

Optična razpoznavna znakov v slikah naravnih scen, ki temelji na metodah HOG

Rok Petek¹, Luka Šajn¹

¹ Faculty of Computer and Information Science, University of Ljubljana
rok5ek@hotmail.com, luka.sajn@fri.uni-lj.si

Abstract

Paper presents some promising feature extractions for character recognition in natural scene images. Our implementations are based on three different methods for feature extraction. The basic HOG method, which also underlies the other two methods is one of the most popular feature extraction methods in object detection and character recognition. PHOG method, which is based on HOG, converts the basic HOG algorithm into a pyramid scheme, due to its structure this method is slower than HOG, but more expressive, since it yields larger feature vectors. The third feature extraction method is Co-HOG algorithm, which inherits the good qualities of HOG method, such as invariance to illumination and geometric changes. Co-HOG differs from HOG and PHOG, by its feature representation, where it also captures the spatial relationship of neighboring pixels in order to describe the character more accurately. All three methods were evaluated on different datasets, such as ICDAR, Chars74K and CVL OCR DB.

1 Uvod

Razpoznavna znakov v slikah naravnih scen je specifično področje, kjer tradicionalne metode razpoznavne znakov ne prinesejo željenih rezultatov. Med tradicionalne metode razpoznavne znakov spadajo sistemi, ki razpoznajo znake iz skeniranih dokumentov, obrazcev, položnic, potovalnih listin, bančnih izpiskov itd., torej v primerih kjer so vnaprej znana in predvidena določena pravila, kot je položaj znakov, tip pisave, barva pisave, dobra osvetljenost, razločnost in ostali parametri, ki pripomorejo k bolj natančni razpoznavi.

Znaki v slikah naravnih scen so lahko različnih barv, vsebujejo šum v ozadju, tipi pisave so različni ali pa osvetlitev ni enakomerna. Te nepravilnosti je potrebno odpraviti na robusten način in zajeti strukturo znaka ter najti podobnosti med različnimi oblikami znakov, s čimer lahko razlikujemo znake. Metode razpoznavne znakov v slikah naravnih scen delimo na dva segmenta, na tiste ki sliko predprocesirajo, ji odpravijo šum in sliko binarizirajo ter na tiste ki ne potrebujejo predprocesiranja. Metode, ki temeljijo na histogramu orientacij gradientov (HOG) ne potrebujejo predprocesiranja in obenem ponujajo dovolj visoko hitrost razpoznavne ter visoko natančnost.

Pridobljene značilke znakov smo testirali na različnih podatkovnih zbirkah slik znakov. Uporabili smo sledeče podatkovne zbirke: ICDAR (International Conference on Document Analysis and Recognition) [1], CVL OCR DB [2] in CHARS74K [3].

V zadnjem koraku razpoznavne znakov smo klasificirali slike znakov v pravilne razrede. Klasifikacijo znakov smo razdelili na 52 ter na 26 razredov, kjer smo v slednjem primeru upoštevali invariantnost velikosti znaka. V obeh primerih smo izključili številke in posebne znake. Uporabili smo klasifikacijske metode kot so ANN, SVM, K-NN in NB ter za vsak klasifikacijski model definirali optimalne parametre, s katerimi smo lahko dosegali najvišjo klasifikacijsko točnost.

2 Metode

Na prvem nivoju OCR sistem prejme kot vhod sliko, ki je zaradi narave naših algoritmov za ekstrakcijo značilke normalizirana na velikost 32x32 slikovnih elementov. Vhodno sliko smo najprej pretvorili v matriko, ki vsebuje 3 barvne kanale ter jo pretvorili v sivinsko sliko. V drugem koraku smo iz vhodne slike izločili relevantne podatke s pomočjo različnih algoritmov. V zadnjem koraku smo algoritme za pridobivanje značilke iz znakov ocenjevali na različnih klasifikacijskih algoritmihih.

2.1 HOG

Deskriptor značilke HOG [4] je izračunan na gosti površini prekrivajočih se mrež. Zasnovan je tako, da je robusten pri majhnih spremembah lokacij obrisov na sliki, robusten na orientacijo smeri ter hkrati robusten na velike spremembe osvetlitve in barve na sliki. Metoda HOG [4] prejme kot vhod celotno sliko oz. njeno matriko. Slednja je bila pretvorjena v sivinsko sliko, za doseganje boljših rezultatov smo ji tudi spremenili višino in širino na velikost 32x32 slikovnih elementov. Sliko smo nato normalizirali, s čimer smo odpravili vpliv neenakomerne osvetlitve slike. Pri normalizaciji slike kompresiramo gama kanal (angl. Gamma Compression/Power Law) in barvo, bodisi z logaritmiranjem vsakega barvnega kanala slike ali z izračunom kvadratnega korena celotne matrike slike.

Po normalizaciji slike sledi izračun gradientov po x in y osi. Predhodno je na področju procesiranja slik veljalo, da naj bi se vsako digitalno zajeto sliko predprocesiralo z gausovim fitrom, saj naj bi z glajenjem slike dosegali boljše rezultate. Vendar se v primeru HOG metode izkaže, da za pridobitev gradienta

slike predhodno glajenje ni potrebno oz. se s povečanjem glajenja celo zmanjša klasifikacijska točnost. Z izračunom gradienta zajamemo silhete in obrise ter informacijo o teksturi slike. Gradient slike je definiran s formulo (1).

$$\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) \quad (1)$$

Parcialni odvod slike je generiran iz originalne slike s konvolucijo jedra čez sliko. Za generiranje gradienta s konvolucijo (2) lahko uporabimo preprosto 1-dimenzionalno jedro, kot je razvidno v enačbi (2).

$$\frac{\partial I}{\partial x} = [1, 0, -1] \otimes I, \quad \frac{\partial I}{\partial y} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \otimes I \quad (2)$$

Poleg gradienta slike izračunamo tudi magnitudo gradienta (3). Gradient nam pove, kako močno se slika spreminja v smeri x ali y, magnituda gradienta pa, kako hitro se slika spreminja. Z magnitudo zaznamo amplitudo robov, kjer slikovni elementi nenadoma spremenijo sivinski nivo. Magnituda gradienta se pri metodi HOG aplicira pri grajenju histograma, saj magnitudo uporabimo za uteževanje histograma.

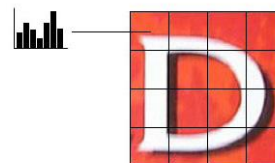
$$\|\nabla I\| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \quad (3)$$

Gradient nam poda tudi informacijo o njegovi smeri, zato moramo izračunati v kateri smeri se slika najbolj spreminja. Orientacijo gradienta izračunamo za gručenje (angl. binning) magnitud v histogramu. Rezultat formule (4) nam vrne vrednosti v radianih med $-\pi$ in π oz. -180° in 180° . Radiane moramo pretvoriti v stopinje ter omejiti njihov razpon na 0° in 180° , saj metoda HOG bolje deluje z nenegativnimi vrednostmi orientacij.

$$\theta = \tan^{-1} \left(\frac{\partial I / \partial x}{\partial I / \partial y} \right) \quad (4)$$

V tretjem koraku metoda HOG razdeli celotno sliko na posamezne celice. Koliko celic naj bi vsebovala slika moramo definirati vnaprej. Sliko smo predhodno normalizirali na velikost 32×32 slikovnih elementov, torej bodo imele vse slike enako število celic. V našem primeru smo sliko razdelili na celice velikost 8×8 slikovnih elementov, torej smo imeli 16 celic v celotni sliki. Na nivoju celic izračunamo 1-dimenzionalni histogram slike. Z definiranim številom celic v sliki sledi lahko zgradimo histogram. Histogram akumuliramo na nivoju vsake celice v sliki (slika 1), celoten histogram je razdeljen na 9 stolpcev oz. na interval od 0° do 180° , kjer vsak stolpec zajema 20° . Za

izračun uteženega histograma smo uporabili magnitudo gradientov, kar pomeni, da je histogram vseboval vrednosti magnitud na soležnih orientacijah gradientov.



Slika 1. Razdelitev slike na posamezne celice v velikosti 8×8 slikovnih elementov. Izgradnja histograma na nivoju celic.

V četrtem koraku moramo celice zgručiti v več med seboj prekrivajočih se blokov. Bloki so enakih velikost, in štejejo 3×3 celice. Notranje celice bodo torej prispevale večim komponentam končnega vektorja značilk. V zadnjem koraku združimo histograme na nivoju blokov, ter jih tudi normaliziramo (5). Z normalizacijo izboljšamo invariantnost algoritma na osvetljevanje, senčenje in kontrast robov. Normalizirane vektorje značilk posameznih blokov nato združimo v 1-dimenzionalni vektor značilk, ki je namenjen za klasifikacijo.

$$H = H / \sqrt{\left(\sum_{i=0}^n H_{(i)}\right)^2 + \varepsilon} \quad (5)$$

Kjer:

$$\varepsilon = 10^{-5}$$

2.2 PHOG

Piramidni histogram orientiranih gradientov [5] je variacija originalnega HOG algoritma. PHOG je primarno namenjen razpoznavi objektov, njegov glavni cilj je prostorska razporeditev oblik, ki bi koristila boljši razpoznavi. Spremenjen PHOG [6], ki smo ga implementirali, temelji na ideji originalne verzije algoritma PHOG in je namenjen razpoznavi znakov v slikah naravnih scen. PHOG metoda vključuje algoritem HOG v prostorski piramidi, njen glavni namen je predstavitev oblike slike in njene prostorske postavitve, tako da je lahko primerjava med dvema oblikama izračunana s SVM klasifikatorjem.

Metoda PHOG za vhod prejme sivinsko sliko, v velikosti 32×32 slikovnih elementov. Vhodno sliko tudi normaliziramo oz. ji uravnamo njeno gamma vrednost, s katero ublažimo neenakomerno osvetljenost. Kot pri metodi HOG, izračunamo tudi magnitudo ter orientacijo gradientov.

Metoda HOG razdeli sliko na več med seboj enakih celic, pri metodi PHOG, pa je porazdelitev celic dinamična in ne statična kot pri HOG metodi, kar pomeni da bo mreža celic z vsakim naslednjim piramidnim nivojem gostejša (6).

$$2^l \times 2^l \quad (6)$$

Kjer:

$$l = \text{piramidni nivo}$$

Algoritem PHOG lahko vsebuje poljubno število piramidnih nivojev, vendar smo se zaradi računske zahtevnosti omejili na 4 nivojsko piramido. Končno dolžino vektorja značilik lahko definiramo z enačbo (7).

$$K \cdot \sum_{l \in L} 4^l \quad (7)$$

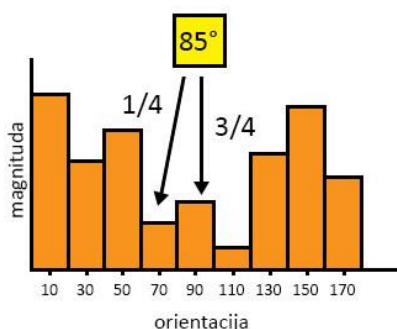
Kjer:

K = število orientacij

L = končno število piramidnih nivojev

l = piramidni nivo

Utežene histograme smo izračunali iz vsake celice na vsakem nivoju. Histograme smo utežili z magnitudo gradientov, kjer smo uporabili bilinearno interpolacijo (slika 2).



Slika 2. Prikaz bilinearne interpolacije med dvema stolpcema histograma, kjer je $\theta = 85^\circ$. Razlika do središča 70 je 15° , razlika do središča stolpca 90 je 5° . Koefficienti za množenje z magnitudo so torej $5/20=1/4$ ter $15/20=3/4$.

Torej, če je bila orientacijska vrednost gradienta 96° , je ta poleg tega, da je prispevala uteževanju v interval med 80° in 100° , prispevala še v bližnji interval oz. stolpce (100° in 120°), saj je vrednost 96° bližje centru desnega stolpca 110° kot centru levega stolpca 70° . Sosednjemu stolpcu prispeva toliko kolikor je zmnožek magnitude z razliko med orientacijsko vrednostjo ter centrom njenega stolpca, deljeno z razliko med centroma stolpcev, kot je ponazorjeno na sliki 2 in z enačbo (8). Svojemu stolpcu prispeva zmnožek magnitude z razliko med to utežjo sosednjega stolpca in vrednostjo 1, kar je razvidno v enačbi (9).

$$H(\theta_2) \leftarrow H(\theta_2) + mag_{x,y} * (\beta / \gamma) \quad (8)$$

$$H(\theta_1) \leftarrow H(\theta_1) + mag_{x,y} * (1 - \beta / \gamma) \quad (9)$$

Kjer:

θ_1 = število orientacij

θ_1 = sosednji interval z najmanjšo razliko orientacije do centra

mag = magnituda gradienta na lokaciji x,y

β = razlika vrednosti orientacije gradienta in njenega centra intervala

γ = razlika med dvema centroma intervalov

V zadnjem koraku metode PHOG sliko razdelimo na več med seboj prekrivajočih se blokov, katere dinamično prilagajamo glede na trenutni nivo piramide. Na prvem nivoju piramide je vsebovana samo ena celica, kar pomeni, da je histogram definiran z enim blokom v velikosti 1×1 , na drugem nivoju z blokom velikosti 2×2 , na tretjem 3×3 in na četrtem 7×7 . Na vsakem piramidnem nivoju se izračuna histograme iz posameznih celic, ter se jih razdeli v več blokov, vektorje iz posameznih blokov nato normaliziramo, normalizirane vektorje pa nato združimo v 1-dimenzionalni vektor značilik.

2.3 Co-HOG

Primarno je Co-HOG [8] namenjen detekciji ljudi, vendar se izkaže kot dober deskriptor za razpoznavo znakov v slikah naravnih scen. Co-HOG [7] temelji na algoritmu HOG, ki je robusten na spremembe osvetlitve ter geometrične spremembe, vendar HOG ne zajema prostorske informacije med sosednjimi slikovnimi elementi. Co-HOG zajema tudi orientacijo gradienta sosednjih slikovnih elementov, zato lahko prikaže več informacij ter izrazi obliko znaka bolj podrobno kot HOG, v katerem se uporablja orientacija gradienta enega slikovnega elementa.

Algoritem Co-HOG prejme za vhod barvno ali sivinsko sliko v velikosti 32×32 slikovnih elementov. V primeru barvne slike, slednjo pretvorimo v sivinsko. Sivinski kanal nato normaliziramo s kvadratnim korenem matrike slike, kot pri metodi HOG in PHOG. Nad sliko izračunamo gradient (1), kjer uporabimo Sobel-ov operator. Implementacija metode Co-HOG vključuje tudi izračun magnitude in orientacije gradientov. Orientacije gradientov so omejene na interval od 0° do 180° , kjer so vrednosti nenegativne oz. nepredznačene. Orientacije so kvantizirane v 9 intervalov.

Co-HOG razdeli sliko na več blokov, vendar ti niso med seboj prekrivajoči, kot pri metodi HOG in PHOG. Matrika slike je razdeljena na 16 blokov, kjer posamezne celice štejejo 8×8 slikovnih elementov. Metoda zajame prostorske informacije s štejem frekvence so-pojavnosti orientiranih gradientov med pari slikovnih elementov, pri čemer so shranjene relativne lokacije. Sredinski element je trenutni element, ki ga obravnavamo, njegovi sosednji elementi so oddaljeni z zamikom dveh elementov. Vsak sosednji element tvori orientacijski par s sredinskim elementom in ustrezno glasuje v so-pojavitveno matriko.

Za izgradnjo histograma, metoda Co-HOG uporablja so-pojavitveno matriko, s katero generira 3-dimenzionalni histogram. So-pojavitvena matrika je kvadratna matrika in je definirana s številom orientacijskih intervalov oz. stolpcev (angl. bins). V našem primeru je dimenzija so-pojavitvene matrike $9 \times 9 \times n$, kjer n predstavlja višino posameznega stolpca in je definirana s formulo (10).

$$H_{x,y}(i, j) = \sum_{(p,q) \in B} \begin{cases} 1 & \text{če } O(p, q) = i \text{ \& } O(p+x, q+x) = j \\ 1 & \text{sicer} \end{cases} \quad (10)$$

Ker pri metodi Co-HOG generiramo so-pojavitveno matriko, nam ta vrne 3-dimenzionalni histogram, ki ga moramo transformirati v 1-dimenzionalni vektor značilk, slednji je normaliziran po L-2 norm formuli. Deskriptor značilk celotne preučevane slike je nato konstruiran z združevanjem vseh normaliziranih vektorjev značilk posameznega bloka.

3 Rezultati

Na podatkovnih zbirkah ICDAR 2003 [1] (5400 slik znakov testne zbirke, 6100 slik znakov učne zbirke), Chars74K [3] (13000 slik znakov naravnih scen), CVL OCR DB [2] (7000 slik znakov) smo ocenjevali algoritme HOG, PHOG ter Co-HOG. V podatkovnih zbirkah smo znake, ki predstavljajo ločila, šumnike in števila, izolirali od našega klasifikacijskega problema. Klasifikacijski problem smo s tem omejili na 52 razredov. Slike vsebovane v podatkovnih zbirkah smo predhodno normalizirali na velikost 32x32 slikovnih elementov.

Tabela 1. Tabela prikazuje klasifikacijske točnosti različnih podatkovnih zbirk ocenjenih z metodami HOG, PHOG in Co-HOG (SVM - Chi-Squared jedro). Zgornje vrstice prikazujejo klasifikacijske rezultate pri 26 razrednem problemu, spodnje vrstice obravnavajo 52 razredni problem.

	ICDAR	CVL OCR DB	Chars74K
HOG	81,80% 77,86%	91,51% 82,54%	68,82% 62,43%
PHOG	82,49% 78,58%	93,23% 84,57%	70,02% 63,91%
Co-HOG	80,33% 76,85%	90,11% 81,26%	67,29% 61,58%

Klasifikacijo smo opravili s SVM, KNN, ANN ter NB algoritmom. Pri klasifikatorju SVM smo uporabili Chi-Squared jedro in metodo »eden proti enemu«. Chi-Squared jedro predvideva, da ima nenegativne podatke, zato moramo negativne vrednosti vektorja značilk pretvoriti v pozitivne, prav tako pa moramo podatke normalizirati z L1-normo. S klasifikatorjem KNN (Manhattan dolžina) smo v povprečju dosegali 5% nižjo klasifikacijsko natančnost kot s SVM klasifikatorjem. S klasifikatorjem ANN smo v povprečju dosegali 4% nižjo klasifikacijsko natančnost kot s klasifikatorjem SVM in Chi-Squared jedrom. Klasifikacijska metoda Navni Bayes nam je prinesla najnižje klasifikacijske rezultate, ki so bili v povprečju za 20% nižji od ostalih.

Najboljši klasifikacijski rezultati [9] so bili doseženi na podatkovni zbirki slik teksta CVL OCR DB, kjer smo uporabili prečno preverjanje (K-fold, K=10). Na zbirki slik teksta CVL OCR DB smo dosegli najvišjo klasifikacijsko natančnost z algoritmom PHOG in klasifikatorjem SVM (Chi-Squared jedro) in sicer 93,23% pri invariantnosti velikosti znaka (26 razredni problem) ter 84,57% natančnost pri razlikovanju velikosti znakov (52 razredni problem). Na podatkovni zbirki ICDAR smo dosegli 82,49% natančnost pri

invariantnosti velikosti znaka ter 78,58% pri variantnosti, ti rezultati so bili doseženi s SVM klasifikatorjem ter metodo PHOG. Podatkovna zbirka Chars74K je bila prav tako evalvirana s klasifikatorjem SVM ter metodo PHOG, tu smo dosegli 70,02% natančnost pri 26 razrednem problemu in 63,91% natančnost pri 52 razrednem problemu. Dobre klasifikacijske rezultate lahko pripisujemo kvalitetnim slikam, ki jih vsebuje posamezna podatkovna zbirka ter prilagojenim metodam za pridobivanje značilk.

4 Zaključek

Izkazalo se je, da je najvišje klasifikacijske točnosti dosegal algoritem PHOG s klasifikacijsko metodo SVM in Chi-Squared jedrom (tabela 1). Visoka klasifikacijska točnost algoritma PHOG je botrovala velikosti vektorja značilk, saj je bil ta 10-krat večji kot pri metodi HOG. Algoritem HOG je zaradi svoje hitrosti bolj primeren za realnočasovno razpoznavo, saj generira manjše vektorje značilk in je obenem procesorsko in časovno manj zahteven kot algoritem PHOG:

Literatura

- [1] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, R. Young, K. Ashida, H. Nagai, M. Okamoto, H. Yamamoto, H. Miyao, Y. Zu, W. Ou, C. Wolf, J. Jolion, L. Todoran, M. Worrington, X. Lin. ICDAR 2003 Robust Reading Competitions: Entries, Results and Future Directions. *International Journal of Document Analysis and Recognition*, 7(2-3):str. 105-122, 2005.
- [2] A. Ikica, P. Peer. Cvl ocr db, an annotated image database of text in natural scenes, and its usability. *Informacije IDEM*, 41(2):str. 150-154, 2011.
- [3] The Chars74K dataset'2016, <http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/>
- [4] B. Triggs, N. Dalal. Histograms of oriented gradients for human detection. In *CVPR*, pages 886-893, 2005.
- [5] X. Munoz, A. Bosch, A. Zisserman. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval, CIVR '07*, pages 401-408, New York, NY, USA, 2007. ACM.
- [6] Z. R. Tan, S. Tian and C. L. Tan. Using pyramid of histogram of oriented gradients on natural scene text recognition. In *International Conference on Image Processing (ICIP)*, IEEE Conference on, pages 2629-2633, Oct 2014.
- [7] B. Su, C. L. Tan, S. Tian, S. Lu, Scene text recognition using co-occurrence of histogram of oriented gradients, In *2013 12th International Conference on Document Analysis and Recognition (ICDAR)*, pages 912-916, Aug 2013.
- [8] K. Yokoi, T. Watanabe, S. Ito. Co-occurrence histograms of oriented gradients for human detection. *IPSP Transactions on Computer Vision and Applications*, 2:39-47, 2010.
- [9] R. Petek. Optična razpoznavna znakov v slikah naravnih scen. Magistrsko delo. Ljubljana: Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, 2016.