

Otkrivanje regiona sa tekstom u slici

Vukota Peković, *RT-RK Computer Based Systems, Novi Sad*

Dragan Mrđan, *Fakultet tehničkih nauka, Novi Sad*

Vladimir Zlokolica, *RT-RK Computer Based Systems, Novi Sad*

Zoran Marčeta, *RT-RK Computer Based Systems, Novi Sad*

Sadržaj — U ovom radu je predstavljen algoritam za brzo i efikasno otkrivanje regiona sa tekstom u slici. Prvo su date osobine teksta na osnovu kojih je razvijen algoritam za otkrivanje. Posle toga je predstavljen sam algoritam. Ovaj algoritam predstavlja prvi i najvažniji korak u postupku optičkog prepoznavanja teksta. Zbog toga je izuzetno važno da je algoritam brz i da ima veliku tačnost, što je i pokazano prilikom ispitivanja. Na kraju su predstavljeni potencijalni dalji koraci u razvoju/poboljšanju algoritma i moguća primena.

Ključne reči — otkrivanje teksta, OCR

I. UVOD

KOLIČINA slika i video-snimaka u globalnoj mreži i u bazama podataka je u stalnom porastu i zbog toga je pravi izazov pronaći efikasne metode za održavanje i pretragu tih resursa po njihovom sadržaju. Tekstualni sadržaji u slici i/ili videu mogu nositi značajne informacije o tipu sadržaja slike/videa. Npr. tekst sa slika sa Interneta odražava sadržaj Internet stranica, naslovi knjiga i časopisa mogu biti korisni za njihovo učitavanje po vrsti digitalnog sadržaja.

Ugrađeni tekst u slike može neposredno da otkrije prirodu sadržaja slike bez komplikovanih proračuna i analiza (otkrivanje i prepoznavanje objekata na slici, proračuni inteziteta regiona, poređenje boja i tekstura...).

Uloga otkrivanja teksta je da se nađu regije slike koje sadrže samo tekst i da ih direktno označi korisniku i/ili prosledi modulu ili aplikaciji koja optički prepoznaje tekst (*Optical Character Recognize – OCR*) i dalje manipuliše slikom/videom. Ovo je najvažniji korak u čitavom postupku.

II. OTKRIVANJE I LOKALIZACIJA TEKSTA

Postoji više pristupa u rešavanju problema otkrivanja i lokalizacije teksta: (1) analiza povezanih komponenti, (2)

Ovaj rad je delimično finansiran od Ministarstva za nauku Republike Srbije, projekat 161003, od 2008. god.

Vukota Peković, Autor, RT-RK Computer Based Systems, Dr. Ilije Đuričića 2a, 21000 Novi Sad, Srbija; (e-mail: vukota.pekovic@rt-rk.com).

Dragan Mrđan, Koautor, Fakultet tehničkih nauka, Novi Sad, Srbija (e-mail: dragan.mrdjan@rt-rk.com).

Vladimir Zlokolica, Koautor, RT-RK Computer Based Systems, Novi Sad, Srbija (e-mail: vladimir.zlokolica@rt-rk.com).

Zoran Marčeta, Koautor, RT-RK Computer Based Systems, Novi Sad, Srbija (e-mail: zoran.marceta@rt-rk.com).

otkrivanje ivica, (3) otkrivanje uglova i (4) metode zasnovane na osobinama teksta.

Robustan algoritam mora da ispuni sledeće zahteve:

- Dobre performanse (mali procenat lažno otkrivenih regiona sa tekstom) za različite fontove, veličinu slova, boju, orijentaciju, jezike i složene pozadine.
- Stalne (konstantne) performanse sa minimalnom potrebom za ručnim podešavanjem parametara ili modifikacijom (*rebuild*) modela.

Da bi ispunili ove zahteve neki autori koriste u svojim radovima neuronske mreže [1] i vektorske mašine [2], [4], no njihovo korišćenje je računski izuzetno zahtevno i neefikasno.

Autori rada [3] su koristili analizu povezanih komponenti i prostornu varijansu koja se primenjuje na horizontalne linije slike. Mana ovog pristupa je nemogućnost otkrivanja teksta ukoliko su slova u jednoj liniji različite boje.

Autori [4] koriste algoritam zasnovan na pronalaženju horizontalnih i vertikalnih ivica slova. Za lokalizaciju koriste podršku vektorskih mašina.

Autori [5] koriste algoritam koji vrši analizu pripadnosti piksela (*pixel – picture element, najmanji deo digitalne slike*) ivici slova. Ta metoda je računski neefikasna i zahteva korišćenje neuronskih mreža prilikom odlučivanja da li grupa piksela predstavlja tekst ili ne.

Autori [6] koriste velvet transformaciju za računanje energije piksela. Zatim se za lokalizaciju teksta koriste razne osobine teksta (gustina promene inteziteta, uglovi...).

Radovi [3]–[6] se oslanjaju na algoritme koji analiziraju sliku piksel po piksel ili linija po linija, što zahteva dodatne proračune za lokalizaciju regiona koji sadrže tekst.

Pojavom HD (*High definition – visoka definicija*) multimedijalnog sadržaja brzina algoritma za otkrivanje naročito dolazi do izražaja.

Razvoj brzog i robustnog algoritma za otkrivanje regiona sa tekstom nije nimalo lak zadatak zbog otežavajućih osobina teksta:

- tekst može da bude na složenoj podlozi,
- veoma je teško naći osobine po kojima se razlikuju tekst i druge stvari koje liče na tekst (prozori, lišće...),
- tekst može da bude različitih boja, veličine, fonta, jezika, orijentacije,
- kvalitet teksta može da opadne usled šuma i procesa kodovanja/dekodovanja.

I pored nabrojanih mana, predstavljeni algoritam se u potpunosti oslanja na osobine teksta:

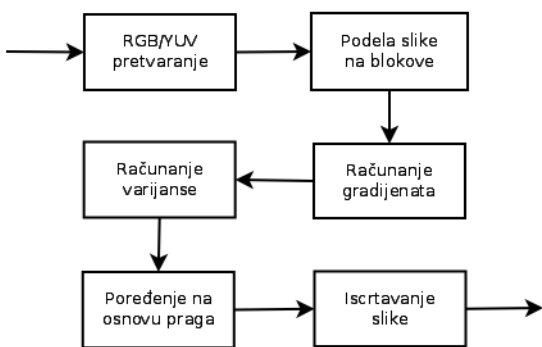
- veliku gustinu promene inteziteta,

- veliki kontrast između teksta i njegove podloge,
- teksture (osobine sastava teksta).

Ovde predstavljeni algoritam prvo pretvara sliku iz RGB u YUV model boja. Zatim se slika deli na blokove (to je najveća razlika u odnosu na [3]–[6]). Potom se računaju gradijenti blokova i njihove varijanse. Na osnovu vrednosti varijanse se ocenjuje da li je to blok sa tekстом ili bez teksta. Rezultat je nova slika sa obeleženim regionima sa tekстом.

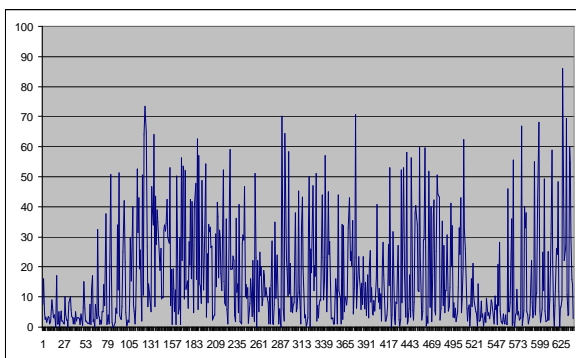
III. OPIS ALGORITMA

Algoritam se sastoji od više koraka koji su prikazani na Sl. 1.

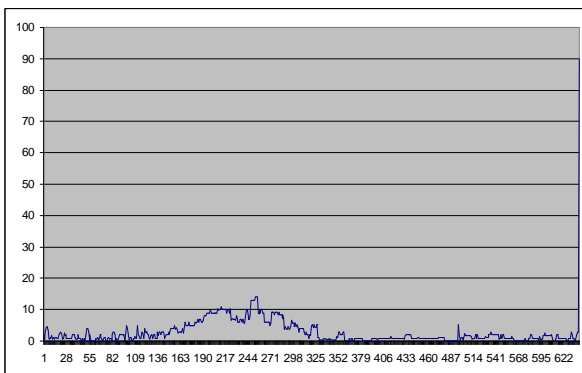


Sl. 1. Koraci algoritma

Algoritam je razvijen na osnovu razlike između gradijenata blokova sa tekстом i bez teksta [6]. Na Sl.2. i Sl. 3. su prikazani grafici gradijenata bloka sa tekстом i bloka bez teksta.



Sl. 2. Gradijenti bloka sa tekстом



Sl. 3. Gradijenti bloka bez teksta

Sa gornjih slika se vidi da su gradijenti kod bloka koji sadrži tekst (maksimalni gradijent ima vrednost 86) mnogo veći nego kod bloka bez teksta (maksimalni gradijent ima vrednost 14). Blok sa tekстом ima mnogo više velikih gradjenata nego blok koji ne sadrži tekst. Ti grafici oslikavaju osobine teksta nabrojane u II poglavlju.

A. Pretvaranje

Nakon učitavanja slike, ona se pretvara iz modela boja RGB u YUV model (računa se samo Y komponenta). Na taj način algoritam gubi zavisnost od višebojnih slova, takođe dobija i na efikasnosti jer je potrebno izračunati samo najznačajniju komponentu (Y – luma, osvetljenje).

B. Podela slike na blokove

Nakon pretvaranja algoritam deli sliku na male blokove veličine 40x16 piksela od kojih će se, ukoliko su ispunjeni svi uslovi, kasnije obrazovati regioni sa tekстом.

Veličina 40x16 je odabrana na osnovu rezultata koji su dobijeni za različite veličine blokova (40x16, 32x16, 20x8, 16x8). Ta veličina predstavlja dimenzije najmanje reči koju algoritam može da otkrije.

C. Računanje gradijenata

Zatim se za svaki blok računa horizontalni gradijent za osvetljenje (Y komponenta) koristeći masku [1 -1] za svaka dva susedna piksela, tj. računa se horizontalna promena inteziteta osvetljenja između dva susedna piksela (računa se apsolutna vrednost te razlike). Velika vrednost gradijenta označava da je došlo do prelaza između teksta i pozadine ili obrnuto.

Ukoliko je gradijent veći od utvrđenog praga (*threshold*) onda se koristi za računanje varijanse bloka, dok se gradijenti manji od praga odbacuju. Taj prag ima vrednost 45, utvrđen je ispitivanjem raznih vrednosti.

Ispitivane su vrednosti iz opsega 25-100. Različite vrednosti praga na jednoj slici mogu dati velike razlike u rezultatima otkrivanja (sadržaj i broj otkrivenih blokova). Za niže vrednosti dolazi do pogrešnog otkrivanja, kada se blokovi koji ne sadrže tekst označe da ga sadrže. To je posledica velikog kontrasta između teksta i njegove podloge. Tada malo složeniji objekat na slici može biti prepoznat kao tekst. Sa druge strane, ukoliko je prag viši od 45, dešava se da algoritam ne prepozna oblasti sa tekстом. Na slikama odnos boja tekst – pozadina nije uvek crno – belo (i obrnuto). Tada kontrast nije izražen, te zbog isuviše visokog praga blok sa takvim tekстом neće biti otkriven..

Ukoliko je broj takvih gradijenata manji od 35 taj blok se odmah označava kao region bez teksta. To je urađeno u cilju eliminacije šuma kao i nekih objekata koji bi mogli biti identifikovani kao tekst (npr. dve paralelne linije). Takođe to doprinosi i računskoj efikasnosti, jer se za te blokove dalje ne uvodi nikakav proračun. Vrednost 35 je određena na osnovu osobina jednog slova da mora imati najmanje 32 prelaza slovo – pozadina. Ukoliko slovo ima manje prelaza, to znači da je izolovano (jedino slovo u bloku – prvo ili poslednje slovo u reči, delovi slova koja imaju donji produžetak) i tada će taj blok biti označen kao blok bez teksta. Da se na taj način ne bi gubila slova, proširuje se čitava oblast koja sadrži tekst.

D. Računanje varijanse

Sledeći korak je računanje varijanse bloka na osnovu gradijenata koji zadovoljavaju uslov iz prethodnog bloka. Pre računanja varijanse se izračuna srednja vrednost gradijenata.

Varijansa se računa po formuli:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

gde je:

- σ^2 – varijansa
- N – broj elemenata za koje se računa varijansa
- x_i – trenutna vrednost gradijenta
- \bar{x} – srednja vrednost gradijenata.

E. Odlučivanje

U ovom koraku se odlučuje da li je to blok sa tekстом ili bez teksta. Ukoliko varijansa bloka prelazi određeni prag, taj blok je označen kao regija sa tekстом. Ovaj korak je zasnovan na osobinama teksta navedenim u II poglavlju.

Prag je postavljen na vrednost 50 i ta vrednost je utvrđena ispitivanjem kao vrednost koja daje najbolje rezultate. Više različitih vrednosti primenjenih na jednoj slici daju različite rezultate. Ukoliko je prag manji od 50,

algoritam može da zaključi da je objekat koji na slici ima veći kontrast ustvari tekst. Suprotno, ukoliko je vrednost praga iznad 50, moguće je da i pravi tekst u bloku neće biti prihvaćen kao tekst.

F. Iscrtavanje slike

Posle pretrage za regionima sa tekстом iscrtavaju se nove slike sa obeleženim regionima gde se nalazi tekst. Granice regiona sa tekстом se proširuju da bi se obuhvatili pikseli van blokova koji pripadaju tekstu (npr. slova koja imaju donji produžetak, prvo ili zadnje slovo u reči nije kompletno obuhvaćeno...). Regioni sa tekстом su iscrtani YUV modelom boja, dok se regioni bez teksta ne vide zato što su prekriveni crnom bojom.

IV. REZULTATI

Na sledećim slikama su pokazani primeri rada algoritma.

Vidi se mali stepen nepravilno otkrivenih oblasti (Sl. 6. i Sl. 9.) kao i mali procenat (2,4%) teksta koji nije otkriven (Sl. 5., Sl. 8., Sl. 6. i Sl. 9.). Nepravilno otkrivanje može da se desi ako objekti na slici imaju slične karakteristike kao tekst, dok je uzrok neotkrivenog teksta mali kontrast između teksta i njegove pozadine.



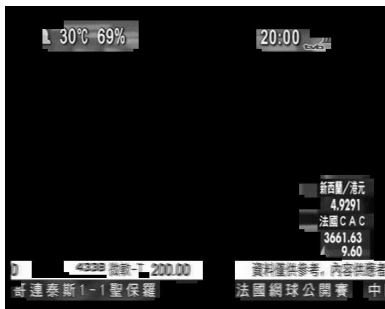
Sl. 4. Primer 1 – Ulazna slika



Sl. 5. – Primer 2 – Ulazna slika



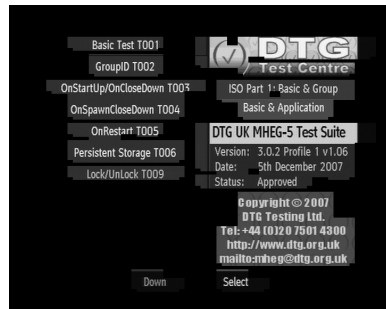
Sl. 6. Primer 3 – Ulazna slika



Sl. 7. Primer 1 – Otkriveni regioni



Sl. 8. Primer 2 – Otkriveni regioni



Sl. 9. Primer 3 - Otkriveni regioni

V. ZAKLJUČAK

Standardni sistemi za optičko prepoznavanje teksta daju loše rezultate za ulazne slike koje ne sadrže samo tekst. Onovni zadatak ovoga algoritma jeste priprema slike za postupak optičkog prepoznavanja teksta. Zbog značaja koji otkrivanje ima u tom postupku, provode se mnoga istraživanja sa ciljem da se dođe do maksimalne tačnosti otkrivanja. Opisani algoritam zbog svoje brzine i tačnosti se može prilagoditi i za otkrivanje teksta u video multimedijalnim sadržajima. Takođe se može i dalje poboljšati njegova tačnost dodatnim analizama blokova (npr. ako blokovi imaju mali kontrast, da li je broj prelaza pozadina-tekst i obrnuto paran...).

LITERATURA

- [1] H. Li, D. Doerman, "Text enhancements in digital video using multiple frame integration," in ACM Multimedia, 1999, pp. 385–395.
- [2] K. I. Kim, K. Jung, H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Transactions on PAMI 25, 2005, pp. 1631 – 1639.
- [3] Y.Zhong, K. Karu, A.K.Kain, "Locating Text in Complex Colour Images," in Pattern Recognition 10, 1995, pp.1523 – 1536 .
- [4] D. Chen, J – M. Odobez, H. Bourlard, "Text detection and recognition in images and video frames," in Pattern Recognition 37, 2004, pp. 595 – 608.
- [5] Q. Liu, C. Jung, S. Kim, Y. Moon, J. Kim, "Stroke filter for text localization in video images," in Proc. Int. Conf. Image Process., 2006, pp. 1473 – 1476.
- [6] Q. Ye, Q. Huang, W. Gao, D. Zhao, "Fast and robust text detection in images and video frames," in Image and Vision Computing 23, 2005, pp. 565 – 576.

ABSTRACT

This paper presents an algorithm for fast and efficient detection of the region with the text in the picture. First we show text features on which we developed our algorithm for text detection. After that we presented details of the algorithm. This algorithm is the first and most important step in the process of optical character recognition. It is therefore very important that the algorithm is fast and has great accuracy, as demonstrated in testing. Finally, we presented potential next steps in the development/improvement of the algorithm and possible implementation.

Detection of regions with the text in images

Vukota Pekovic, RT-RK Computer Based Systems
Dragan Mrdjan, Faculty of Technical Sciences
Vladimir Zlokolica, RT-RK Computer Based Systems
Zoran Marceta, RT-RK Computer Based Systems