proceedings of Second ECCV*, pp.92-96, 1992.
- [14] P. Sinha, "Object Recognition via Image Invariants: A Case Study", *Investigative Ophthalmology and Visual Science*, vol. 35, no.4, pp.1735-1740, 1994.
- [15] Y.H. Kwon, N. da Vitoria Lobo, "Face Detection Using Templates", *IEEE Proceedings of ICPR*, vol.1, pp.764-767, 1994.
- [16] T. Agui, Y. Kokubo, H. Nagashashi, T. Nagao, "Extraction of Face Recognition from Monochromatic Photographs Using Neural Networks", *IEEE Proceedings of Second International Conference on Automation, Robotics and Computer Vision*, vol.1, pp.18.8.1-18.8.5, 1992.
- [17] H.A. Rowley, S. Baluja and T. Kanade, "Neural Network – Based Face Detection", *IEEE Proceedings of CVPR*, vol.1, pp.203-208 ,1996.
- [18] H.A.Rowley, S. Baluja, T. Kanade, "Rotation Invariant Neural Network-Based Face Detection", *IEEE Proceedings of CVPR*, pp.38-44, 1998.
- [19] K. K. Sung, T. Poggio, "Example-Based Learning for View-Based Human Face Detection", *IEEE PAMI*, vol.20, no.1, pp.39-51, 1998.
- [20] S.H. Lin, S.Y.Kung, L.J. Lin, "Face Recognition/Detection by Probabilistic Decision-Based Neural Network", *IEEE Transactions on Neural Networks*, vol. 8, no.1, pp.114-132, 1997.
- [21] C. Zhang, Z. Zhang, "Improving Multiview Face Detection with Multi-Task Deep Convolutional Neural Networks", *IEEE WACV*, pp.1036-1041, 2014.
- [22] X. Zhu, D.Ramanan, "Face Detection, Pose Estimation and Landmark Localization in the Wild", *IEEE Proceedings of CVPR*, vol. 1, pp. 2879-2886, 2012.
- [23] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, "Multi-PIE", *Elsevier Journal of Image and Vision Computing*, vol.24, issue:5, pp.807-813, 2010.
- [24] V. Jain, E. Learned-Miller, "FDDB: A Benchmark for Face Detection in Unconstrained Settings", *Technical Report UM-CS-2010-009*, Dept. of Computer Science, University of Massachusetts, Amherst, 2010.
- [25] S. Liao, A. K. Jain, S. Z. Li, "A Fast and Accurate Unconstrained Face Detector", *IEEE PAMI*, vol. 38, no. 2, pp. 211-223, 2016.
- [26] H.Schneiderman, T. Kanade, "A Statistical Method for 3D Object Detection Applied to Faces and Cars, *IEEE CVPR*, vol. 1, pp. 746-751, 2000.
- [27] T. Rikert, M. Jones, P. Viola, "A Cluster-Based Statistical Model for Object Detection", *IEEE Proceeding of ICCV*, vol.2, pp. 1046-1053, 1999.
- [28] J. Friedman, T. Hastie, R. Tibshirani, "Additive Logistic Regression:A Statistical View of Boosting", *Technical Report*, Dept. of Statistics, Stanford University, 1998.
- [29] R. E. Schapire, Y. Singer, "Improved Boosting Algorithms Using Confidence-Rated Predictions", *Journal of Machine Learning*, vol. 37, no.3, pp.297-336, Kluwer Academic Publishers, 1999.
- [30] S. Li, L. Zhu, Z. Zhang, A. Blake, H. Zhang, H. Shum, "Statistical Learning of Multi-View Face Detection", *IEEE Proceedings of ECCV*, pp.67-81, 2002.
- [31] P. Viola, M. Jones, "Rapid Object Detection Using Boosted Cascade of Simple Features", *IEEE Proceedings of CVPR*, vol. 1, pp. 511-518, 2001.
- [32] M. Jones, P.Viola, "Fast Multi-View Face Detection", *Mitsubishi Electric Research Laboratories report TR 2003-96*, 2003.
- [33] B. Fröba, A. Ernst, "Fast Frontal-View Face Detection Using a Multi-Path Decision Tree", *Proceedings of Audio and Video Based Biometric Person Authentication*, pp.921-928, 2003.
- [34] C. Huang, H. Ai, Y. Li, S. Lao, "Learning Sparse Features in Granular Space for Multi-View Face Detection", *IEEE International Conference on Automatic Face and Gesture Recognition*, pp.401-406, 2006.
- [35] T. Mita, T. Kaneko, O. Hori, "Joint Haar-Like Features for Face Detection", *IEEE Proceedings of ICCV*, pp.1619-1628 ,2005.
- [36] S. Baluja, M. Sahami, H.A. Rowley, "Efficient face orientation discrimination", *IEEE Proceedings of ICIP*, pp.589-592 ,2004.
- [37] J. Meynet, V. Popovici, J.P. Thiran, "Face Detector with Boosted Gaussian Features", *IEEE Pattern Recognition*, vol. 40, no.8, pp. 2283-2291, 2007.
- [38] S. Yan, S. Shan, X. Chen, W. Gao, "Locally assembled binary (LAB) Feature with Feature-Centric Cascade for Fast and Accurate Face Detection", *IEEE Proceedings of CVPR*, pp.1-7 ,2008.
- [39] C. Zhang, Z. Zhang, "A Survey of Recent Advanes in Face Detection", *Microsoft Technical Report MSR-TR-2010-66*, July 2010.
- [40] E. Osuna, R. Freund, F. Girosi, "Training Support Vector Machines: An Application to Face Detection", *IEEE Proceedings of CVPR*, pp.130-136, 1997.
- [41] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, T. Poggio, "Pedestrian Detection Using Wavelet Templates", *IEEE Proceedings of CVPR*, pp. 193-199, 1997.
- [42] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection", *IEEE CVPR*, vol.1, pp.886-893, 2005.
- [43] P. Fezlschwald, R. Girschick, D.McAllister, D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models", *IEEE PAMI*, vol.32, no.9, pp. 1627-1645, 2010.
- [44] P. Ott, M. Everingham, "Shared Parts for Deformable Part-Based Models", *IEEE Proceedings of CVPR*, vol. 1, pp.1513 - 1520 , 2011.
- [45] Y. Yang, D. Ramanan, "Articulated Pose Estimation with Flexible Mixtures of Parts", *IEEE Proceedings of CVPR*, vol.1, pp.1385-1392., 2011.
- [46] P.N. Belhumeur, D.W. Jacobs, D. Kriegman, N. Kumar, "Localizing Parts of Faces Using a Consensus of Exemplars", *IEEE Proceedings of CVPR*, vol. 1, pp. 545-552, 2011.
- [47] E. Learned-Miller, G. Huang, A. R. Chowdhury, H. Li, G. Hua, "Labeled Faces in the Wild:A Survey", <http://vis-www.cs.umass.edu/lfw/>
- [48] S. Walk, N. Mayer, K. Schindler, B. Schiele, "New Features and Insights for Pedestrian Detection", *IEEE CVPR*, pp. 1030-1037, 2010.

- [49] G. Galdi, A. Prati, R. Cucchiara, "Multistage Particle Windows for Fast and Accurate Object Detection", IEEE PAMI, vol. 34, no.8, pp.1589-1604, 2012.
- [50] C.H. Lampert, M.B. Blaschko, T Hoffman, "Beyond Sliding Windows: Object Localization by Efficient Subwindow Search", IEEE Proceedings of CVPR, vol.1, pp.1-8 ,2008.
- [51] P. Dollar,Z. Tu, P. Perona, S. Belongie, "Integral Channel Features", IEEE Proceedings of BMVC, vol.1,2009.
- [52] S. Zhang, C. Bauckhage, A.B.Cremers, "Informed Haar-like Features Improve Detection", IEEE Proceedings of CVPR, vol.1, pp. 947-954, 2014.
- [53] S. Zhang, R. Benenson, B. Schiele, "Filtered Channel Features for Pedestrian Detection", IEEE Proceedings of CVPR, vol.1, pp.1751-1760, 2015.