

Detekcija ljudi na slikama primenom HOG-LBP obeležja i SVM klasifikatora

Branko Brkljač, Fakultet tehničkih nauka, Novi Sad

Sadržaj — U radu su prikazane teorijske osnove i praktične performanse metoda za detekciju ljudi na slikama koji se zasniva na izdvajanju karakterističnih HOG-LBP obeležja iz slike. Obeležja se formiraju kombinovanjem obeležja dobijenih na osnovu histograma orijentisanih gradijenata (HOG) i lokalnih binarnih oblika (LBP). Detekcija ljudi se vrši primenom support vector machine (SVM) klasifikatora. Metod je implementiran u programskom jeziku C++ uz korišćenje OpenCV biblioteke, namenjene razvoju aplikacija za kompjutersku viziju.

Ključne reči — detekcija ljudi, histogrami orijentisanih gradijenata (HOG), lokalni binarni oblici (LBP), Support vector machine (SVM) klasifikator, OpenCV.

I. UVOD

Tema rada je ispitivanje delotvornosti metoda za detekciju ljudi u uspravnom, stojećem, položaju na digitalnim slikama koji se zasniva na izdvajanju HOG-LBP obeležja iz slike i korišćenju SVM klasifikatora. Metod je u prvobitnom obliku predložen u radu [1], a kasnije je proširen i poboljšan u [2]. U radu [3] izvršena je detaljna analiza teorijskih osnova i programska realizacija navedenog metoda.

Postojeće primene obuhvataju bezbednosne sisteme za video nadzor i analizu ponašanja, detekciju pešaka i drugih objekata namenjene učesnicima u saobraćaju i sisteme za analizu slika i multimedijalnih sadržaja.

Problem detekcije može se posmatrati kao najjednostavniji slučaj klasifikacije objekata ili događaja sa samo dve kategorije, odnosno klase, [4]. U slučaju detekcije ljudi na slikama pod klasama se podrazumevaju grupe slika koje sadrže ljude ili slike pozitiva i grupe slika bez ljudi ili slike negativna. Ispravna klasifikacija podrazumeva detekciju ljudi u slici u slučaju slike pozitiva, odnosno potvrdu klasifikatora da u slici negativna nema ljudi. Stoga se ukupan rezultat klasifikacije opisuje sa dve vrednosti kojima se izražava prosečan broj ispravno klasifikovanih slika pozitiva i slika negativna, ili na ekvivalentan način sa vrednostima koje izražavaju prosečan broj pogrešno klasifikovanih pozitiva i negativna. Navedene vrednosti predstavljaju prosečan broj promašaja i lažnih pozitiva na korišćenom test skupu i koriste se za ocenjivanje i poređenje performansi klasifikatora.

Kako bi se označilo izdvajanje informacija iz slike, koje

su u konkretnom slučaju od interesa za klasifikaciju, uvodi se pojam obeležja, kao istaknute kvalitativne ili kvantitativne osobine koja na određeni način karakteriše sliku ili neki njen deo. Moguće je istovremeno uzeti u obzir više različitih obeležja, kada se govori o kombinaciji obeležja koja karakteriše dati oblik, odnosno vektoru obeležja koji pripada nekom višedimenzionalnom vektorskom prostoru obeležja. Pitanje izbora, pripreme i dobijanja dobrih obeležja je ključno i u najvećoj meri određuje ukupne performanse sistema za detekciju.

Kao posledica velike raznovrsnosti i promenljivosti slika na kojima se pojavljuju ljudi, detekcija ljudi u slici predstavlja izazovan zadatak i zahteva konstruisanje robustnog obeležja i dobro obučenog klasifikatora.

Velika raznovrsnost potiče od velikog broja položaja koje ljudsko telo može zauzeti, različitog izgleda i načina oblačenja, složenosti i karakteristika scene, velikog opsega mogućih vrednosti osvetljaja, prisustva i stepena okluzija, pomeranja ljudi, kamere i objekata u pozadini.

Ono što je karakteristično i zajedničko za sve te slike je prisustvo značajnih ivica koje odgovaraju ljudskim udovima i koje su, lokalno gledano, uglavnom slično raspoređene u prozoru u kome se vrši analiza. Taj aspekt je iskorišćen kao ideja za formiranje HOG obeležja.

II. IZDVAJANJE HOG-LBP OBELEŽJA

Osnovna ideja za primenu HOG-LBP obeležja na problem detekcije ljudi u ovom radu proistekla je iz rada Wang-a i dr. [2], u kome je prvi put predloženo korišćenje kombinovanog HOG-LBP obeležja, kao i jedan mogući način rešavanja problema detekcije u slučaju delimičnog zaklanjanja, okluzije čovekove siluete nekim drugim objektom u slici. Iako je i ovaj drugi aspekt, predložen u [2], veoma značajan, ovde neće biti razmatran.

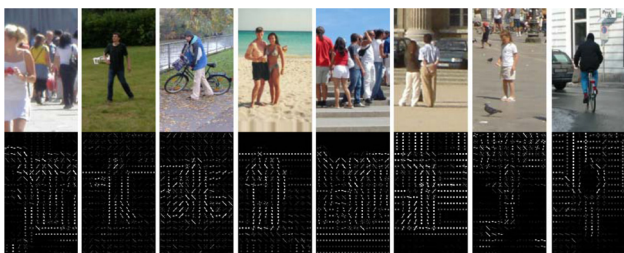
Ideja o HOG obeležju kao dobrom izboru za detekciju ljudi potekla je iz rada Dalal-a i Triggs-a [1] u kome je uspešno korišćeno izdvajanje HOG obeležja u kombinaciji sa SVM klasifikatorom i izvršeno sveobuhvatno poređenje sa do tada postojećim metodama za detekciju ljudi, koje se zasnivaju na nekim drugim kombinacijama izbora obeležja i klasifikatora.

Rezultati koje je HOG pokazao, na do tada najčešće korišćenom skupu slika za obuku i testiranje klasifikatora, bili su gotovo bez greške, pa je formiran izazovniji INRIA trening/test skup slika [5], na kome su rezultati uglavnom bili za red veličine bolji u odnosu na ranije.

Dodavanje LBP obeležja u [2], kao izuzetno dobrog deskriptora tekture, pokazalo je još bolje rezultate i uz predloženi način za rešavanje spornih situacija nastalih

Branko Brkljač, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, Trg Dositeja Obradovića 6, 21000 Novi Sad, Srbija, član IEEE (telefon: +381642472212, e-mail: [branco1mail\(at\)gmail.com](mailto:branco1mail(at)gmail.com)). Rad je proistekao iz master rada čiji mentor je bio dr Vladimir Crnojević, vanr. prof.

zaklanjanjem, prema navodima autora [2], trenutno predstavlja rešenje sa najboljim performansama za detekciju ljudi na pomenutom trening/test skupu, Sl. 1.



Sl. 1. Nekoliko primera slika pozitivna iz korišćenog skupa, uz generisane slike histograma orijentisanih gradijenata

Kao referentni klasifikator za poređenje performansi u [1] i [2] je korišćen linearni SVM implementiran u SVMlight i LibSVM biblioteci.

U ovom radu je korišćen isti skup slika i linearni SVM klasifikator dat u OpenCV biblioteci otvorenog koda, [6].

Izbor linearnog SVM je opravdan zbog visoke dimenzionalnosti korišćenog HOG-LBP vektora obeležja, što izostavlja potrebu za dodatnim nelinearnim preslikavanjem u prostor veće dimenzije koje bi odgovaralo nelinearnom SVM. Pored toga SVM klasifikator, zbog dobrih performansi i efikasnosti, spada u vodeću grupu klasifikatora.

A. Histogrami orijentisanih gradijenata (HOG)

Pod histogramom orijentisanih gradijenata podrazumeva se vektor obeležja sastavljen od određenog broja diskretnih histograma vrednosti gradijenata, dobijenih obradom slike nad ćelijama u prozoru i normalizacijom po blokovima sastavljenim od više ćelija.

Gradijent u digitalnoj obradi slike predstavlja uopštenje gradijentnog operatora na diskretni signal slike i vektor koji pokazuje pravac i smer najbržeg porasta osvetljaja u svakom pikselu, (1). Može se predstaviti u vidu dve matrice, tačnije slike modula ili vrednosti gradijenata G i slike uglova θ .

$$\nabla f(x,y) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right]^T \Leftrightarrow G = |\nabla f| = \sqrt{G_x^2 + G_y^2}, \theta = \arctg \frac{G_y}{G_x} \quad (1)$$

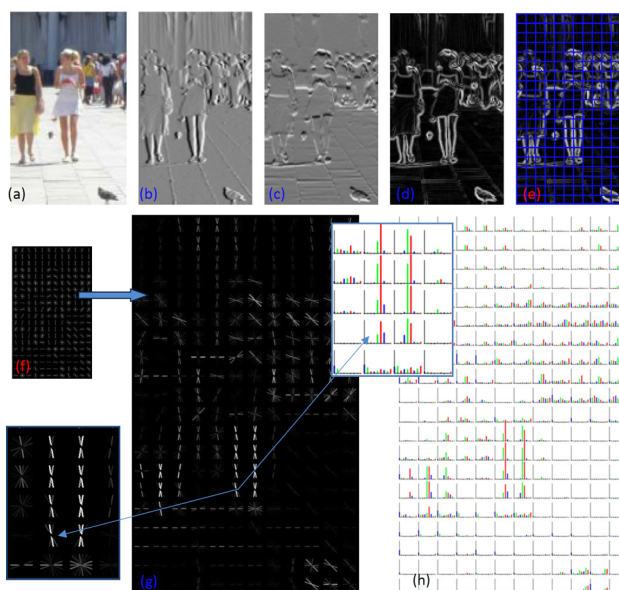
Vrednosti gradijenta odgovaraju združenim promenama osvetljaja duž horizontalnog, Sl.2.b) i vertikalnog, Sl.2.c), pravca u posmatranim pikselima, a slika modula gradijenata, Sl.2.d), značajnim ivicama u slici. Orijentacija gradijenta u svakom pikselu je sadržana u slici uglova.

Cilj izračunavanja HOG obeležja je dobiti diskretne histograme orijentacije gradijenata koji odgovaraju ćelijama. Stoga je prvo potrebno izvršiti diskretizaciju mogućih vrednosti pravaca gradijenata, a zatim njihovo „prebrojavanje“ po ćelijama. Postavlja se pitanje na koji način odabrati dimenzije ćelije i koliko mogućih pravaca gradijenata treba dozvoliti.

Dimenzije ćelija su uslovljene dimenzijom prozora u kome se vrši analiza, ali i odnosom dimenzija prozora i veličine čoveka u slici, izraženim u pikselima. U [1] je utvrđeno da za dati skup slika, najbolje rezultate daju ćelije dimenzija 6×6 ili 8×8 piksela, zavisno od veličine bloka za normalizaciju histograma (3×3 ili 2×2 ćelije).

Nakon definisanja dozvoljenih pravaca koje gradijent može zauzeti u svakom pikselu se vrši „projekcija“ gradijenta na dva najbliža, susedna pravca. Pod projekcijom se podrazumeva preraspodela vrednosti modula gradijenta na dozvoljene, diskretne pravce u odnosu koji odgovara ugaonoj udaljenosti gradijenta od svakog od njih, tačnije vrši se interpolacija vrednosti gradijenta. U [1] je utvrđeno da u slučaju neoznačenih pravaca ($0^\circ \leq \theta \leq 180^\circ$) nema potrebe za više od devet, jer dalje povećanje broja pravaca ne utiče značajno na krajnji rezultat klasifikacije.

Dobijeni histogrami, na primeru jedne slike iz trening/test skupa, prikazani su grafički na Sl.2.g) i Sl.2.h) pomoću posebnog programa kreiranog u [3].



Sl. 2. Ilustracija dobijanja histograma orijentisanih gradijenata na primeru slike iz korišćenog INRIA skupa

Histogrami se izračunavaju „prebrojavanjem“ koliko kojih od devet mogućih pravaca ima u svakoj ćeliji. Pri tome se pod prebrojavanjem podrazumeva sumiranje vrednosti modula gradijenata koji odgovaraju posmatranom pravcu, po svim pikselima u ćeliji za svaki od devet diskretnih pravaca. Karakteristika dobijenih histograma je da nemaju celobrojne vrednosti.

Vrednosti histograma opisuju prisustvo dominantnih pravaca u posmatranoj ćeliji i na taj način, lokalno gledano, prisustvo ivica u slici za više susednih ćelija.

Sa Sl. 2 može se uočiti da su najznačajniji pravci kod čoveka u stojećem položaju ivice na prelazu između čoveka i pozadine, odnosno pravci koji odgovaraju ekstremitetima, glavi i ramenima. Karakteristične vrednosti ovih histograma (sa izraženim vertikalnim pravcima kao na uvećanim delovima Sl. 2) i položaj ćelija kojima pripadaju u prozoru čine osnovu za uspešnu primenu HOG obeležja na detekciju ljudi u slici.

Pošto je dobijenih devet matrica gradijenata (za svaki od diskretnih pravaca) dosta retko, u [2] je predloženo da se pre formiranja histograma izvrši prostorno filtriranje matrica gradijenata pomoću posebne konvolucione maske.

Vrednosti histograma između prostorno bliskih ćelija mogu značajno da variraju, kao posledica lokalnih

varijacija u osvetljaju i kontrastu, pa je neophodna lokalna normalizacija vrednosti histograma za svaku od ćelija.

Normalizacija se vrši po blokovima koji obuhvataju nekoliko bliskih ćelija, pri čemu je bitno da postoji preklapanje između blokova u smislu da pojedine ćelije mogu ulaziti u sastav više blokova. Preklapanjem blokova i normalizacijom vrednosti histograma postiže se da svaki histogram više puta ulazi u sastav konačnog HOG vektora obeležja. U [1] je analiziran uticaj geometrije i stepena preklapanja blokova na performanse klasifikatora.

Za pomenuti trening/test skup slika, uzimajući u obzir rezultate dobijene u [1] i [2], za potrebe ovog rada korišćeni su sledeći parametri HOG obeležja: dimenzije prozora 64×128 piksela, dimenzije ćelije 8×8 piksela, diskretizacija gradjenata na 9 pravaca u opsegu od 0 do 180 stepeni, pravougaoni blokovi dimenzija 2×2 ćelije sa preklapanjem u širini jedne ćelije i normalizacija vrednosti histograma po blokovima pomoću L_2 norme.

Kao rezultat dobija se: 105 blokova, sastavljenih od 4 histograma sa po 9 vrednosti, koji se upisuju red po red u konačni vektor obeležja. Dimenzija konačnog HOG vektora obeležja je: $105 \cdot (9 \cdot 4) = 3780$ vrednosti.

B. Lokalni binarni oblici (LBP)

Pod lokalnim binarnim oblicima se podrazumeva obeležje kojim se opisuje lokalna struktura vrednosti piksela u jednokanalnoj, sivjoj slici, odnosno tekstura u blizini posmatranog piksela.

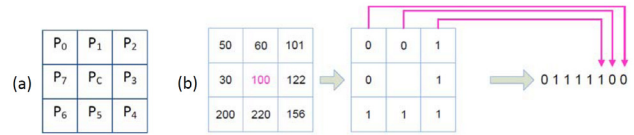
Tekstura podrazumeva karakterističan raspored vrednosti piksela u lokalnoj oblasti i izračunava se poređenjem vrednosti centralnog P_c i okolnih piksela P_i . Pri tome je neophodno definisati šta se podrazumeva pod okolinom, njenu geometriju i način poređenja vrednosti. U ovom radu je korišćena pravougaona geometrija sa okolnim pikselima na jediničnom rastojanju, Sl. 3a).

Za meru sličnosti prilikom poređenja uglavnom se uzima razlika vrednosti okolnih i centralnog piksela, koja se kvantizuje kao 0 ili 1. Ako se kvantizovana razlika vrednosti piksela označi sa $a(\cdot)$ binarni oblik $LBP_{n,r}$ koji opisuje teksturu je:

$$LBP_{n,r} = \sum_{i=0}^{n-1} a(P_i - P_c) 2^i, \quad a(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

gde n označava broj okolnih piksela na radijusu r koji se porede sa centralnim. Pod uniformnim oblicima $LBP_{n,r}^u$ se podrazumevaju oni kod kojih broj prelaza 0-1, odnosno 1-0, ne prelazi unapred definisan broj u , dok se svi ostali smatraju neuniformnim. Na Sl. 3b) preuzetaj iz [2] je prikazan primer uniformnog oblika $LBP_{8,1}^2$, dok bi oblik 01010100 odgovarao uniformnom obliku $LBP_{8,1}^6$.

Neosetljivost LBP obeležja na monotone promene nivoa osvetljaja u slici i računaska efikasnost uz veliku diskriminativnost ga čini veoma pogodnim za detekciju ljudi, [2]. Izračunavanje LBP obeležja se vrši na sličan način kao u prethodnom poglavlju, analizom slike u prozoru određenih dimenzija i formiranjem normalizovanih histograma uniformnih oblika po ćelijama.



Sl. 3. Geometrija pravougaonog binarnog oblika i primer uniformnog oblika sa ne više od dva prelaza iz [2]

LBP obeležje korišćeno u radu predstavlja vektor obeležja sastavljen od vrednosti diskretnih histograma binarnih oblika za svaku od ćelija. Histogrami se dobijaju izračunavanjem binarnih oblika za svaki od piksela u ćeliji i njihovim prebrojavanjem. Pri tome uniformni oblici predstavljaju posebne vrednosti u histogramu, dok se svi ostali oblici ubrajaju u jednu istu klasu, kojoj odgovara poslednja vrednost u histogramu. Broj vrednosti u histogramu zavisi od izbora parametara n i u .

Za korišćeni oblik $LBP_{8,1}^2$, dobija se histogram sa 58 uniformnih vrednosti, pa je ukupna dimenzija generisanog vektora obeležja za prozor dim. 64×128 i ćelije dim. 16×16 piksela, nakon normalizacije vrednosti histograma po ćelijama sa L_2 normom: $(58 + 1) \cdot 8 \cdot 4 = 1888$.

Konačni vektor obeležja, koji se formira za svaku od slika iz trening/test skupa [5] i prosleđuje SVM klasifikatoru, nastaje spajanjem na red dobijenih HOG i LBP vektora i njegova dimenzija je: 5668 vrednosti.

III. SVM KLASIFIKATOR

SVM ili mašina sa nosećim vektorima predstavlja specifičan metod nadgledanog učenja za određivanje granica regiona odlučivanja linearnog klasifikatora ili rešavanje regresionih problema estimacije vrednosti funkcija, proizašao iz statističke teorije učenja čiji je glavni tvorac V. Vapnik, [7].

Zasniva se na konceptu direktnog učenja na osnovu prikupljenih podataka u visokodimenzionalnom prostoru obeležja, bez potrebe za poznavanjem ili estimacijom gustina raspodela verovatnoća.

Veliki broj dimenzija podrazumeva problem dimenzionalnosti, koji se odlikuje time da je broj uzoraka često nedovoljno velik i da je prostor obeležja i suviše prazan, što utiče na mogućnost generalizacije dobijenog rešenja na nove, slučajne uzorke koji nisu bili obuhvaćeni konačnim trening skupom, a pojavljuju se pri radu.

Stoga je jedna od ideja autora SVM bila prevazilaženje ovog problema uvođenjem pojma strukturalnog rizika koji obuhvata verovatnoću greške klasifikatora na trening skupu (empirijski rizik) i njegovu sposobnost uspešne klasifikacije u realnom radu. Minimizacijom strukturalnog rizika dobija se optimalni klasifikator.

U [7] i [8] je pokazano da linearni SVM predstavlja optimalni klasifikator u smislu da je klasifikator sa maksimalnom marginom razdvajanja, odnosno da maksimizacija rastojanja uzoraka od optimalne hiperravnine razdvajanja određene linearnom diskriminantnom funkcijom (3) odgovara minimizaciji strukturalnog rizika.

$$g(\mathbf{x}) = \hat{\mathbf{w}}^T \mathbf{x} + \hat{w}_0 = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i^T \mathbf{x}_i + \frac{1}{N_{SV}} \sum_{s=1}^{N_{SV}} \left(\frac{1}{y_s} - \hat{\mathbf{w}}^T \mathbf{x}_s \right) \quad (3)$$

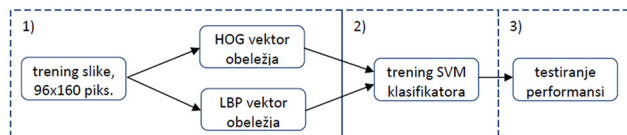
Jednačina $g(x)=0$ odgovara hiperravni pa znak funkcije (3) definiše pravilo odlučivanja, pri čemu je n ukupan broj uzoraka iz trening skupa. Funkcija (3) je u potpunosti određena težinskim vektorom normale hiperravni \hat{w} i skalarom \hat{w}_0 , odnosno vrednostima N_{SV} nosećih vektora x_i kojima odgovaraju Lagranžovi množitelji $\alpha_i \neq 0$ za $i=1, \dots, N_{SV}$. Ostali trening uzorci kojima odgovaraju $\alpha_i = 0$ ne utiču na položaj hiperravni. Sa y_i su označene diskretne labele klasa pridružene uzorcima iz trening skupa.

Noseći vektori x_i i njima pridruženi Lagranžovi množitelji se dobijaju nakon završene obuke klasifikatora na trening skupu i predstavljaju rešenje statičkog problema kvadratnog programiranja, [8], kojim se maksimizuje margina razdvajanja. Postavljeni optimizacioni problem sa ograničenjima u vidu linearnih nejednakosti, koje diktira svaki od uzoraka iz trening skupa, se rešava numerički primenom nekog od iterativnih algoritama za obuku SVM.

Pretpostavka za postojanje rešenja u vidu funkcije (3) je da su klase linearno razdvojive što u praksi najčešće nije slučaj pa se postavljeni zahtevi relaksiraju uvođenjem linearnog SVM klasifikatora meke margine, čija implementacija je korišćena u ovom radu.

IV. REZULTATI DETEKCIJE

Programska realizacija klasifikatora je predstavljena na Sl. 4. i sastoji se iz tri faze čiji cilj je dobijanje obučenog SVM klasifikatora i merenje postignutih performansi.



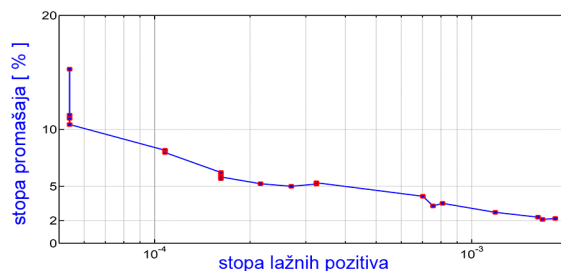
Sl. 4. Opšta skica programa za trening i testiranje, [3]

Ukupan broj slika pozitivna koji je bio na raspolaganju za trening i test klasifikatora je 3542, a ukupan broj slika negativna zavisi od načina njihovog izdvajanja iz „velikih” slika negativna koje su u [5] date u originalnoj, većoj rezoluciji. Pod pretpostavkom da se iz svake velike slike na slučajan način iseca ne više od dvadeset slika negativna, dobija se 18536 slika negativna koje su na raspolaganju za trening i test klasifikatora.

Da bi se povećala pouzdanost dobijenih rezultata prilikom obuke i testiranja vršena je krosvalidacija sa deset particija uzoračkog skupa, pri čemu je devet particija korišćeno za trening, a jedna za test. Promenom težinskih koeficijenata pridruženih klasama u algoritmu za obuku dobijena je radna karakteristika na Sl. 5.

Performanse dobijenog linearnog SVM klasifikatora, bez uvođenja težinskih koeficijenata, pri vrednosti parametra meke margine $C=10$, iznose: 2,17758% promašaja, odnosno 97,82242% pogodaka pri stopi od $18,3451 \cdot 10^{-4}$ lažnih pozitivna i predstavljaju referentne rezultate prikazanog metoda na skupu [5] u ovom radu.

Kao posledica usvojenog načina merenja radnih karakteristika i činjenice da je rezolucija merenja stope



Sl. 5. Radna karakteristika klasifikatora realizovanog u [3] lažnih pozitivna ograničena konačnim brojem negativna u test skupu, na Sl. 5. nije bilo moguće dobiti radne tačke pri veoma malim vrednostima lažnih pozitivna.

V. ZAKLJUČAK

Rezultati klasifikacije dobijeni u ovom radu potvrđuju primenljivost opisanog metoda, predstavljenog u [1] i [2], na problem detekcije ljudi. Kao mogućnost za dalja poboljšanja predlaže se korišćenje novog trening/test skupa i drugačije metodologije merenja performansi, koji su prikazani u [9]. Za primenu dobijenog detektora neophodan je rad u realnom vremenu, pa se kao jedno od rešenja predlaže implementacija metoda na nekoj od platformi sa više procesora.

LITERATURA

- [1] N. Dalal, B. Triggs, „Histograms of Oriented Gradients for Human Detection”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, str. 886-893, 2005.god.
- [2] X. Wang, T.X. Han, S. Yan, „An HOG-LBP Human Detector with Partial Occlusion Handling”, *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, 2009.god.
- [3] B. Brkljač, „Detekcija ljudi na slikama primenom LBP-HOG obeležja i SVM klasifikatora”, Master rad, Fakultet tehničkih nauka, Novi Sad, 2010.god.
- [4] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, 2000.god.
- [5] Institut national de recherche en informatique et automatique (INRIA) Person Dataset, <http://pascal.inrialpes.fr/data/human/>
- [6] Intel Open Source Computer Vision Library (OpenCV) 1.0.01, <http://opencv.willowgarage.com>
- [7] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 2000.god.
- [8] V. Kecman, *Learning and Soft Computing - Support Vector Machines, Neural Networks, Fuzzy Logic Systems*, MIT press, Cambridge MA, 2001.god.
- [9] P. Dollar, C. Wojek, B. Schiele, P. Perona, „Pedestrian Detection: A Benchmark”, *IEEE CVPR*, 2009.god.

ABSTRACT

Work presents theoretical details and applied performances of one method for detection of humans in images based on extraction of characteristic HOG-LBP features from image. Features are formed by combining features obtained from histograms of oriented gradients (HOG) and local binary patterns (LBP). Detection of humans is made by utilizing support vector machine (SVM) classifier Method is implemented in C++ programming language using OpenCV library, designed for development of computer vision software applications.

DETECTION OF HUMANS IN IMAGES BY USING HOG-LBP FEATURES AND SVM CLASSIFIER

Branko Brkljač, Faculty of Technical Sciences, Novi Sad