Detekcija teksta v slikah naravnih scen

Andrej Ikica

doktorska disertacija

predana

Fakulteti za računalništvo in informatiko

kot del izpolnjevanja pogojev za pridobitev naziva

doktor znanosti

s področja

računalništva in informatike

Ljubljana, 2013

IKS
Računalniški inženiring

Naložba v vašo prihodnost
Operacijo delno financira Evropska unija
Evropski socialni sklad

Detekcija teksta v slikah naravnih scen

Andrej Ikica

doktorska disertacija
predana
Fakulteti za računalništvo in informatiko
kot del izpolnjevanja pogojev za pridobitev naziva
doktor znanosti
s področja
računalništva in informatike

Ljubljana, 2013

IZJAVA

Izjavljam, da sem avtor doktorske disertacije z naslovom Detekcija teksta v slikah naravnih scen, ki sem jo izdelal samostojno pod vodstvom mentorja, in da ta ne vsebuje materiala, ki bi ga kdorkoli predhodno že objavil ali oddal v obravnavo za pridobitev naziva na univerzi ali na drugem visokošolskem zavodu, razen v primerih, kjer so navedeni viri. Soglašam z javno objavo elektronske oblike doktorske disertacije, ki je identična s tiskano obliko doktorske disertacije.

— Andrej Ikica —

oktober 2013

Oddajo so odobrili

dr. Peter Peer

docent za računalništvo in informatiko

MENTOR IN ČLAN OCENJEVALNE KOMISIJE

dr. Franc Solina

profesor za računalništvo in informatiko

PREDSEDNIK OCENJEVALNE KOMISIJE

dr. Božidar Potočnik

izredni profesor za računalništvo in informatiko

ZUNANJI ČLAN OCENJEVALNE KOMISIJE

Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko

dr. Stanislav Kovačič

profesor za elektrotehniko

ZUNANJI ČLAN OCENJEVALNE KOMISIJE

Univerza v Ljubljani, Fakulteta za elektrotehniko

PREDHODNA OBJAVA

Izjavljam, da so bili rezultati obravnavane raziskave predhodno objavljeni/sprejeti za objavo v recenzirani reviji ali javno predstavljeni v naslednjih primerih:

[1] A. Ikica and P. Peer. SWT voting-based color reduction for text detection in natural scene images. *EURASIP Journal on Advances in Signal Processing*, 2013(1):1–13, 2013. doi: 10.1186/1687-6180-2013-95

[2] A. Ikica and P. Peer. CVL OCR DB, an annotated image database of text in natural scenes, and its usability. *Info. MIDEM*, 41(2):150–154, 2011.

[3] A. Ikica and P. Peer. An improved edge profile based method for text detection in images of natural scenes. In *Proc. of IEEE Conference on Computer as a Tool (EUROCON)*, pages 1–4, 2011. doi: 10.1109/EUROCON.2011.5929289

Potrjujem, da sem pridobil pisna dovoljenja vseh lastnikov avtorskih pravic, ki mi dovoljujejo vključitev zgoraj navedenega materiala v pričujočo disertacijo. Potrjujem, da zgoraj navedeni material opisuje rezultate raziskav, izvedenih v času mojega podiplomskega študija na Univerzi v Ljubljani.

Univerza *v Ljubljani*
Fakulteta *za računalništvo in informatiko*
Andrej Ikica
Detekcija teksta v slikah naravnih scen

POVZETEK

Detekcija teksta v slikah naravnih scen je razmeroma novo področje računalniškega vida. Zaradi njene aplikativnosti – od avtomatičnega preiskovanja vizualnih vsebin ter pomoči slepim in slabovidnim do prevajanja z mobilnimi telefoni posnetih napisov – se ji v zadnjem času namenja precej pozornosti. Namen detekcije teksta je poiskati regije v sliki, ki vsebujejo tekst, zato jo je od same razpoznave teksta treba razlikovati. Kljub temu sta detekcija in razpoznava teksta povezani, saj se detektirane regije lahko uporabljajo kot vhod modulom za razpoznavo teksta.

Zaradi raznovrstnosti in kompleksnosti slik naravnih scen je detekcija teksta izredno težavna. Tekst je lahko v slikah na poljubnih mestih in v poljubnih oblikah, velikostih ter barvah. Poleg tega je podvržen številnim geometrijskim transformacijam, ki so posledica narave zajema slik. Ne nazadnje, slike naravnih scen tipično vsebujejo izredno kompleksna ozadja, ki dodatno otežujejo postopek detekcije teksta.

V doktorski disertaciji poleg pregleda področja predstavljamo metodi barvne redukcije na podlagi glasovanja SWT in detekcije smeri SWT, ki predstavljata izvirni znanstveni prispevek. Prva izmed naštetih je segmentacijska metoda, prilagojena detekciji teksta. Z uporabo strukturne informacije (SWT) metoda usmerja postopek barvne redukcije in izboljšuje natančnost segmentacije v primerjavi z ostalimi primerljivimi metodami, ki trenutno dosegajo ene izmed najboljših rezultatov. Barve, ki so bogate s slikovnimi elementi SWT, po vsej verjetnosti pripadajo tekstu, zato jih metoda v fazi barvne redukcije blokira in jim ne dovoli, da bi konvergirale proti barvam ozadja.

Pomanjkljivost metode SWT je detekcija pravilne smeri iskanja. Metoda namreč paralelne robove črk v sliki išče v smeri gradientov robnih točk. V primeru temnih tekstov na svetli podlagi gradienti pravilno kažejo v notranjost črk, v primeru svetlih tekstov na temni podlagi pa kažejo v nasprotno smer, kar povzroča nepravilno detekcijo teksta. Avtorji metode omenjeni problem sicer rešujejo z dvakratnim poganjanjem

celotnega postopka v gradientni in protigradientni smeri ter z integracijo rezultatov obeh smeri, vendar je takšen način nenatančen in časovno kompleksen – zahteva namreč dvakratno poganjanje celotne metode. Da bi se problemu izognili, v doktorski disertaciji predlagamo metodo detekcije smeri SWT, ki na podlagi analize histogramov posameznih blokov slike SWT določi pravilno smer iskanja v enem koraku.

Najpogosteje se za evalvacijo metod detekcije teksta uporabljata zbirki ICDAR 2003 in ICDAR 2011. Glavna pomanjkljivost obeh je anotacija posameznih besed teksta s pravokotniki. Takšen tip anotacije zahteva, da metoda sama poskrbi za pravilno grupiranje detektiranega teksta v besede, kar je s stališča objektivne evalvacije problematično. Z namenom objektivnejše evalvacije smo postavili lastno javno dostopno zbirko slik teksta v naravnih scenah CVL OCR DB, ki uporablja anotacijo z n-poligoni in binarno anotacijo. Binarna anotacija omogoča evalvacijo na nivoju posameznih črk in s tem odpravlja potrebo po dodatnem grupiranju teksta v besede.

Eksperimentalni rezultati kažejo, da metoda barvne redukcije na podlagi glasovanja SWT na zbirki CVL OCR DB dosega boljše segmentacijske rezultate kot metoda barvne redukcije, prilagojene detekciji teksta. To v fazi segmentacije uporablja metoda strukturne particije in grupiranja, ki trenutno dosega ene izmed najboljših rezultatov med obstoječimi metodami detekcije teksta. S problemom ugotavljanja pravilne smeri SWT se literatura eksplicitno ne ukvarja, zato metode detekcije smeri SWT ni mogoče primerjati z ostalimi metodami. Kljub temu metoda na zbirki CVL OCR DB dosega visoko stopnjo detekcije in je sposobna pravilno določiti smer SWT tudi, ko so v sliki prisotni tako temni teksti na svetlih podlagah kot svetli teksti na temnih podlagah.

V splošnem doktorska disertacija služi tudi kot dobro izhodišče za pregled področja detekcije teksta v slikah naravnih scen.

Ključne besede: računalniški vid, detekcija teksta, slike naravnih scen, evalvacijske zbirke, SWT, barvna redukcija, glasovanje SWT, detekcija smeri SWT, CVL OCR DB

University *of Ljubljana*
Faculty *of Computer and Information Science*
Andrej Ikica
Text detection methods in images of natural scenes

ABSTRACT

Text detection in natural scene images has gained much attention in the last years due to its enormous applicative potential in many areas such as content-based image retrieval, PDA signboard translators and applications for assisting blind and visually impaired people. A clear distinction, however, has to be made between text detection and text recognition. The task of the former is to locate text regions in an image, not to recognize them. Nevertheless, text detection and text recognition are closely related since the detected text regions can be subsequently fed into the text recognition modules.

Due to diversity and complexity of natural scene images, the text detection task is considerably challenging. Text can appear at arbitrary image locations in arbitrary shapes, sizes and colors. Additionally, it is often subject to numerous geometric transformations. Finally, natural scene images contain very complex backgrounds, which make text detection even more difficult.

In this dissertation, we present two novel methods: *SWT voting-based color reduction* and *SWT direction determination*. The first is a text detection-oriented segmentation method, that supervises the color reduction process by integrating additional SWT information. It improves segmentation accuracy compared to the other state-of-the-art methods. Colors rich with SWT pixels most likely belong to text and are therefore blocked from being mean-shifted away towards background colors.

One of the disadvantages of the SWT method is the search direction problem. The method searches for parallel character edges in the gradient directions. In case of a dark text on a light background gradients correctly point towards character interiors, whereas in case of a light text on a dark background they point in the opposite directions and cause incorrect text detection. In order to solve the problem, the authors of the SWT method run the algorithm twice – in gradient and counter-gradient di-

rections. Such approach, however, is imprecise and time consuming since the whole method has to be run twice. To avoid the search direction issue, we present a novel SWT direction determination method. By analyzing SWT sub-block histograms of both gradient and counter-gradient directions, the method is able to determine the correct SWT direction in one step.

Usually, ICDAR 2003 and ICDAR 2011 datasets are used for text detection evaluation. Their disadvantage is rectangular annotation of single words in images, which requires, that the detected text is already grouped into words. Since text detection and word grouping are separate subjects, such annotation is problematic from the perspective of objective evaluation. Therefore, we created our own public annotated dataset of text in natural scene images CVL OCR DB. The dataset supports two types of annotation: *n-polygon annotation* and *binary annotation*. The latter allows *per character evaluation* and makes word grouping unnecessary.

Experimental results on the CVL OCR DB dataset indicate that the SWT voting-based color reduction method outperforms the text-oriented color reduction method, which is used in the segmentation phase of the state-of-the-art text detection method of structure-based partition and grouping. Literature does not explicitly address SWT search direction issue; thus, the SWT direction determination method cannot be directly compared to the other methods. Nevertheless, the method achieves high detection rate on the CVL OCR DB dataset and is able to determine correct SWT directions when both dark text on light backgrounds and light text on dark backgrounds appear in the image.

Generally speaking, the dissertation can also serve as a survey of text detection in natural scene images.

Key words: computer vision, text detection, natural scene images, evaluation datasets, SWT, color reduction, SWT voting, SWT direction determination, CVL OCR DB

ZAHVALA

Iskreno se zahvaljujem vsem, ki so mi pomagali pri nastajanju doktorske disertacije. Na prvem mestu gre zahvala mentorjema, doc. dr. Petru Peeru in mag. Branku Ikici (mentor iz gospodarstva), ki sta mi v času študija stala ob strani in mi pomagala s številnimi idejami in nasveti. Zahvala gre tudi članom Laboratorija za računalniški vid za konstruktivne diskusije in nasvete. Prav tako se zahvaljujem vsem članom komisije, prof. dr. Francu Solini, prof. dr. Božidarju Potočniku in prof. dr. Stanislavu Kovačiču, za koristne pripombe in nasvete ter trud, ki so ga vložili v pregled doktorske disertacije.

Na koncu gre zahvala tudi moji družini, ki mi je ves čas študija stala ob strani in me podpirala.

— Andrej Ikica, Ljubljana, oktober 2013.

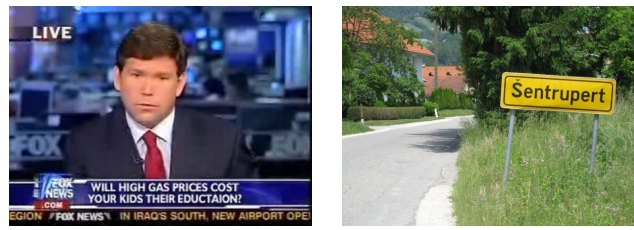
KAZALO

I

Uvod

1.1 Detekcija teksta v slikah naravnih scen

Tekst v slikah predstavlja pomembno vizualno in semantično informacijo, ki jo je iz ostalih visokonivojskih značilk slike težko izluščiti. Za ilustracijo navedimo primer superponiranega teksta v televizijskem posnetku, ki podaja trenutni izid športne tekme, ime osebe, ki je v sceni prisotna, naslov videospota, ki se predvaja, ali celo kratek povzetek prispevka – na primer *"Will high gas prices cost your kids their education?"* na sliki 1.1a. Podobno fotografija table z imenom kraja jasno podaja, kje je bila slika posneta (slika 1.1b). Brez pripadajočega teksta bi praktično nemogoče prišli do informacij, ki jih ponuja tekst.



(a) (b)

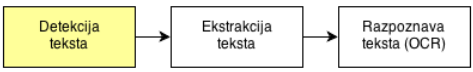
Slika 1.1
Pomen teksta v slikah.

Avtomatična razpoznava teksta v slikah je domena računalniškega vida, ki se z njo ukvarja že več desetletij. Sprva se je razvoj osredotočal na preprosto razpoznavo tiskanih črk, kasneje pa so se začele pojavljati kompleksnejše rešitve za razpoznavo pisanih črk in razpoznavo teksta v slikah kompleksnejših dokumentov. Z digitalizacijo videovsebin ob koncu 90. let so se začele pojavljati rešitve za razpoznavo teksta v videoposnetkih, zadnja leta pa je razvoj vse bolj usmerjen tudi v razpoznavo teksta v t. i. slikah naravnih scen. Z izrazom *slike naravnih scen* označujemo široko paleto raznovrstnih slik vsakdanjih scen, ki so tipično zajete s splošno razširjeno opremo za zajem slik. Za razliko od standardne razpoznave črnih črk na beli podlagi predstavlja razpoznava teksta v slikah naravnih scen trd oreh, saj gre večinoma za zelo različne slike, zajete pod različnimi vremenskimi in svetlobnimi pogoji ter z raznovrstno opremo – od mobilnih telefonov, fotoaparatov do videokamer. Tekst v slikah naravnih scen je na poljubnih mestih, podvržen je številnim transformacijam, vsebuje ogromno različic pisav in barv ter, ne nazadnje, slike naravnih scen vsebujejo kompleksna ozadja, ki so zlahka na-

(a) (b) (c) (d) (e) (f)

Slika 1.2

Primeri teksta v slikah naravnih scen. (a) Primer napisa na trgovini. (b) Primer teksta z različnimi barvami črk. (c) Primer različnih pisav. (d) Primer poševnega teksta. (e, f) Primera ukrivljenega teksta.



Slika 1.3

Tipični koraki avtomatične razpoznave teksta. Z rumeno barvo je označena faza detekcije teksta, s katero se ukvarjamo v doktorski nalogi.

pačno klasificirana kot tekst. Nekaj primerov teksta v slikah naravnih scen prikazuje slika 1.2.

Tipično razpoznava teksta v slikah naravnih scen poteka v treh korakih (slika 1.3). Tekst se v sliki najprej poišče (detekcija teksta), nato izlušči (ekstrakcija teksta) in na koncu klasificira z uporabo klasifikatorjev OCR (angl. *Optical Character Recognition*), ki so prilagojeni naravnim scenam. Vsi trije koraki so za natančno razpoznavo teksta izredno pomembni, hkrati pa tudi izredno zahtevni, zato se raziskovalci pogosto ukvarjajo le z enim od teh korakov. V doktorski nalogi se ukvarjamo s prvim korakom, tj. detekcijo teksta v slikah naravnih scen.

Kljub specifičnosti je detekcija teksta v slikah naravnih scen močno prepletena z detekcijo teksta v dokumentih [1–6] in z detekcijo teksta v videoposnetkih. Pri slednji ločimo detekcijo superponiranega teksta (angl. *captions*) [7], detekcijo teksta, ki je del scene (angl. *scene text*), in detekcijo obeh vrst teksta hkrati [8–13]. Detekcija teksta v videoposnetkih sega že v 90. leta prejšnjega stoletja, vendar še vedno ne dosega

želenih rezultatov. Tako se tudi v zadnjih letih pojavljajo izpopolnjene metode, ki tekst v videoposnetkih detektirajo natančneje [14–21]. Množična dostopnost naprav za zajem slik (npr. mobilni telefoni z integrirano kamero) in izredna aplikativnost področja sta razlog, da je detekciji teksta v slikah naravnih scen zadnja leta namenjeno veliko pozornosti. Med številnimi potencialnimi področji, kjer jo lahko uporabimo, jih naštejmo nekaj: preiskovanje in indeksiranje vizualnih vsebin, pomoč slabovidnim in slepim [22, 23], prevajanje napisov, posnetih z mobilnimi napravami [24–26], ter hitra in učinkovita digitalizacija besedil – namesto z dragimi in nerodnimi skenerji lahko besedilo zajamemo in digitaliziramo z mobilnim telefonom.

Podrobnejši pregled področja detekcije teksta v slikah in videoposnetkih je podan v [27–29]. Tipično se metode detekcije teksta delijo na teksturne (angl. *texture-based*) in regijske (angl. *region-based*). Ideja teksturnih metod [17, 30–32] je premikanje preiskovalnega okna vzdolž celotne slike in iskanje območij, ki ustrezajo definirani teksturi teksta. Pri tem se lahko opazuje gostota robov v okolici centra preiskovalnega okna, obnašanje gradientov v okolici teksta, varianca intenzitet slikovnih elementov, porazdelitev valčnih ali DCT-koeficientov (prednost DCT-koeficientov je v tem, da so neposredno dostopni, če je slika v formatu JPEG) in podobno. Poleg visoke časovne kompleksnosti je slabost teksturnih metod občutljivost na različne velikosti teksta, omejenost na horizontalne tekste ter nenatančno določanje dejanskih robov teksta. Regijske metode [33–37] temeljijo na izgradnji povezanih komponent (angl. *connected components*) s podobno barvno porazdelitvijo ali podobno intenziteto robov. Regijske metode so hitre in niso omejene le na horizontalne tekste, vendar pa so občutljive na kompleksna ozadja in šum (splošno znano je namreč, da so lokalne značilke občutljive na šum). Poleg teksturnih in regijskih obstajajo tudi hibridne metode [38–43], ki izkoriščajo prednosti obeh pristopov – hitrost in orientacijsko neodvisnost regijskega ter robustnost teksturnega pristopa.

Metode detekcije teksta se dodatno delijo na omejitveno usmerjene (angl. *constraint-based*) in učno usmerjene (angl. *learning-based*). Ideja omejitveno usmerjenih je predefinirana množica pravil oz. hevristik, ki se uporabljajo pri filtriranju netekstovnih regij. Tipično gre za omejitve velikosti kandidatnih regij, razmerja med višino in širino kandidatne regije, ekvidistančnosti regij ipd. Pri učno usmerjenih metodah se namesto množice hevristik uporablja klasifikator, ki kandidatne regije klasificira kot tekst oz. kot ozadje. Marsikatera rešitev vključuje oba pristopa, vendar lahko v grobem pristope v [8, 9, 22, 23] uvrstimo v prvo, pristope v [11, 12, 17, 39] pa v drugo skupino.

Najbolj razširjena zbirka za evalvacijo metod detekcije teksta je zbirka ICDAR 2003 [44], ki je bila postavljena za potrebe tekmovanja v detekciji in razpoznavi teksta v slikah naravnih scen v okviru konference ICDAR leta 2003. Največji problem zbirke ICDAR je anotacija teksta s pravokotniki, pri čemer vsak pravokotnik ustreza posamezni besedi teksta. Poleg nenatančnosti anotacije in omejenosti na anotacijo le horizontalnih tekstov takšen način anotacije predvideva, da metoda detektiran tekst že pravilno grupira v posamezne besede, kar je izredno težavno. Tudi v primeru natančne detekcije lahko metoda zaradi slabega grupiranja teksta v besede dosega zelo slabe rezultate. Poleg zbirke ICDAR, ki je bila leta 2011 sicer posodobljena [45], a problema anotacije s pravokotniki še vedno ni odpravila, ne obstaja veliko javno dostopnih zbirk. Ena izmed njih je zbirka SVT [46], generirana iz slik aplikacije *Google Street View*[1], ki pa ni splošno razširjena.

1.2 Namen disertacije in pregled vsebine

Glavni namen doktorske disertacije je proučiti trenutno stanje na področju detekcije teksta v slikah naravnih scen, identificirati prednosti in slabosti določenih metod ter predlagati izboljšave. Kljub temu, da se področje šele razvija, literatura navaja veliko število raznovrstnih metod, zato se v doktorski nalogi omejimo na tiste, ki trenutno dosegajo najboljše rezultate.

V poglavju 2 podajamo podroben opis obstoječih zbirk za evalvacijo metod detekcije teksta in identificiramo njihove pomanjkljivosti. Skupna slabost obstoječih zbirk je anotacija s pravokotniki, ki jo skušamo odpraviti z lastno javno dostopno zbirko CVL OCR DB [47]. Zbirka uporablja anotacijo z n-poligoni in binarno anotacijo. Slednja omogoča evalvacijo detektiranega teksta na nivoju posameznih črk in ne besed. Poleg slik teksta v naravnih scenah zbirka vsebuje tudi posamezne sličice črk, kar ji daje širšo uporabno vrednost, saj jo je prav tako možno uporabljati za učenje in testiranje klasifikatorjev OCR. Velik problem zbirke ICDAR 2003 [44], na katerega opozarjajo tudi avtorji zbirke ICDAR 2011 [45], je izbira mer natančnosti. Preciznost (angl. *precision*) in priklic (angl. *recall*), ki se uporabljata, izhajata iz teorije signalov in nista najbolj primerna za podajanje natančnosti detekcije teksta. Pri detekciji teksta namreč težko govorimo o absolutnih zadetkih (angl. *hits*) in zgrešitvah (angl. *misses*), saj lahko delno zamaknjen pravokotnik še vedno pravilno pokriva določen tekst. Različna metrika pri

[1] http://www.google.com/streetview

evalvaciji preciznosti in priklica lahko povzroči odstopanje rezultata od 0,81 do 1,00 [39], zaradi česar lahko natančnost določene metode pade za nekaj razredov.[2] Mere natančnosti, ki jih uporabljamo v zbirki CVL OCR DB, ta problem rešujejo, saj tekst obravnavajo kot množico črk.

Metode detekcije teksta so evalvirane zelo različno. Celo v primeru evalvacije metod na zbirki ICDAR 2003 se dogaja, da se za evalvacijo uporabljajo različni deli zbirke ali celo da se namesto predpisanih povprečnih mer računajo mere preko vseh anotiranih in detektiranih pravokotnikov v vseh slikah zbirke, kar nekoliko dvigne natančnost detekcije. Zaradi neenotnega načina evalvacije je metode težko objektivno primerjati. Kljub temu v poglavju 3 podajamo rangiranje trenutno najuspešnejših metod in jih tudi podrobneje opišemo.

Zaradi časovne kompleksnosti teksturnega pristopa se pri zasnovi izboljšav opiramo na hitrejši regijski pristop, saj je primernejši za integracijo metod na mobilno platformo, ki predstavlja prihodnost aplikacij za detekcijo in razpoznavo teksta. Večina regijskih metod se v fazi segmentacije pri analizi slikovnih elementov opira bodisi na strukturo bodisi na barvo teksta, zato raziskave usmerjamo v razvoj hibridne rešitve, ki združuje tako strukturno kot barvno informacijo. Velik potencial za nadgradnjo imata metoda SWT [33] in metoda barvne redukcije, prilagojene detekciji teksta [34, 48], saj obe dosegata zelo dobre rezultate. Prav tako sta strukturna orientiranost metode SWT in barvna orientiranost metode barvne redukcije, prilagojene detekciji teksta, pisani na kožo našemu konceptu hibridizacije.

Poglavje 4 namenjamo opisu predlaganih metod barvne redukcije na podlagi glasovanja SWT [49] in detekcije smeri SWT [49]. Metoda barvne redukcije na podlagi glasovanja SWT je segmentacijska metoda, ki je prilagojena detekciji teksta. V povezavi z moduloma za filtriranje in grupiranje povezanih komponent je primerna za uporabo pri regijskem pristopu detekcije teksta. Metoda postopek barvne redukcije usmerja s strukturno informacijo SWT [33] in s tem izboljšuje delovanje osnovne segmentacijske metode barvne redukcije, prilagojene detekciji teksta [48]. Problem metode SWT je nezmožnost določanja prave smeri iskanja paralelnih robov črk. V primeru temnih tekstov na svetli podlagi je smer iskanja pravilna, saj gradienti robnih točk kažejo v notranjost črk, medtem ko v primeru svetlih tekstov na temni podlagi gradienti kažejo

[2] V poglavju 2, kjer podajamo rezultate tekmovanj ICDAR, in v poglavju 3, kjer rangiramo trenutno najuspešnejše metode, je jasno razvidno, kako različne vrednosti preciznosti in priklica vplivajo na rangiranje določene metode.

v napačno smer in povzročajo nepravilno delovanje metode. Metoda detekcije smeri SWT, ki jo predlagamo, razbije sliko SWT na bloke in z analizo histogramov določi pravilne smeri iskanja v posameznih blokih.

V disertaciji se skušamo evalvacije metod detekcije teksta lotiti še z drugega zornega kota. Metode so tipično evalvirane na določeni evalvacijski zbirki, pri čemer so rezultati podani v obliki takšne ali drugačne natančnosti. Poleg klasične evalvacije se zato v poglavju 5 ukvarjamo tudi z vplivom ostrine in šuma na obnašanje metod. Takšna vrsta testiranja je smiselna, saj raznovrstnost pogojev zajema slik naravnih scen pogosto vpliva na različne stopnje šuma v sliki.

Zaključki disertacije so strnjeni v poglavju 6.

2

Evalvacijske zbirke

2.1 Uvod

Detekcija teksta v slikah naravnih scen je relativno novo področje, zato javno dostopnih zbirk za evalvacijo metod detekcije teksta ni veliko. Najbolj razširjena med njimi je zbirka ICDAR 2003 [44], ki je bila do nedavnega praktično edina primerna za resnejšo evalvacijo. Zadnjih nekaj let sta na voljo tudi javno dostopni zbirki ICDAR 2011 [45] in SVT [46]. Zbirka ICDAR 2011 je bila postavljena z namenom korenite prenove zbirke ICDAR 2003, medtem ko je zbirka SVT zgenerirana iz slik aplikacije *Google Street View*. Skupna pomanjkljivost vseh treh zbirk je anotacija posameznih besed teksta s pravokotniki. Takšna anotacija zahteva, da metoda detektiran tekst pravilno grupira v posamezne besede, kar je s stališča objektivnosti evalvacije detekcije teksta problematično. Kljub temu, da določena metoda zelo dobro detektira tekst, lahko zaradi slabega modula za grupiranje teksta dosega slabe rezultate. Da bi se problemu izognili, smo postavili lastno zbirko slik teksta CVL OCR DB [47], ki uporablja anotacijo z *n*-poligoni in binarno anotacijo.

2.2 ICDAR 2003

Zbirka ICDAR 2003 [44] je bila leta 2003 postavljena za potrebe tekmovanja *Robust Reading Competition* v okviru konference ICDAR (*International Conference on Document Analysis and Recognition*). Namen tekmovanja je bil oceniti takratno stanje področja detekcije teksta v slikah naravnih scen. Kljub temu, da je bilo tekmovanje razdeljeno v tri kategorije, to so *detekcija teksta*, *razpoznava znakov* in *razpoznava besed*, so se vsi tekmovalci prijavili le v kategorijo detekcije teksta.

Organizatorji tekmovanja so zbirko razdelili v štiri sklope: *Sample*, *TrialTrain*, *TrialTest* in *Competition*. Sklop *Sample* (20 slik) je bil na voljo pred začetkom tekmovanja, da so si potencialni tekmovalci lahko ustvarili vtis, za kakšno vrsto slik gre. Sklopa *TrialTrain* (258 slik) in *TrialTest* (251 slik) sta bila namenjena učenju, testiranju in finalnim nastavitvam metod, ki so jih morali tekmovalci poslati do predvidenega roka. Po preteku roka so vse prejete metode organizatorji evalvirali na sklopu *Competition* (544 slik). Rezultati tekmovanja so predstavljeni v poglavju 2.2.2.

Zbirka ICDAR 2003 ni bila omejena le na tekmovanje, temveč so organizatorji z njo želeli spodbuditi nadaljnjo evalvacijo metod. Pri tem je bila postavljena splošna smernica, da se za učenje metod uporablja sklop *TrialTrain*, medtem ko je sklop *TrialTest* namenjen evalvaciji in objavljanju rezultatov [44]. Avtorji se teh smernic držijo

zelo različno. Tako nekateri za evalvacijo uporabljajo le del slik določenega sklopa, oba sklopa skupaj ali celo oba sklopa skupaj s sklopom *Competition*.

Sklopi *Sample*, *TrialTrain*, *TrialTest* in *Competition* so organizirani enako. Vsak sklop vsebuje podmapo s slikami in anotacijsko datoteko *locations.xml*. Izsek anotacijske datoteke sklopa *Competition* je prikazan na sliki 2.1. V anotacijski datoteki vsaki sliki v sklopu ustreza pripadajoča značka (angl. *tag*) <*image*> s podznačkami <*imageName*>, <*resolution*> in <*taggedRectangles*>, ki označujejo ime slike, njeno ločljivost ter anotacijske pravokotnike v sliki. Posamezni anotacijski pravokotnik <*taggedRectangle*> ima 6 atributov: *x*, *y*, *width*, *height*, *offset*, *rotation*, *userName*, ki ustrezajo koordinatama *x* in *y* zgornjega levega kota anotacijskega pravokotnika, njegovi dolžini in širini, nagibu in rotaciji[1] pravokotnika ter imenu uporabnika, ki je anotiral sliko. Primeri anotiranih slik zbirke so prikazani na sliki 2.2.

```
  <image>
    <imageName>sml_01.08.2002/IMG_1202.JPG</imageName>
    <resolution x="1600" y="1200" />
    <taggedRectangles>
      <taggedRectangle x="219.0" y="763.0" width="715.0" height="69.0"
offset="0.0" rotation="0.0" userName="admin" />
      <taggedRectangle x="952.0" y="762.0" width="115.0" height="70.0"
offset="0.0" rotation="0.0" userName="admin" />
      <taggedRectangle x="1080.0" y="765.0" width="365.0" height="68.0"
offset="0.0" rotation="0.0" userName="admin" />
      <taggedRectangle x="584.0" y="856.0" width="487.0" height="66.0"
offset="0.0" rotation="0.0" userName="admin" />
      <taggedRectangle x="75.0" y="957.0" width="503.0" height="60.0"
offset="0.0" rotation="0.0" userName="admin" />
      <taggedRectangle x="593.0" y="964.0" width="405.0" height="57.0"
offset="0.0" rotation="0.0" userName="admin" />
      <taggedRectangle x="1015.0" y="963.0" width="134.0" height="52.0"
offset="0.0" rotation="0.0" userName="admin" />
      <taggedRectangle x="1160.0" y="968.0" width="392.0" height="44.0"
offset="0.0" rotation="0.0" userName="admin" />
    </taggedRectangles>
  </image>
```

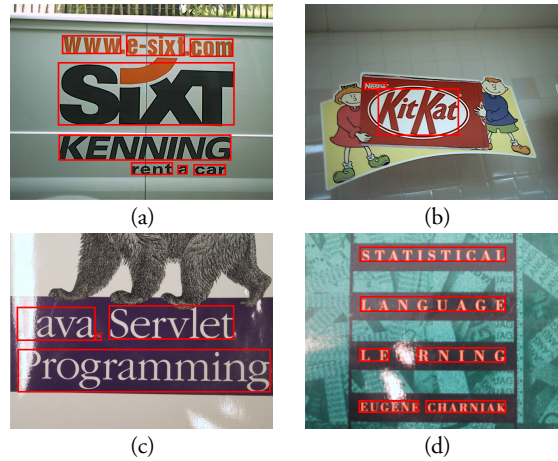
Slika 2.1

Izsek anotacijske datoteke *locations.xml* sklopa *Competition* zbirke ICDAR 2003.

2.2.1 *Evalvacijska shema*

Za ocenjevanje uspešnosti detekcije teksta zbirka ICDAR 2003 uporablja priklic r (angl. *recall*) in preciznost p (angl. *precision*). Priklic in preciznost sta standardni meri za

[1] Nagib se uporablja za anotacijo nagnjenih (angl. *slanted*) tekstov in določa horizontalni odmik med zgornjim levim in spodnjim levim ogliščem anotacijskega pravokotnika, medtem ko rotacija določa nagib celotnega anotacijskega pravokotnika. Nagib in rotacija se v kategoriji detekcije teksta ne upoštevata.



Slika 2.2

Primeri anotiranih slik zbirke ICDAR 2003. Nekonsistentnost anotacije na slikah (c) in (d): na sliki (c) je osvetljena črka "J" anotirana, medtem ko osvetljena črka "A" na sliki (d) ni anotirana.

ocenjevanje uspešnosti klasifikacije in sta v splošnem definirani kot:

$$r = \frac{TP}{FN + TP}, \tag{2.1}$$

$$p = \frac{TP}{FP + TP}, \tag{2.2}$$

pri čemer TP (angl. *True Positives*) označuje število pravilno napovedanih pozitivnih primerov, FN (angl. *False Negatives*) število nepravilno napovedanih negativnih primerov in FP (angl. *False Positives*) število nepravilno napovedanih pozitivnih primerov. Zaradi specifičnosti detekcije teksta sta priklic in preciznost v zbirki prilagojena in definirana na naslednji način:

$$r = \frac{c}{|T|}, \tag{2.3}$$

$$p = \frac{c}{|E|}, \tag{2.4}$$

pri čemer T označuje množico anotiranih pravokotnikov[2] (rdeči pravokotniki na sliki 2.3), E množico detektiranih pravokotnikov[3] (zeleni pravokotniki na sliki 2.3) in c

[2]Črka T izhaja iz angleške besede *targets*.

[3]Črka E izhaja iz angleške besede *estimates*.

število pravilno detektiranih pravokotnikov v sliki. Ker je lahko določen pravokotnik pravilno detektiran kljub temu, da se z anotiranim pravokotnikom ne ujema popolnoma, zbirka ICDAR 2003 koncept ujemanja razširja. Ujemanje m med dvema pravokotnikoma je definirano kot razmerje med ploščino njunega preseka (moder pravokotnik na sliki 2.3) in ploščino minimalnega pravokotnika, ki ju obdaja (črtkan pravokotnik na sliki 2.3). Na podlagi definicije ujemanja sta priklic in preciznost definirana kot:

$$r = \frac{\sum_{rect_t \in T} m(rect_t, T)}{|T|}, \tag{2.5}$$

$$p = \frac{\sum_{rect_e \in E} m(rect_e, T)}{|E|}, \tag{2.6}$$

pri čemer $m(rect, T)$ označuje najboljše ujemanje pravokotnika $rect$ z vsemi pravokotniki iz množice T. Poleg priklica in preciznosti zbirka ICDAR 2003 uporablja dodatno mero f, ki je obteženo harmonično povprečje priklica in preciznosti:

$$f = \left(\frac{\alpha}{p} + \frac{1-\alpha}{r} \right)^{-1}, \tag{2.7}$$

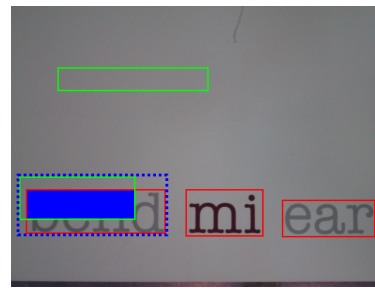
pri čemer je α enako 0,5. Evalvacija nad celotno zbirko ICDAR 2003 poteka tako, da se mere p, r in f določijo neodvisno za vsako sliko v zbirki posebej, kot končni rezultat pa se vzamejo povprečne mere p, r in f preko vseh slik zbirke. Nekateri avtorji se tega pravila ne držijo in namesto povprečnih mer računajo mere preko vseh anotiranih in detektiranih pravokotnikov v vseh slikah zbirke ter s tem povečajo vpliv slik z več teksta. Ker zahtevnejše slike tipično vsebujejo manj teksta, metode na ta način dosegajo višje rezultate.

2.2.2 *Rezultati tekmovanja ICDAR 2003*

Rezultati tekmovanja ICDAR 2003 [44] so prikazani v tabeli 2.1. Iz tabele je razvidno, da je bilo na tekmovanje prijavljenih zelo malo tekmovalcev, kar kaže, da gre za zelo zahtevno področje. Temu primerni so tudi rezultati, ki večinoma ne presegajo 50% natančnosti.

2.2.3 *Pomanjkljivosti zbirke*

Kljub temu, da je zbirka ICDAR 2003 precej razširjena, ima kar nekaj pomanjkljivosti. Velik problem zbirke je anotacija teksta s pravokotniki, ki onemogoča natančno anotacijo nehorizontalnih tekstov, med katere spadajo poševni (slika 1.2d) in ukrivljeni



Slika 2.3

Primer anotiranih in detektiranih pravokotnikov v zbirki ICDAR 2003. Rdeči pravokotniki ustrezajo anotiranim besedam, medtem ko zelena pravokotnika ustrezata detektiranemu tekstu. Polni modri pravokotnik označuje presek med anotiranim in detektiranim pravokotnikom, črtkani pravokotnik pa minimalni pravokotnik, ki ju obdaja.

Tabela 2.1

Rezultati tekmovanja ICDAR 2003 [44]. Poimenovanja metod so enaka kot v [44]. "Full" označuje slepo detekcijo, pri kateri je vedno detektiran le en pravokotnik, ki prekriva celotno sliko.

Metoda	*p*	*r*	*f*	*t (sek)*
Ashida	0,55	0,46	0,50	8,70
HWDavid	0,44	0,46	0,45	0,30
Wolf	0,30	0,44	0,35	17,00
Todoran	0,19	0,18	0,18	0,30
Full	0,10	0,06	0,08	0,20

teksti (sliki 1.2e in 1.2f). Primer anotacije nehorizontalnih tekstov je prikazan na sliki 2.4.

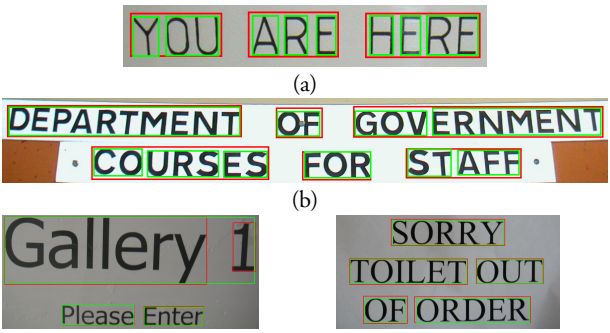
Ker so v zbirki ICDAR 2003 s pravokotniki anotirane posamezne besede, mora detektor teksta sam poskrbeti za pravilno grupiranje detektiranega teksta v besede, kar je s stališča objektivnosti evalvacije problematično. Metoda detekcije lahko v osnovi deluje zelo dobro, a zaradi slabega grupiranja v besede dosega slabe rezultate. Slike 2.5a, 2.5b in 2.5c prikazujejo primere popolnoma pravilne detekcije teksta, ki pa so zaradi nepravilnega grupiranja v besede kaznovani z nizkimi ocenami. Primer na sliki 2.5d prikazuje popolnoma pravilno detekcijo teksta in grupiranje v besede, vendar zaradi rahlega odstopanja med anotiranimi in detektiranimi pravokotniki detekcija dosega le 89% natančnost.

Pomanjkljivost zbirke ICDAR 2003 je tudi nekonsistentnost anotacije. Kljub specifikaciji, da sklopi *TrialTrain*, *TrialTest* in *Competition* vsebujejo 258, 251 in 544 slik, jih je v anotacijskih datotekah dejansko vključenih le 250, 249 in 501. Zanimivo je

(a) (b)

Slika 2.4

Primer problematične anotacije nehorizontalnih tekstov s pravokotniki v zbirki ICDAR 2003. Mali rdeči kvadratek znotraj velikega rdečega pravokotnika na sliki (a) označuje znak "®".



(a) (b) (c) (d)

Slika 2.5

Problem anotacije posameznih besed v zbirki ICDAR 2003. Z rdečo barvo so označeni anotirani pravokotniki, z zeleno pa detektirani. (a-c) Tekst na sliki je detektiran pravilno, vendar so zaradi napačnega grupiranja v besede preciznost p in priklic r le (a) 0,27 in 0,39, (b) 0,49 in 0,66 ter (c) 0,91 in 0,70. (d) Tekst na sliki je pravilno detektiran in grupiran v besede. Zaradi rahlega odstopanja med anotiranimi in detektiranimi pravokotniki sta preciznost p in priklic r le 0,89 in 0,89.

tudi, da na primer sklop *TrialTrain* vsebuje kar 12 dupliciranih slik, ki so navedene pod različnimi imeni z različnimi anotacijami, kljub temu, da gre za iste slike.

2.3 ICDAR 2005

Leta 2005 so v okviru konference ICDAR tekmovanje v detekciji teksta ponovili [50], pri čemer je bila za evalvacijo uporabljena ista zbirka kot leta 2003. Kadar literatura kot evalvacijsko zbirko navaja ICDAR 2005, gre dejansko za zbirko ICDAR 2003.

2.3.1 Rezultati tekmovanja ICDAR 2005

Rezultati tekmovanja ICDAR 2005 [50] so prikazani v tabeli 2.2. Glede na rezultate tekmovanja ICDAR 2003 je opaziti kar precejšnje izboljšanje natančnosti detekcije, saj

Tabela 2.2

Rezultati tekmovanja ICDAR 2005 [50]. Poimenovanja metod so enaka kot v [50]. "Full" označuje slepo detekcijo, pri kateri je vedno detektiran le en pravokotnik, ki prekriva celotno sliko. Z zvezdico so označeni rezultati tekmovanja ICDAR 2003.

Metoda	*p*	*r*	*f*	*t (sek)*
Hinnerk Becker	0,62	0,67	0,62	14,40
Alex Chen	0,60	0,60	0,58	0,35
Qiang Zhu	0,33	0,40	0,33	1,60
Jisoo Kim	0,22	0,28	0,22	2,20
Nobuo Ezaki	0,18	0,36	0,22	2,80
Ashida*	0,55	0,46	0,50	8,70
HWDavid*	0,44	0,46	0,45	0,30
Wolf*	0,30	0,44	0,35	17,00
Todoran*	0,19	0,18	0,18	0,30
Full	0,10	0,06	0,08	0,20

najvišje uvrščena metoda dosega natančnost, višjo od 60%.

2.4 ICDAR 2011

Po nekajletnem premoru so leta 2011 tekmovanje v robustnem branju teksta *Robust Reading Competition* v okviru konference ICDAR ponovili [45]. S tekmovanjem so želeli preveriti napredek metod detekcije teksta glede na leti 2003 in 2005. Organizatorji so tekmovanje razdelili v dve kategoriji: kategorijo branja računalniško ustvarjenih slik (angl. *born-digital images*) in kategorijo branja teksta v slikah naravnih scen. V nadaljevanju se ukvarjamo le z drugo kategorijo.

Zaradi številnih pomanjkljivosti zbirke ICDAR 2003, na katere so opozarjali raziskovalci, so se organizatorji tekmovanja odločili zbirko popolnoma prenoviti. Predvsem je šlo za pripombe v zvezi s problemom rahlo večjih detektiranih pravokotnikov (slika 2.5d), z nekonsistentnim označevanjem besed, neanotacijo nekaterih slik, nekonsistentnostjo pri označevanju posebnih znakov ter samo evalvacijsko shemo. Evalvacijska shema ICDAR 2003 namreč ni primerna za evalvacijo ujemanj *ena proti mnogo* (angl. *one-to-many matches*) in *mnogo proti ena* (angl. *many-to-one matches*), torej ujemanj, ko anotacijski pravokotnik ustreza več detektiranim pravokotnikom (slika 2.5a) oziroma ko detektiran pravokotnik pokriva več anotiranih pravokotnikov (slika 2.5c). Poleg tega priklic in preciznost, definirana v evalvacijski shemi ICDAR 2003, nista dovolj intuitivna. Kot navaja Wolf [51], lahko priklic z vrednostjo 0,5 pomeni, (a)

da je bilo detektiranih le 50% anotiranih pravokotnikov, (b) da so bili detektirani vsi anotirani pravokotniki, vendar le s 50% ujemanjem, ali pa (c) neki vmesni scenarij med ekstremnima primeroma (a) in (b). Prav tako je računanje povprečnega priklica in preciznosti glede na vse slike pomanjkljivo, saj zaradi nefaviriziranja slik z več teksta daje dvoumne rezultate.

Pri prenovi zbirke so avtorji uporabili del slik zbirke ICDAR 2003 in jim dodali nekaj novih slik, pri čemer so bile vse slike zaradi konsistentnosti na novo anotirane. Prenovljena je bila tudi evalvacijska shema, ki je opisana v nadaljevanju.

Učni in testni del zbirke ICDAR 2011 [45] vsebujeta 229 in 255 slik. Zbirka za anotacijo namesto datotek XML uporablja klasične tekstovne datoteke, pri čemer vsaki sliki v zbirki pripada anotacijska datoteka z istim korenom imena. Posamezna vrstica anotacijske datoteke ustreza določeni besedi v sliki in vsebuje koordinate pravokotnika, ki besedo obdaja, ter pripadajoči niz (angl. *string*) (slika 2.6).

(a)

```
130,250,206,283,"KEY"
129,209,206,246,"FOR"
96,172,235,211,"GLASS"
98,130,238,170,"BREAK"
```

(b)

Slika 2.6

Anotacija zbirke ICDAR 2011. (a) Primer slike. Anotirani pravokotniki so prikazani z rdečo barvo. (b) Pripadajoča anotacijska datoteka.

2.4.1 *Evalvacijska shema*

Zbirka ICDAR 2011 uporablja evalvacijsko shemo, ki jo predlagata Wolf in Jolion [51]. Za razliko od evalvacijske sheme ICDAR 2003 za izračun mer ne uporablja ploščin presekov najboljših ujemanj, temveč le število pravilnih ujemanj. Če T in E označujeta množico anotiranih in množico detektiranih pravokotnikov v posamezni sliki, sta priklic r in preciznost p nove evalvacijske sheme definirana kot:

$$r(T, E, t_r, t_p) = \frac{\sum_i m_T(T_i, E, t_r, t_p)}{|T|}, \tag{2.8}$$

$$p(T, E, t_r, t_p) = \frac{\sum_j m_E(E_j, T, t_r, t_p)}{|E|}, \tag{2.9}$$

pri čemer praga t_r in t_p določata minimalno stopnjo ujemanja ter ju bomo podrobneje opisali v nadaljevanju, funkciji ujemanja m_T in m_E pa sta definirani kot:

$$m_T(T_i, E, t_r, t_p) = \begin{cases} 1 & \textit{ena proti ena}, \\ 0 & \textit{ni ujemanja}, \\ f_{sc}(k) & \textit{ena proti mnogo}. \end{cases} \tag{2.10}$$

$$m_E(E_j, T, t_r, t_p) = \begin{cases} 1 & \textit{ena proti ena}, \\ 0 & \textit{ni ujemanja}, \\ f_{sc}(k) & \textit{mnogo proti ena}. \end{cases} \tag{2.11}$$

Preprosto povedano, funkcija m_T določa, ali se anotirani pravokotnik T_i ujema s katerim od pravokotnikov v množici E. Če se pravokotnik ujema z natanko enim detektiranim pravokotnikom (ujemanje *ena proti ena*), funkcija m_T dobi vrednost 1, če ujemanja ni, pa vrednost 0. Podobno funkcija m_E določa, ali se detektirani pravokotnik E_j ujema s katerim od pravokotnikov v množici T. Če se pravokotnik ujema z natanko enim anotiranim pravokotnikom (ujemanje *ena proti ena*), funkcija m_E dobi vrednost 1, če ujemanja ni, pa vrednost 0. Primera *ena proti mnogo* (enačba (2.10)) in *mnogo proti ena* (enačba (2.11)) sta posebna primera ujemanja. Primer *ena proti mnogo* nastopi, kadar več detektiranih pravokotnikov pravilno pokriva posamezni anotiran pravokotnik (slika 2.5a), medtem ko primer *mnogo proti ena* nastopi, ko detektiran pravokotnik pravilno pokriva več anotiranih pravokotnikov (slika 2.5c). Oba primera evalvacijska shema kaznuje s funkcijo kazni $f_{sc}(k)$, pri čemer k ustreza številu pravokotnikov, ki jih posamezni pravokotnik pokriva. Evalvacijska shema ICDAR 2011 uporablja konstantno funkcijo kazni f_{sc} z vrednostjo 0,8.

Za razliko od evalvacijske sheme ICDAR 2003, ki primerov *ena proti mnogo* in *mnogo proti ena* ne obravnava pravilno, nova evalvacijska shema takšne primere obravnava kot pravilne, kljub temu pa jih še vedno kaznuje.

Za ugotavljanje, ali gre za ujemanje med določenim anotiranim in detektiranim pravokotnikom, in če gre, za kakšno vrsto ujemanja gre (*ena proti ena*, *ena proti mnogo*, *mnogo proti ena*), evalvacijska shema uporablja koncept matrik prekrivanja (angl. *overlap matrix*) [52]. Matrika prekrivanja je matrika velikosti $|T| \times |E|$, pri čemer T in E označujeta množico anotiranih in množico detektiranih pravokotnikov v posamezni sliki, element matrike (i, j) pa označuje par, ki ga sestavljata i-ti anotirani pravokotnik iz množice T in j-ti detektirani pravokotnik iz množice E. Za vsako sliko posebej se

generirata matriki prekrivanja σ in τ z naslednjimi vrednostmi:

$$\sigma_{ij} = \frac{area(T_i \cap E_j)}{area(T_i)}, \tag{2.12}$$

$$\tau_{ij} = \frac{area(T_i \cap E_j)}{area(E_j)}, \tag{2.13}$$

pri čemer T_i in E_j označujeta i-ti anotirani pravokotnik iz množice T in j-ti detektirani pravokotnik iz množice E, $area()$ pa označuje površino regije. Na podlagi vrednosti matrik σ in τ evalvacijska shema določi, za kakšno vrsto ujemanja gre, in informacijo uporabi pri izračunu funkcij ujemanja v enačbah (2.10) in (2.11):

a) Ujemanje *ena proti ena*

Če i-ta vrstica matrike σ in matrike τ vsebuje natanko en element, ki ustreza pogoju v enačbi (2.14), in če j-ta kolona obeh matrik vsebuje natanko en element, ki ustreza istemu pogoju, gre za ujemanje ena proti ena.

$$\begin{aligned} \sigma_{ij} &> t_r \,, \\ \tau_{ij} &> t_p \,. \end{aligned} \tag{2.14}$$

b) Ujemanje *ena proti mnogo*

Ujemanje ena proti mnogo nastopi, kadar se anotiran pravokotnik T_i ujema z množico detektiranih pravokotnikov v E:

$$\begin{aligned} \sum_{j \in E} \sigma_{ij} &\geq t_r \,, \\ \forall j \in E : \tau_{ij} &\geq t_p \,. \end{aligned} \tag{2.15}$$

c) Ujemanje *mnogo proti ena*

Ujemanje mnogo proti ena nastopi, kadar detektiran pravokotnik E_j pravilno pokriva

množico anotiranih pravokotnikov v T:

$$\begin{aligned} &\forall i \in T : \sigma_{ij} \geq t_r \,, \\ &\sum_{i \in T} \tau_{ij} \geq t_p \,. \end{aligned} \tag{2.16}$$

Kot smo v poglavju 2.2.1 že omenili, evalvacijska shema ICDAR 2003 priklic in preciznost računa neodvisno za vsako sliko posebej in ju nato povpreči preko vseh slik zbirke. Ker so na takšen način vse slike obravnavane enakovredno, ne glede na to, koliko črk vsebujejo, evalvacijska shema ICDAR 2011 priklic in preciznost računa glede na vse pravokotnike v vseh slikah zbirke:

$$r(\bar{T}, \bar{E}, t_r, t_p) = \frac{\sum_k \sum_i m_T(T_i^k, E^k, t_r, t_p)}{\sum_k |T^k|}, \tag{2.17}$$

$$p(\bar{T}, \bar{E}, t_r, t_p) = \frac{\sum_k \sum_j m_E(E_j^k, T^k, t_r, t_p)}{\sum_k |E^k|}. \tag{2.18}$$

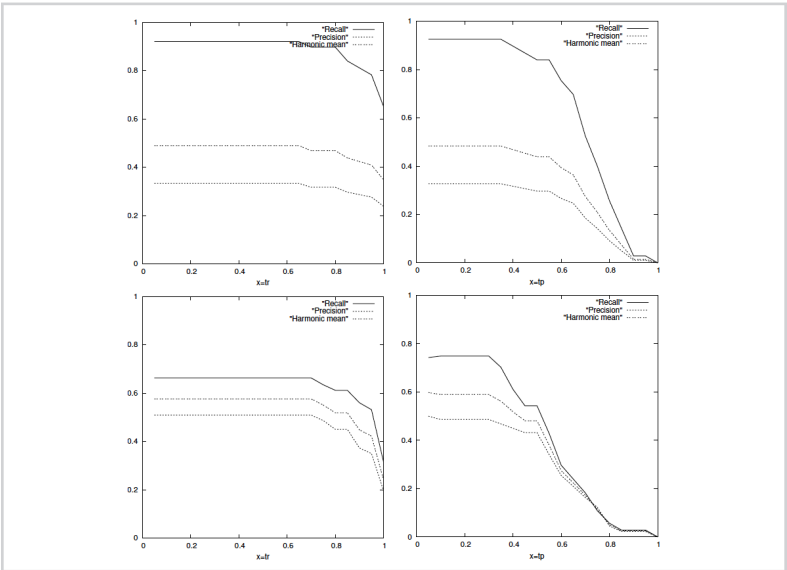
pri čemer $\bar{T}$ in $\bar{E}$ označujeta množici vseh anotiranih in vseh detektiranih pravokotnikov v vseh slikah.

Gibanje evalvacijskih mer r in p pri različnih pragih t_r in t_p lahko vizualno predstavimo s krivuljami mer. Primeri krivulj mer za metodi [53] in [54] so prikazani na sliki 2.7. Opazimo lahko, da so mere močno odvisne od vrednosti pragov. Tako se lahko zgodi, da določena metoda pri nekem pragu dosega boljše rezultate od neke druge metode, medtem ko pri drugačnem pragu dosega slabše rezultate. Da bi bile končne evalvacijske mere čim bolj objektivne in da bi odražale uspešnost posamezne metode preko celotnega spektra različnih pragov, sta končni priklic r' in končna preciznost p' definirana kot povprečje r oziroma p pri vseh vrednostih pragov t_r in t_p:

$$r' = \frac{1}{2d} \sum_{i=1}^{d} r(\bar{T}, \bar{E}, i/d, t_p) + \frac{1}{2d} \sum_{i=1}^{d} r(\bar{T}, \bar{E}, t_r, i/d), \tag{2.19}$$

$$p' = \frac{1}{2d} \sum_{i=1}^{d} p(\bar{T}, \bar{E}, i/d, t_p) + \frac{1}{2d} \sum_{i=1}^{d} p(\bar{T}, \bar{E}, t_r, i/d). \tag{2.20}$$

pri čemer parameter d določa stopnjo kvantizacije. Avtorji v [51] uporabljajo stopnjo



Slika 2.7

Krivulje mer, ki prikazujejo gibanje evalvacijskih mer ICDAR 2011 (slika je povzeta po [51]). Zgornja vrstica ustreza algoritmu [53], spodnja pa algoritmu [54]. Leva kolona prikazuje gibanje evalvacijskih mer pri spremenljivem pragu t_r, pri čemer je prag t_p fiksiran na vrednost 0,4. Desna kolona prikazuje gibanje evalvacijskih mer pri spremenljivem pragu t_p, pri čemer je prag t_r fiksiran na vrednost 0,8.

kvantizacije 20. Skupna mera učinkovitosti f' se izraža kot:

$$f' = 2 \cdot \frac{p' \cdot r'}{p' + r'}. \tag{2.21}$$

2.4.2 *Rezultati tekmovanja ICDAR 2011*

Tekmovanje ICDAR 2011 [45] je potekalo v odprtem načinu (angl. *open mode*). Štirinajst dni pred iztekom roka za oddajo rezultatov detekcije je bil javno objavljen učni del zbirke z 229 slikami, ki so ga tekmovalci lahko uporabili za učenje svojih metod, testni del zbirke z 255 slikami pa je bil javno objavljen tri dni pred iztekom roka. Do izteka roka so morali tekmovalci posredovati rezultate detekcije na testnem delu zbirke. Rezultati tekmovanja so prikazani v tabeli 2.3. Za primerjavo so v tabeli 2.4 prikazani rezultati tekmovanja ICDAR 2003 na podlagi nove metrike ICDAR 2011.

2.4.3 *Pomanjkljivosti zbirke*

Problem evalvacijske sheme zbirke ICDAR 2003 je napačno obravnavanje ujemanj *ena proti mnogo* in *mnogo proti ena*. Evalvacijska shema zbirke ICDAR 2011 ta problem

Tabela 2.3

Rezultati tekmovanja ICDAR 2011 [45]. Poimenovanja metod so enaka kot v [45].

Metoda	*r' (%)*	*p' (%)*	*f' (%)*
Kim	62,47	82,98	71,28
Yi	58,09	67,22	62,32
TH-TextLoc	57,68	66,97	61,98
Neumann	52,54	68,93	59,63
TDM_IACS	53,52	63,52	58,09
LIP6-Retin	50,07	62,97	55,78
KAIST AIPR	44,57	59,67	51,03
ECNU-CCG	38,32	35,01	36,59
Text Hunter	25,96	50,05	34,19

Tabela 2.4

Primerjava rezultatov tekmovanja ICDAR 2003 pri uporabi metrike ICDAR 2003 in metrike ICDAR 2011. Poimenovanja metod so enaka kot v [51].

Metoda	*Metrika ICDAR 2003*			*Metrika ICDAR 2011*		
	r	*p*	*f*	*r'*	*p'*	*f'*
Ashida	46,00	55,00	50,00	41,70	55,30	47,50
H. W. David	46,00	44,00	45,00	46,60	39,60	42,80
Wolf et al.	44,00	30,00	35,00	44,90	19,40	27,10
Todoran	18,00	19,00	18,00	17,90	14,30	15,90

odpravlja le delno, saj takšna ujemanja sicer pravilno detektira, kljub temu pa jih še vedno kaznuje. Na ta način vpliv nepravilnega grupiranja na evalvacijo zmanjšuje (v primerjavi z zbirko ICDAR 2003), v celoti pa ga ne odpravlja.

Podobno kot zbirka ICDAR 2003 tudi zbirka ICDAR 2011 za anotacijo teksta uporablja pravokotnike, kar onemogoča pravilno anotacijo nehorizontalnih tekstov.

2.5 SVT

Ena izmed novejših zbirk slik teksta v naravnih scenah je zbirka SVT (angl. *Street View Text*) [46]. Zbirka vsebuje 350 slik, ki so pridobljene iz spletne aplikacije *Google Street View*. Učni del zbirke obsega 100 slik, testni del pa 250 slik. Primarno je zbirka SVT namenjena odkrivanju besed v slikah[4] (angl. *word spotting*), ki se od klasične detekcije

[4]Pri odkrivanju besed v sliki (angl. *word spotting*) gre za problem lociranja ene ali več vhodnih besed na določeni sliki.

teksta precej razlikuje, zato je temu primerna tudi drugačna anotacija slik v zbirki. Avtorji so za vsako sliko posebej na podlagi njene geolokacije z orodjem *Search Nearby* poiskali vse trgovsko-poslovne lokacije (angl. *business locations*) v njeni bližini in na podlagi 20 najvišje uvrščenih lokacij generirali leksikon s 50 besedami. Vse besede leksikona, ki so bile prisotne v sliki, so anotirali s pravokotniki.

Primer slike zbirke SVT s pripadajočo anotacijsko datoteko XML je prikazan na sliki 2.8. Poleg informacije o imenu, ločljivosti in geolokaciji slike anotacijska datoteka vsebuje leksikon besed (značka <*lex*>) in seznam pravokotnikov okoli tistih besed na sliki, ki so v leksikonu (značka <*taggedRectangles*>). Na sliki 2.8 so anotirane le besede "THE", "triple" in "door" (beseda "THE" se v sliki pojavi trikrat), saj so edine besede leksikona, ki so prisotne na sliki. Ostale besede na sliki niso anotirane, saj se ne pojavljajo v leksikonu.

2.5.1 Pomanjkljivosti zbirke

Zbirka SVT je sicer heterogena in kompleksna, vendar nepopolna anotacija (anotacija le določenih besed) predstavlja veliko pomanjkljivost pri evalvaciji metod detekcije teksta. Vprašanje je namreč, kako obravnavati pravilno detektiran tekst, ki ni anotiran.

Podobno kot pri zbirkah ICDAR 2003 in ICDAR 2011 je problem zbirke SVT tudi anotacija teksta s pravokotniki.

2.6 CVL OCR DB

CVL OCR DB [47] je javno dostopna zbirka anotiranih slik teksta v naravnih scenah.[5] Za postavitev zbirke smo se odločili zaradi že omenjenih pomanjkljivosti zbirk ICDAR 2003, ICDAR 2011 in SVT.

Velik problem zbirke ICDAR 2003 je nepravilno obravnavanje ujemanj *ena proti mnogo* in *mnogo proti ena*. Zbirka ICDAR 2011 kljub mehanizmu, ki jih pravilno detektira, takšna ujemanja še vedno kaznuje. Višina kazni je definirana v funkciji kazni f_{sc} (enačbi (2.10) in (2.11)). Posledično je rezultat evalvacije posamezne metode močno odvisen od modula za grupiranje teksta v besede, kar poveča subjektivnost evalvacije. Kot sta pokazala Neumann in Matas [55], grupiranje teksta zelo vpliva na končni rezultat evalvacije.

[5] http://lrv.fri.uni-lj.si/andreji/CVLOCRDB/

(a)

```
<image>
   <imageName>img/18_01.jpg</imageName>
   <address>? 216 Union Street Seattle UA</address>
   <lex>THE,TRIPLE,DOOR,WILD,GINGER,ASIAN,RESTAURANT,GELATIAMO,SEATTLE,ART,MUSEUM,BENAROYA,HALL,POST,OFFICE,
        SYMPHONY,SHOWBOX,THE,MARKET,BYRNIE,UTZ,HATS,DOWNTOWN,EPICENTER,WASHINGTON,MUTUAL,INC,FINANCIAL,SERVICES,
        HARRIED,HUNGRY,CATERING,CAFE,ROSS,DRESS,FOR,LESS,REPUBLIC,PARKING,UNION,WILLIAM,TRAVER,GALLERY,MARCELLA,
        BOUTIQUE,SKINNER,BUILDING,BRUNO,MEXICAN,ITALIAN</lex>
   <Resolution x="1280" y="800"/>
   <taggedRectangles>
      <taggedRectangle height="79" width="186" x="539" y="349">
         <tag>DOOR</tag>
      </taggedRectangle>
      <taggedRectangle height="38" width="71" x="614" y="633">
         <tag>THE</tag>
      </taggedRectangle>
      <taggedRectangle height="41" width="68" x="527" y="666">
         <tag>THE</tag>
      </taggedRectangle>
      <taggedRectangle height="86" width="189" x="462" y="257">
         <tag>TRIPLE</tag>
      </taggedRectangle>
      <taggedRectangle height="32" width="50" x="562" y="217">
         <tag>THE</tag>
      </taggedRectangle>
   </taggedRectangles>
</image>
```

(b)

Slika 2.8

Anotacija zbirke SVT. (a) Primer slike. Anotirani pravokotniki so prikazani z rdečo barvo. (b) Izsek pripadajoče anotacijske datoteke XML.

Zbirka SVT je namenjena odkrivanju vnaprej določenih besed v sliki, kar pomeni, da so anotirane le določene besede slike in ne vse. Tudi če bi na zbirki SVT testirali optimalno metodo detekcije teksta, ki bi pravilno detektirala celoten tekst v sliki, rezultati ne bi bili optimalni. Glede na to, da bi pravilno detektirali vse anotirane pravokotnike, bi dosegli maksimalen možen priklic, vendar pa bi po drugi strani dosegli tudi zelo nizko preciznost, saj bi število detektiranih pravokotnikov (kljub temu, da bi ti dejansko ustrezali tekstu) močno presegalo število anotiranih.

Skupna pomanjkljivost zbirk ICDAR 2003, ICDAR 2011 in SVT je anotacija teksta s pravokotniki, ki močno otežuje anotacijo nehorizontalnih tekstov (poševnih in ukrivljenih), ki so v naravnih scenah zelo pogosti – bodisi zaradi same nehorizontalne

oblike teksta bodisi zaradi perspektivnih projekcij, ki se pojavljajo pri zajemu slik (slika 2.4).

Zbirka CVL OCR DB omenjene pomanjkljivosti ostalih zbirk odpravlja z dvema tipoma anotacije, in sicer z anotacijo z *n*-poligoni in z binarno anotacijo, ki ju podrobno opisujemo v nadaljevanju.

2.6.1 Organizacija CVL OCR DB

Zbirka CVL OCR DB [47] vsebuje 341 anotiranih slik teksta v formatu JPEG. Glede na svetlobne pogoje zajema so slike v zbirki razdeljene v tri glavne kategorije: *day*, *night* in *artifical*, ki ustrezajo zajemu podnevi, ponoči in pri umetni osvetlitvi (npr. v nakupovalnih središčih). Kategorija *day* se dodatno deli na podkategorije *normal*, *fog*, *rain* in *sun*, ki ustrezajo zajemu pri normalnih pogojih ter v meglenem, deževnem in sončnem vremenu, kategorija *night* pa se deli na podkategoriji *dusk* in *night*, ki ustrezata zajemu ob mraku ter zajemu pri popolni temi (seveda ob prisotnosti umetnih svetil, kot so ulične svetilke, bliskavica in podobno). Primeri slik posameznih kategorij oz. podkategorij so prikazani na sliki 2.9.

Zbirka CVL OCR DB je urejena v obliki hierarhične direktorijske strukture, ki jo ponazarja slika 2.10. Na najnižjem nivoju so direktoriji z dejanskimi podatki, ki jih imenujemo podatkovna vozlišča. Podatkovna vozlišča so določena z enoličnim imenom, ki je sestavljeno iz prve črke imena ustrezne kategorije, prve črke imena ustrezne podkategorije ter petmestne zaporedne številke podatkovnega vozlišča v posamezni podkategoriji. Tako *d_n_00001* označuje prvo podatkovno vozlišče kategorije *day* in podkategorije *normal* (slika 2.10a).

Vsako podatkovno vozlišče vsebuje natanko eno (originalno) sliko, anotacijsko datoteko XML in binarno anotacijsko datoteko. Vse datoteke so poimenovane po podatkovnem vozlišču, ki mu pripadajo. Tako so v primeru podatkovnega vozlišča *d_n_00001* pripadajoča originalna slika, anotacijska datoteka XML in binarna anotacijska datoteka poimenovane z *d_n_00001_full.jpg*, *d_n_00001.xml* ter *d_n_00001_gt.bmp*[6] (slika 2.10a).

Vsako podatkovno vozlišče dodatno vsebuje tudi izrezane sličice posameznih črk v sliki. Te sicer nimajo neposredne povezave s samo detekcijo teksta, vsekakor pa prispevajo k dodani vrednosti zbirke, saj lahko služijo za učenje in testiranje klasifikatorjev

[6]Niz "_gt" v poimenovanju binarne anotacijske datoteke označuje, da gre za anotacijsko datoteko. Črki "g" in "t" ustrezata začetnicama angleškega poimenovanja *Ground Truth*.

(a) (b) (c) (d) (e) (f) (g)

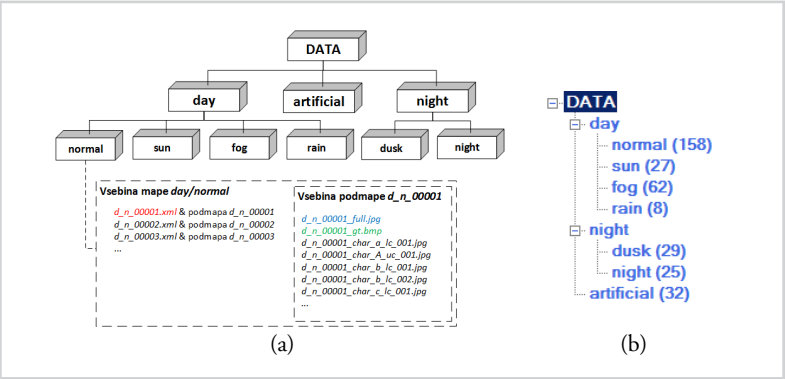
Slika 2.9

Primeri slik posameznih kategorij in podkategorij zbirke CVL OCR DB: (a) *day*|*normal*, (b) *day*|*fog*, (c) *day*|*rain*, (d) *day*|*sun*, (e) *night*|*dusk*, (f) *night*|*night* in (g) *artificial*.

OCR. Poimenovanje sličic črk je podobno poimenovanju ostalih datotek v podatkovnem vozlišču. Za primer vzemimo ime *d_n_00001_char_a_lc_001.jpg* na sliki 2.10a. Predponi *d_n_00001* sledi medpona *char_a*, ki označuje, da gre za sličico črke "a". Medpona *lc*, ki sledi, označuje, da gre za malo tiskano črko (angl. *lowercase*). V primeru velike tiskane črke bi bila uporabljena medpona *uc* (angl. *uppercase*). Ime datoteke se zaključi z zaporedno številko *001*, ki označuje, da gre za sliko prve male tiskane črke "a" v tekstu. Zaporedna številka je zelo pomembna, saj lahko v sliki nastopa več malih tiskanih črk "a".

2.6.2 *Anotacija z n-poligoni*

Kot smo na začetku poglavja omenili, je ena izmed slabosti zbirk ICDAR 2003, ICDAR 2011 in SVT anotacija nehorizontalnih tekstov s pravokotniki. Da bi omogočili na-



Slika 2.10

Hierarhična direktorijska struktura zbirke CVL OCR DB. Številke poleg kategorij na sliki (b) označujejo število slik v posamezni kategoriji.

tančnejšo anotacijo, so slike v zbirki anotirane z *n*-poligoni[7]. Slika 2.11a prikazuje anotirano sliko zbirke ICDAR 2011 z nehorizontalnim tekstom, ki je posledica perspektivne projekcije pri zajemu slike. Anotacija s pravokotniki ni ustrezna, saj pravokotniki zajemajo tudi del slike, kjer teksta ni. Slika 2.11b prikazuje anotacijo iste slike z *n*-poligonoma, ki se natančneje prilegata tekstu.

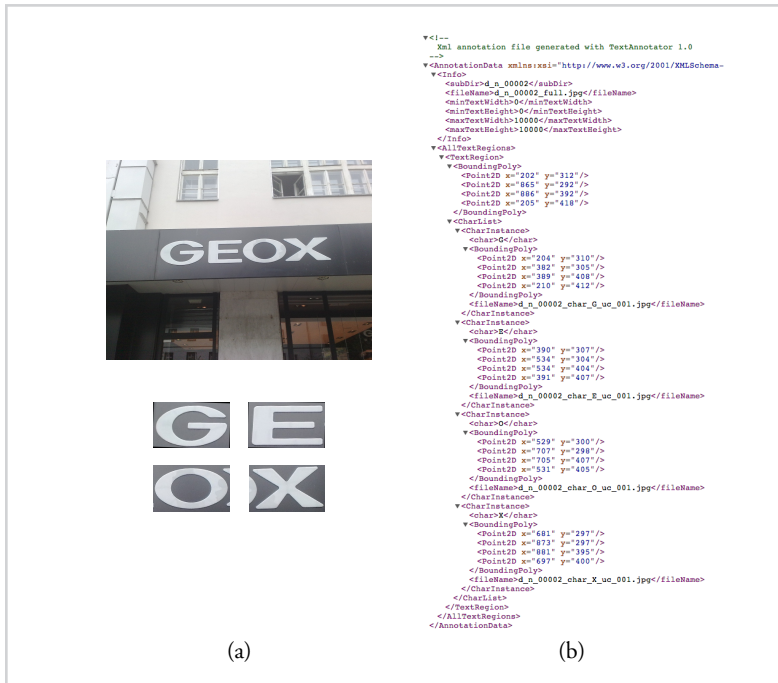
Slika 2.11

Primerjava anotacije s pravokotniki in *n*-poligoni. (a) Slika zbirke ICDAR 2011, anotirana s pravokotniki. (b) Ista slika, anotirana z *n*-poligoni.

Primer slike zbirke CVL OCR DB s pripadajočo anotacijsko datoteko XML in izrezanimi sličicami črk je prikazan na sliki 2.12. Anotacijska datoteka XML na najvišjem nivoju vsebuje značko <*Info*> s splošnimi podatki ter značko <*AllTextRegions*>, ki vsebuje seznam posameznih anotiranih regij <*TextRegion*>. Vsaka regija je definirana z *n*-poligonom <*BoundingPoly*>, ki ga sestavlja poljubno število oglišč <*Point2D*>.

[7]Kljub temu, da gre pri *n*-poligonu dejansko za klasičen poligon, smo s predpono *n* želeli poudariti, da je definiran s poljubnim številom oglišč.

Dodatno regija vsebuje tudi seznam pripadajočih črk <*CharList*>. Vsaka črka je predstavljena z značko <*CharInstance*>, ki vsebuje ime črke, *n*-poligon, ki črko obdaja, ter ime datoteke z izrezano sličico črke. Podrobna struktura anotacijske datoteke XML je prikazana na sliki 2.13.

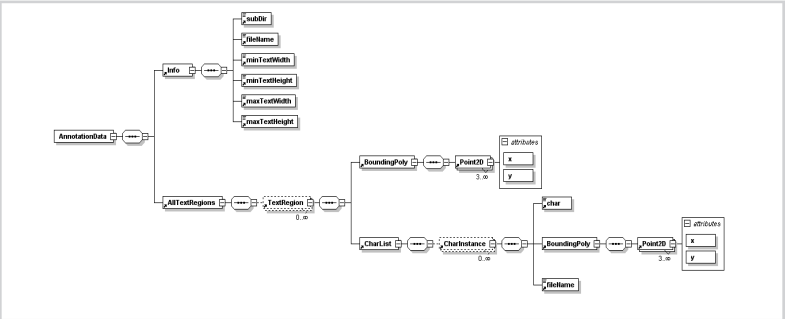


Slika 2.12

(a) Primer slike zbirke CVL OCR DB z izrezanimi sličicami posameznih črk. (b) Izsek pripadajoče anotacijske datoteke XML.

Za poenostavitev anotacije slik z *n*-poligoni smo razvili poseben anotacijski program *TextAnnotator*, ki omogoča označevanje tekstovnih regij v sliki s poligoni, ekstrakcijo sličic posameznih črk teksta ter shranjevanje anotacijskih podatkov v obliki datotek XML.

Na prvi pogled se zdi, da anotacija z *n*-poligoni zaplete postopek evalvacije, ki je v primeru anotacije s pravokotniki zelo enostaven (glej evalvacijski shemi ICDAR 2003 in ICDAR 2011), vendar se izkaže, da je koncept evalvacije možno aplicirati tudi na *n*-poligone. Vse, kar je potrebno, je, da metoda detekcije teksta namesto pravokotnikov generira minimalne konveksne poligone, ki obdajajo detektiran tekst v sliki, in jih



Slika 2.13

XSD-shema anotacijske datoteke XML zbirke CVL OCR DB.

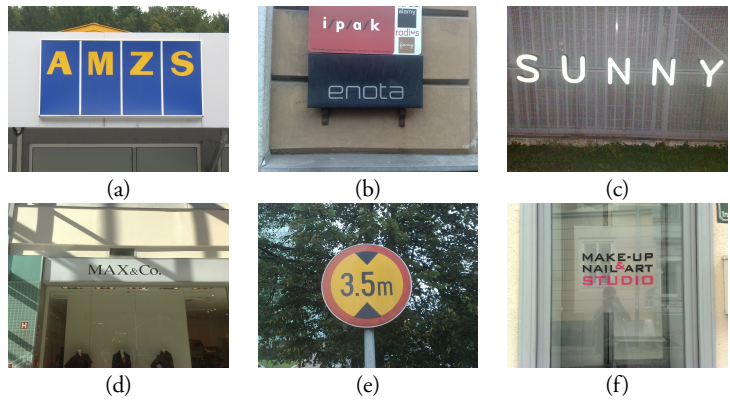
shrani v obliki *n*-poligonov. V takšnem primeru je možno uporabiti tako evalvacijsko shemo ICDAR 2003 kot evalvacijsko shemo ICDAR 2011, le koncept ujemanja pravokotnikov nadomesti koncept medsebojnega ujemanja poligonov.

Na koncu omenimo, da kljub nadgradnji koncepta anotacijskih pravokotnikov z anotacijskimi *n*-poligoni zbirka CVL OCR DB še vedno ostaja kompatibilna z evalvacijskima shemama zbirk ICDAR 2003 in ICDAR 2011. Vsak anotiran *n*-poligon je namreč možno predstaviti z minimalnim pravokotnikom, ki ga obdaja, in zbirko tako evalvirati na klasičen način (ICDAR).

2.6.3 Binarna anotacija

Anotacija z *n*-poligoni omogoča natančnejšo anotacijo teksta kot anotacija s pravokotniki, še vedno pa ne rešuje problema osnovne anotacijske enote oziroma, z drugimi besedami, kateri je osnovni gradnik teksta, ki se ga anotira. So to vrstice, besede, posamezne črke? Je pomišljaj del besede ali jo razdeli na dve besedi? Je pika del besede ali ne? Kako velik mora biti presledek, da loči dve besedi? Tudi če se eksaktno definira, kaj je osnovna anotacijska enota, takšna evalvacijska shema še vedno zahteva, da je detektiran tekst pravilno grupiran v ustrezne anotacijske enote, kar je izredno težavno tako za implementacijo kot za objektivno evalvacijo. Slika 2.14 prikazuje nekaj primerov, kjer je težko definirati, kaj natančno so osnovne anotacijske enote in kako tekst pravilno grupirati. Binarna anotacija, ki jo uporablja zbirka CVL OCR DB, potrebo po pravilnem grupiranju teksta popolnoma odpravlja.

Primera originalne slike in binarno anotirane slike sta prikazana na slikah 2.15a in 2.15b. Pri binarni anotaciji so slikovni elementi slike, ki pripadajo tekstu, označeni z



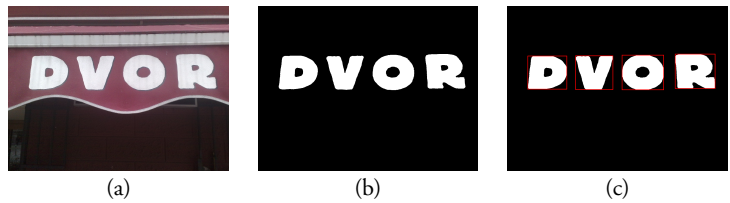
(a) (b) (c)

(d) (e) (f)

Slika 2.14

Problem izbire ustrezne anotacijske enote. (a) "AMZS" lahko anotiramo kot eno besedo ali kot štiri ločene črke. (b) Napis "ipak" lahko ustreza eni anotirani besedi ali štirim anotiranim črkam. (c) "SUNNY" je beseda z velikimi razmaki med posameznimi črkami. (d) Je "&" del skupne besede ali deli tekst na "MAX" in "Co"? (e) Dejansko gre za dve besedi: "3.5" in "m". Anotacija to upošteva? (f) Kakšna je funkcija znakov "-" in "&"?

neničelno vrednostjo, vsi ostali pa z nič. S povezovanjem neničelnih slikovnih elementov v povezane komponente (angl. *connected-components*) dobimo posamezne črke v sliki (slika 2.15c), kar omogoča neposredno evalvacijo metod detekcije teksta na nivoju posameznih črk. Tako dodatno grupiranje detektiranega teksta ni več potrebno. Še več, binarna anotacija omogoča zelo natančno anotacijo nehorizontalnih tekstov, saj ni omejena z medsebojno lego posameznih črk v sliki.



(a) (b) (c)

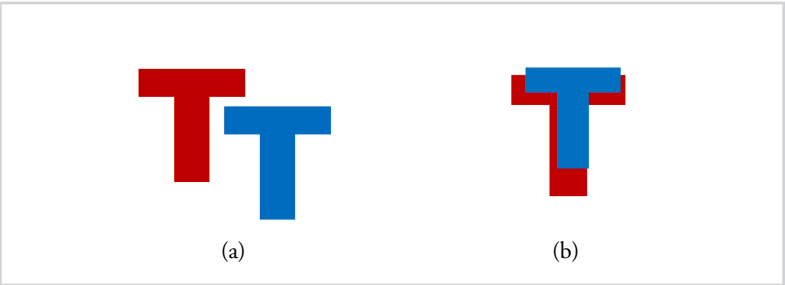
Slika 2.15

Primer binarne anotacije zbirke CVL OCR DB. (a) Originalna slika. (b) Binarno anotirana slika. (c) Povezane komponente, ki ustrezajo črkam.

Evalvacija metod detekcije teksta na binarno anotiranih slikah poteka na podoben način kot v [48]. Označimo z G množico vseh anotiranih črk v sliki in z D množico vseh detektiranih črk v sliki. Za vsako črko C_G v množici G poiščemo njeno najboljše ujemanje C_D v množici D. Če je razmerje ploščin črk C_G in C_D večje od praga T_R[8]

[8]Črka "R" v T_R ustreza prvi črki angleške besede *rectangle* in označuje, da gre za prag ujemanja pravokotni-

in če je razmerje ploščin $C_D \cap C_G$ in črke C_G večje od praga T_R, je anotirana črka pravilno detektirana in govorimo o ujemanju. V nasprotnem primeru govorimo o neujemanju. Razmerje med številom pravilno detektiranih in številom vseh anotiranih črk v sliki imenujemo stopnja detekcije (angl. *detection rate*) na posamezni sliki. Primer neujemanja in ujemanja anotirane in detektirane črke je prikazan na sliki 2.16.



Slika 2.16

Primer neujemanja in ujemanja detektiranih in anotiranih črk. Z rdečo barvo so označene anotirane, z modro pa detektirane črke. (a) Črki se ne ujemata. (b) Ujemanje črk je odvisno od praga T_R. Če je prag dovolj nizek, beležimo ujemanje črk, v nasprotnem primeru beležimo neujemanje.

Uspešnost metod detekcije teksta na celotni zbirki CVL OCR DB merimo s povprečno in z absolutno stopnjo detekcije. Povprečna stopnja detekcije označuje povprečje stopenj detekcije na posameznih slikah zbirke, medtem ko absolutna stopnja detekcije označuje razmerje med številom vseh pravilno detektiranih črk in številom vseh anotiranih črk v vseh slikah zbirke.

Povprečna stopnja detekcije, ki jo uporabljata Nikolaou in Papamarkos [48], obravnava vse slike enakovredno, ne glede na število črk, ki se pojavljajo v njih. Vzemimo primer zbirke z dvema slikama, pri čemer vsebuje prva slika 4 črke, druga pa 400 črk. Če metoda detekcije teksta na prvi sliki pravilno detektira 4 črke, na drugi pa 200 črk, dobimo povprečno stopnjo detekcije $(4/4 + 200/400)/2 = 0,75$. Stopnja detekcije je relativno visoka, kljub temu, da metoda praktično ni detektirala skoraj polovice vseh črk. Absolutna stopnja detekcije uspešnost meri bolj realno, saj je za zgornji primer enaka $204/404 \simeq 0,5$.

2.7 Diskusija

Glavna pomanjkljivost zbirk ICDAR 2003, ICDAR 2011 in SVT je anotacija teksta s pravokotniki, ki ustrezajo posameznim besedam v tekstu. Takšen način anotacije

kov.

zahteva, da metoda detekcije teksta sama poskrbi za pravilno grupiranje teksta v besede. Določena metoda lahko tekst detektira zelo dobro, vendar zaradi slabega grupiranja (ki je lahko vseeno pravilno s stališča klasifikatorja OCR) dosega zelo nizke rezultate (slika 2.5).

Binarna anotacija zbirke CVL OCR DB problem grupiranja odpravlja, saj omogoča evalvacijo na nivoju posameznih črk. Pravzaprav je grupiranje teksta v besede samo sebi namen, saj je namenjeno izključno evalvaciji in nima direktnega vpliva na delovanje tovrstnih sistemov v praksi. Šele ko so črke razpoznane s klasifikatorji OCR, jih je možno (skoraj) pravilno grupirati v besede z uporabo leksikonov.

Kljub številnim prednostim je treba poudariti, da je postopek binarne anotacije mnogo zahtevnejši od anotacije slik s pravokotniki. To je najbrž tudi razlog, da se marsikdo odloči za anotacijo slik s pravokotniki. Seveda pa je binarna anotacija kljub vsemu veliko bolj natančna od anotacije s pravokotniki in omogoča bolj objektivno evalvacijo.

3

Obstoječe metode detekcije teksta

3.1 Uvod

Obstaja veliko metod detekcije teksta v slikah naravnih scen. Kljub njihovi raznovrstnosti prevladujeta dva glavna pristopa: regijski in teksturni. Pri regijskem pristopu se slikovni elementi s karakteristikami teksta povezujejo v povezane komponente, izmed katerih se izberejo tiste, ki ustrezajo vnaprej določenim geometrijskim pravilom. Za razliko od regijskega pristopa, ki deluje od spodaj navzgor (angl. *bottom-up approach*), teksturne metode delujejo od zgoraj navzdol (angl. *top-down approach*). S premikanjem preiskovalnega okna (angl. *sliding window*) pri različnih skalah piramide iščejo regije, ki ustrezajo določenim lastnostim teksta. Teksturne metode so zaradi piramidnega pristopa časovno kompleksnejše, medtem ko so regijske metode tipično hitrejše. Poleg omenjenih obstajajo tudi hibridne metode, ki z združevanjem regijskega in teksturnega pristopa izkoriščajo prednosti obeh.

Dodatno se metode detekcije teksta delijo na omejitveno in učno usmerjene. Omejitveno usmerjene metode za filtriranje netekstovnih regij uporabljajo množico geometrijskih pravil, medtem ko učno usmerjene metode temeljijo na uporabi klasifikatorjev strojnega učenja.

Rangiranje obstoječih metod ni enostavno, saj za evalvacijo uporabljajo različne zbirke. Tudi kadar je določena metoda evalvirana na zbirki ICDAR 2003, še vedno ni natančno določeno, na katerem delu zbirke je in na kakšen način. Nekatere metode za evalvacijo namreč uporabljajo le del zbirke [32, 34, 40, 56] ali celo kombinacijo zbirke in dodatnih slik [35, 57]. Avtorji metod pogosto ne specificirajo, kateri del zbirke je bil uporabljen za učenje in kateri del za evalvacijo. Metoda [38] za učenje celo uporablja tako učni kot testni del zbirke, evalvira pa jo na delu *Competition*. V primeru omejitveno usmerjenih metod, ki ne uporabljajo klasifikatorjev, pogosto ni specificirano, kako je potekala evalvacija: ali je bil uporabljen le testni del zbirke (glede na to, da faza učenja ni potrebna) ali celotni del zbirke. Prav tako se za podajanje evalvacije lahko uporabljajo bodisi povprečne mere bodisi posamezne mere na podlagi vseh pravokotnikov v vseh slikah [33], kot je opisano v poglavju 2.2.1. Na podlagi povedanega se upravičeno postavlja vprašanje, kako se je izvajala evalvacija preostalih metod, tj. metod, pri katerih je napisano le, da je bila za evalvacijo uporabljena zbirka ICDAR 2003. Kljub temu v tabeli 3.1 podajamo lestvico trenutno najuspešnejših metod na zbirki ICDAR 2003, pri čemer jih rangiramo glede na preciznost p (v primeru, ko imata metodi enako vrednost p, ju dodatno rangiramo glede na priklic r). V nadaljevanju vse metode

Tabela 3.1

Evalvacija metod na zbirki ICDAR 2003. Kadar metoda nima splošnega imena, za poimenovanje metode uporabljamo eno izmed njenih karakteristik.

Metoda	*Avtor*	*Vrsta metode*	*p*	*r*	*f*
Grupiranje mejnih točk [56]	Yi in Tian	regijska	0,73	0,67	0,66
F-HOG [40]	Minetto in drugi	hibridna	0,73	0,61	0,67
SWT [33]	Epshtein, Ofek in Wexler	regijska	0,73	0,60	0,66
MSER [37]	Chen in drugi	regijska	0,73	0,60	0,66
MSER++ [55]	Neumann in Matas	regijska	0,72	0,62	0,67
Strukt. part. in grupiranje [34]	Yi in Tian	regijska	0,71	0,62	0,62
CRF [38]	Pan	hibridna	0,67	0,70	0,69
AdaBoost [32]	Lee in drugi	teksturna	0,66	0,75	0,70
SnooperText [39]	Minetto in drugi	hibridna	0,63	0,61	0,61

tudi opisujemo.

3.2 Teksturne metode

Ideja teksturnega pristopa je premikanje preiskovalnega okna vzdolž slike pri različnih skalah piramide in iskanje regij, ki ustrezajo predefinirani teksturi teksta. Zaradi detekcije tekstovnih regij pri različnih skalah piramide mora metoda ustrezno poskrbeti za pravilno integriranje rezultatov posameznih skal. Slabosti teksturnega pristopa so nenatančnost določanja robov teksta, nezmožnost detekcije nehorizontalnih tekstov in visoka časovna kompleksnost. Tipični predstavniki teksturnega pristopa so [17, 30–32].

Ker so v zadnjem času popularnejše regijske metode, ki jih opisujemo v poglavju 3.3, se v podrobnosti teksturnih metod ne spuščamo. V nadaljevanju opisujemo metodo detekcije teksta s klasifikatorjem AdaBoost [32], ki je tipičen (in v tabeli 3.1 tudi edini) predstavnik teksturnih metod, na koncu podpoglavja pa podajamo še krajši pregled ostalih teksturnih metod.

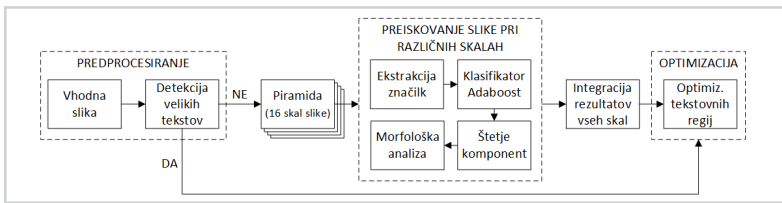
3.2.1 Detekcija teksta s klasifikatorjem AdaBoost

Diagram poteka metode detekcije teksta s klasifikatorjem AdaBoost [32] je prikazan na sliki 3.1. Njen glavni del predstavlja faza preiskovanja slike pri različnih skalah (angl. *multi-scale sequential search*), ki je značilna za teksturne metode. Metoda uporablja 16 različnih skal slike, pri čemer se pri vsaki skali posebej s preiskovalnim oknom, velikosti

64×32 slikovnih elementov, sekvenčno premika vzdolž slike in klasificira bloke pod preiskovalnim oknom. Za klasifikacijo metoda uporablja 6 tipov značilk:

- povprečje in standardna deviacija odvodov v horizontalni in vertikalni smeri,
- lokalna energija Gaborjevega filtra,
- statistična mera teksture (angl. *statistical texture measure*) histograma bloka,
- standardna deviacija valčnih koeficientov diskretne valčne transformacije DWT (angl. *Discrete Wavelet Transform*),
- povprečje in standardna deviacija horizontalnih intervalov med robnimi točkami bloka ter
- geometrijske lastnosti povezanih komponent v bloku.

Bloki, ki so pri posameznih skalah klasificirani kot tekst, se po končani fazi preiskovanja ustrezno integrirajo v osnovno skalo, v fazi optimizacije (angl. *result optimisation*) pa metoda razmetane bloke na podlagi minimalnih in maksimalnih robov oblikuje v pravokotnike. Faza predprocesiranja (angl. *pre-processing*) je namenjena detekciji zelo velikih tekstov v slikah. V tej fazi metoda sliko zmanjša na velikost 64×32 slikovnih elementov in jo klasificira s klasifikatorjem AdaBoost. Če je slika klasificirana kot tekst, metoda fazo preiskovanja slike pri različnih skalah preskoči.



Slika 3.1

Diagram poteka metode detekcije teksta na podlagi klasifikatorja AdaBoost [32].

3.2.2 *Ostale teksturne metode*

Poleg metode detekcije teksta na podlagi klasifikatorja AdaBoost [32] obstajajo še druge teksturne metode. Metoda, ki jo predlagajo Li in drugi [17], temelji na filtriranju slike s posebno strukturno masko (angl. *stroke filter*) v horizontalni, vertikalni ter obeh

diagonalnih smereh. Metoda nato s premikanjem preiskovalnega okna vzdolž slike regije pod oknom klasificira s klasifikatorjem SVM. Razmetane regije, ki so klasificirane kot tekst, se na podlagi geometrijskih pravil grupirajo in oblikujejo v pravokotnike, ki jih avtorji imenujejo tekstovne črte[1]. Vse detektirane tekstovne črte metoda dodatno klasificira s klasifikatorjem SVM. Postopek detekcije se izvaja pri različnih skalah slike.

Chen in Yuille [31] v svoji metodi uporabljata klasifikator AdaBoost, s katerim klasificirata regije pod preiskovalnim oknom pri 14 različnih skalah vhodne slike. Za klasifikacijo uporabljata tri tipe značilk, ki jih generirata iz blokovnih particij (angl. *block patterns*) posameznih regij. Značilke temeljijo na (1) odvodih v horizontalni in vertikalni smeri, (2) histogramih intenzitet, smeri gradientov in intenzitet gradientov ter (3) povezanih robnih točkah znotraj bloka.

Primer teksturne metode je tudi [30]. Metoda temelji na večskalni valčni dekompoziciji (angl. *multiscale wavelet decomposition*). Vsi slikovni elementi z valčno energijo (angl. *wavelet energy*) nad dinamičnim pragom so označeni kot kandidatni slikovni elementi. V fazi razraščanja regij (angl. *region growing*) se okoli kandidatnih slikovnih elementov generirajo regije, velikosti 16×10 slikovnih elementov, ki se iterativno združujejo v tekstovne črte. Z uporabo projekcijskih profilov (angl. *projection profiles*) metoda tekstovne črte razbije na posamezne vrstice in jih dodatno klasificira s klasifikatorjem SVM na podlagi valčnih značilk. Metoda rezultate posameznih skal na koncu ustrezno integrira v končni rezultat.

3.3 Regijske metode

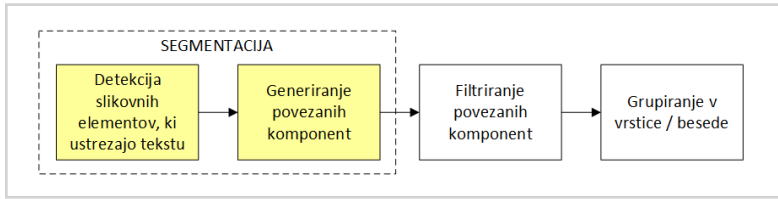
Diagram poteka tipične regijske metode je prikazan na sliki 3.2. Za regijske metode je značilno, da v sliki najprej poiščejo slikovne elemente (ali manjše regije), ki imajo določene značilnosti teksta in jih povežejo v povezane komponente (angl. *connected components*). Te na podlagi vnaprej določenih geometrijskih pravil filtrirajo in izločijo tiste, ki ne ustrezajo geometrijskim lastnostim črk, preostale pa grupirajo v vrstice in besede. Fazo detekcije slikovnih elementov in fazo generiranja povezanih komponent v nadaljevanju pogosto označujemo s skupnim izrazom segmentacija (slika 3.2). Poleg tega, da imajo nižjo časovno kompleksnost kot teksturne metode, regijske metode omogočajo natančno določanje robov črk in detekcijo nehorizontalnih tekstov. Mednje se uvršča večina novejših metod [33–37].

[1] Poudariti je treba, da ne gre za klasične tekstovne črte, ki jih omenjamo skozi doktorsko disertacijo. Avtorji v [17] izraz tekstovne črte uporabljajo za poimenovanje detektiranih blokov teksta.

V nadaljevanju podrobneje opisujemo predstavnike regijskih metod. Še posebej se osredotočamo na metodo SWT [33] in metodo barvne redukcije, prilagojene detekciji teksta [48], saj sta ključni za razumevanje predlagane metode, ki je opisana v poglavju 4. Metoda barvne redukcije, prilagojene detekciji teksta, sicer pokriva le prva dva koraka regijskega pristopa (slika 3.2), vendar jo je z moduloma za filtriranja in grupiranje možno ustrezno razširiti, zato jo v nadaljevanju uvrščamo med regijske metode. Primer njene razširitve je regijska metoda strukturne particije in grupiranja [34], ki dosega ene izmed najboljših rezultatov (tabela 3.1) in jo tudi podrobneje opisujemo.

Slika 3.2

Diagram poteka tipične regijske metode. Z rumeno barvo sta označena koraka, ki ju pokrivata metoda barvne redukcije, prilagojene detekciji teksta, in predlagana metoda, opisana v poglavju 4.

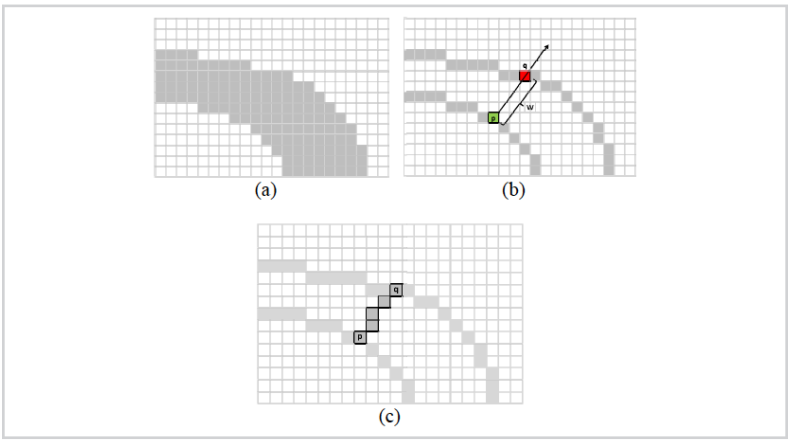


3.3.1 Metoda SWT

Metoda SWT (angl. *Stroke Width Transform*) [33] temelji na predpostavki o konstantni debelini črk (angl. *stroke width constancy assumption*), ki pravi, da se tekst od ostalih struktur v sliki razlikuje po značilni paralelni obliki konstantne debeline. Če pogledamo primer manjšega segmenta črke, ki je prikazan na sliki 3.3a, vidimo, da segment ohranja približno enako (konstantno) debelino vzdolž celotne slike, pri čemer sta njegov zunanji in notranji rob bolj ali manj vzporedna.

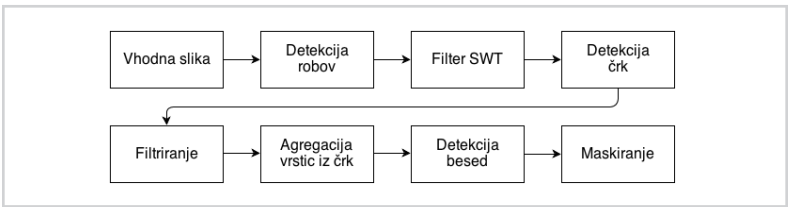
Diagram poteka metode SWT je prikazan na sliki 3.4. Jedro metode predstavlja filter SWT, ki sledi fazi detekcije robov. Filter SWT za vsako robno točko p v sliki v smeri pripadajočega gradienta d_p poišče najbližjo robno točko q z gradientom d_q (slika 3.3b). Če sta gradienta d_p in d_q približno nasprotna[2] ($d_q = -d_p \pm \pi/6$), se vsem slikovnim elementom na poti med p in q (vključno s p in q) priredi vrednost w, pri čemer w označuje evklidsko razdaljo med p in q (slika 3.3c). Prirejanje poteka tako, da se vrednost w shrani na pripadajoča mesta v ločeni sliki SWT I_{SWT}, ki je enake velikosti kot vhodna slika. Vsi slikovni elementi slike I_{SWT} so ob inicializaciji

[2]Če gradienta nista nasprotna, se par robnih točk ne upošteva in filter SWT izvajanje nadaljuje v naslednji robni točki.



Slika 3.3

Delovanje filtra SWT [33] (slika je povzeta po [33]). (a) Segment črke (angl. *stroke*). Metoda SWT predpostavlja, da so robovi črke vzporedni. (b) Iz robnega slikovnega elementa p potujemo v smeri gradienta, dokler ne naletimo na robni slikovni element q s približno nasprotnim gradientom. Razdaljo med p in q označimo z w. (c) Vsem slikovnim elementom na poti med p in q priredimo vrednost w.

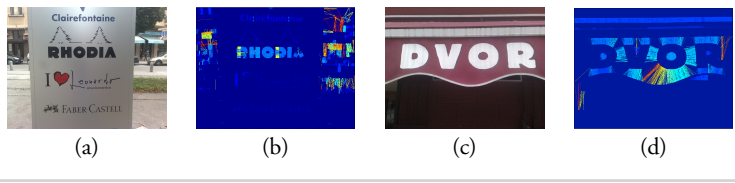


Slika 3.4

Diagram poteka metode SWT [33].

postavljeni na vrednost ∞. Po končanem postopku slika I_{SWT} (vsaj teoretično) vsebuje debeline črk, ki jim slikovni elementi pripadajo. Če ima določen slikovni element slike I_{SWT} vrednost ∞, po vsej verjetnosti ne leži na nobeni črki oziroma paralelni strukturi. Primera slik SWT sta prikazana na sliki 3.5.

V fazi detekcije črk, ki sledi filtru SWT, se slikovni elementi slike I_{SWT} povežejo v povezane komponente. Dva sosednja slikovna elementa pripadata isti povezani komponenti, če imata podobni vrednosti SWT. V fazi filtriranja se izločijo vse prevelike in premajhne povezane komponente ter vse tiste, ki imajo bodisi previsoko varianco vrednosti SWT bodisi previsoko ali prenizko razmerje med višino in širino, bodisi vse tiste, ki vsebujejo več kot dve povezani komponenti. Povezane komponente, ki niso izločene, se naknadno grupirajo v vrstice in besede. Povezani komponenti pripadata isti vrstici, če imata podobno povprečje vrednosti SWT, če sta podobnih dimenzij, če ležita dovolj skupaj in če imata podobno barvo. Na podlagi analize histogramov



(a) (b) (c) (d)

Slika 3.5

Primera slik SWT. Originalni sliki (a, c) in pripadajoči sliki SWT (b, d). Slikovni elementi na sliki SWT, ki so pobarvani z isto barvo, imajo enake vrednosti SWT. V primeru (d) zaradi svetlega teksta na temni podlagi gradienti robnih točk črk kažejo ven iz črk, zato metoda SWT detektira napačne slikovne elemente.

medčrkovnih razdalj posameznih vrstic se vrstice grupirajo v besede.

Kljub intuitivnosti ima metoda SWT kar nekaj pomanjkljivosti. V primeru slabih, nekontrastnih slik metoda zaradi slabo detektiranih robov velikokrat ne najde vseh črk v sliki. Ker nekateri robovi posameznih delov črk niso popolnoma paralelni, se pogosto dogaja tudi, da metoda določene črke detektira le parcialno. Primer parcialne detekcije črke "A" v besedi "RHODIA" je prikazan na sliki 3.5b.

V primeru temnega teksta na svetli podlagi gradienti robnih točk črke kažejo proti vzporednim robnim točkam črke (slika 3.3b), kar v primeru svetlega teksta na temni podlagi ne drži (slika 3.5d). V tem primeru so gradienti robnih točk usmerjeni navzven (ven iz črk), kar povzroča iskanje v napačni smeri in posledično nepravilno delovanje metode. Avtorji problem sicer rešujejo z dvakratnim poganjanjem metode v gradientni in protigradientni smeri, vendar je takšen pristop neprimeren, saj dopušča detekcijo medčrkovnih regij. Problem nepravilne smeri iskanja je velika pomanjkljivost metode SWT, zato se mu v poglavju 4.4.3 zelo podrobno posvečamo in predlagamo metodo detekcije smeri SWT.

3.3.2 Metoda TOCR

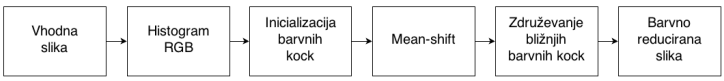
Metoda barvne redukcije, prilagojene detekciji teksta (v nadaljevanju TOCR[3]), ki sta jo predlagala Nikolaou in Papamarkos [48], reducira število barv v sliki na manjše število

[3]Kratica TOCR se za poimenovanje metode barvne redukcije, prilagojene detekciji teksta, v literaturi ne uporablja. V doktorski disertaciji jo uporabljamo izključno zaradi jasnosti razlage. Kratico TOCR tvorimo iz prvih črk angleškega poimenovanja *Text Oriented Color Reduction*.

osnovnih barv slike in s tem poenostavi postopek segmentacije[4]. Metodo pri detekciji teksta uporabljata tudi Yi in Tian [34] in z njo dosegata zelo dobre rezultate na zbirki ICDAR 2003 (tabela 3.1).

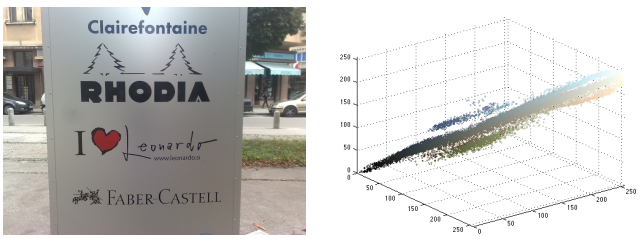
Diagram poteka metode TOCR je prikazan na sliki 3.6. Metoda na podlagi vhodne slike najprej generira histogram RGB (primer histograma je prikazan na sliki 3.7b). V fazi inicializacije barvnih kock metoda izbere naključno celico histograma in okoli nje generira začetno kocko z dolžino stranice h. Vse celice histograma, ki jih kocka pokriva, metoda označi kot obiskane in postopek generiranja začetnih kock ponavlja, dokler niso obiskane vse celice histograma. Primer začetnih barvnih kock s pripadajočimi barvami in začetno barvno reducirano sliko je prikazan na slikah 3.8a, 3.8b in 3.8c.

V fazi *mean-shift* metoda začetne barvne kocke iterativno pomika proti gravitacijskim centrom histograma RGB, tj. proti dominantnim barvam v sliki. Kocke, ki po fazi *mean-shift* ležijo dovolj skupaj, metoda naknadno združi. Središča barvnih kock po končanem združevanju ustrezajo končnim barvam, na podlagi katerih se generira končna barvno reducirana slika. Slika 3.8d prikazuje primer končnih barvnih kock s pripadajočimi barvami (slika 3.8e) in končno barvno reducirano sliko (slika 3.8f).



Slika 3.6

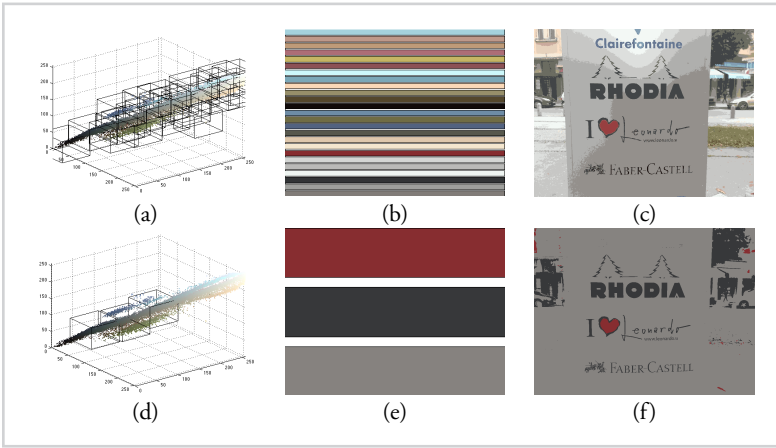
Diagram poteka metode TOCR [48].



Slika 3.7

Primer histograma RGB. Originalna slika (a) in njen histogram RGB (b).

[4]Dva sosednja slikovna elementa pripadata isti povezani komponenti, kadar imata v barvno reducirani sliki enako barvo.



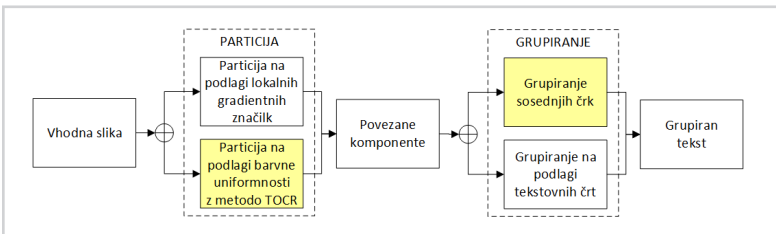
Slika 3.8

Postopek barvne redukcije slike 3.7a. (a) Začetne barvne kocke. (b) Barve, ki ustrezajo kockam v (a). (c) Začetna barvno reducirana slika na podlagi začetnih barv v (a). (d) Končne barvne kocke. (e) Barve, ki ustrezajo kockam v (d). (f) Končna barvno reducirana slika na podlagi končnih barv v (d).

Žal se pri barvni redukciji dogaja, da nedominantne barve teksta, ki v barvnem prostoru RGB ležijo preblizu barvam ozadja in ki ustrezajo manjšim tekstom v sliki, v fazi *mean-shift* gravitirajo proti barvam ozadja. Takšen primer prikazuje slika 3.8f. Modra barva besede "Clairefontaine" se v fazi barvne redukcije izgubi, zato končna barvno reducirana slika te besede ne vsebuje.

3.3.3 *Metoda strukturne particije in grupiranja*

Metoda strukturne particije in grupiranja (angl. *structure-based partition and grouping*), ki jo predlagata Yi in Tian [34], temelji na barvni redukciji slike z metodo TOCR. Diagram poteka metode je prikazan na sliki 3.9. Metoda v fazi particije določi slikovne elemente, ki pripadajo tekstu, in jih poveže v povezane komponente. V fazi grupiranja povezane komponente povezuje v nize črk (angl. *text strings*).



Slika 3.9

Diagram poteka metode strukturne particije in grupiranja [34]. Na mestih, kjer se pojavljajo razcepi, metoda izvaja enega izmed obeh možnih korakov. Z rumeno barvo je označena kombinacija, pri kateri metoda dosega najboljše rezultate.

Particija se lahko izvaja na dva načina: na podlagi lokalnih gradientnih značilk ali na podlagi barvne uniformnosti črk (slika 3.9). Pri particiji na podlagi lokalnih gradientnih značilk se podobno kot pri filtru SWT [33] poiščejo vsi pari robnih točk s približno nasprotnimi gradienti. Za vsak najdeni par robnih točk se določi eksponentna porazdelitev magnitud slikovnih elementov na poti med njima:

$$g(G_{mag}; \lambda) = \lambda \exp(-\lambda G_{mag}), \tag{3.1}$$

pri čemer G_{mag} označuje vektor magnitud slikovnih elementov na poti, λ pa stopnjo razkroja (angl. *decay rate*), ki je enaka:

$$\lambda = \frac{1}{\sum G_{mag}}. \tag{3.2}$$

Višja stopnja razkroja je indikator, da pot poteka preko slikovnih elementov z nizkimi magnitudami, kar je v skladu s predpostavko o uniformnosti intenzitet v notranjosti črke. Zato se slikovni elementi z dovolj visoko stopnjo razkroja označijo kot tekstovni. Rezultat particije je seznam povezanih komponent, dobljenih iz binarne slike tekstovnih in netekstovnih slikovnih elementov.

Particija na podlagi barvne uniformnosti temelji na metodi TOCR, ki je opisana v poglavju 3.3.2. Rezultat particije so povezane komponente, pri čemer posamezni povezani komponenti ustrezata sosednja slikovna elementa, ki imata v barvno reducirani sliki isto barvo.

Avtor predlaga dva načina grupiranja: grupiranje sosednjih črk in grupiranje na podlagi tekstovnih črt. Pri grupiranju sosednjih črk se za vsako povezano komponento C določi množica sosedov. Povezana komponenta C' je sosed komponente C, če imata komponenti približno enaki višini, če horizontalna in vertikalna razdalja med njima ni prevelika, če imata približno enako površino in če sta podobnih barv. Sosednje komponente, ki ležijo na levi strani C, tvorijo množico levih sosedov F_L, medtem ko preostale tvorijo množico desnih sosedov F_R. Vsaka povezana komponenta, ki ima neprazni množici F_L in F_R, ki se v številu elementov razlikujeta za največ 3, skupaj s F_L in F_R tvori grupo sosedov. V iterativnem postopku se vse grupe sosedov medsebojno združujejo, pri čemer se dve grupi lahko združita, kadar vsebujeta vsaj dve isti povezani komponenti. Po končanem postopku so vse grupe s standardnimi deviacijami ploščin, standardnimi deviacijami medčrkovnih razdalj in standardnimi deviacijami debelin črk nad določenim pragom naknadno izločene.

Grupiranje na podlagi tekstovnih črt je namenjeno detekciji nehorizontalnih tekstov. Iz množice povezanih komponent metoda izbere tri naključne komponente m_i, m_j in m_k s pripadajočimi daljicami $m_i m_j$ in $m_j m_k$ ter izračuna razliko dolžin daljic Δd in razliko njunih naklonov $\Delta\Theta$. Točke m_i, m_j in m_k so kolinearne, kadar sta Δd in $\Delta\Theta$ pod določenim pragom. Za vsako detektirano trojico kolinearnih točk se s Houghovo transformacijo poiščejo vse povezane komponente v sliki, ki ležijo na pripadajoči premici. Rezultat postopka je množica tekstovnih črt, ki ustrezajo horizontalnim in nehorizontalnim tekstom v sliki. Podobno kot pri grupiranju sosednjih črk se tudi pri grupiranju na podlagi tekstovnih črt dodatno preverijo standardne deviacije ploščin, medčrkovnih razdalj in debelin črk v grupi.

Glede na vrsto particije in vrsto grupiranja so možne štiri kombinacije izvajanja metode: (a) gradientna particija z grupiranjem sosednjih črk (GS), (b) barvna particija z grupiranjem sosednjih črk (BS), (c) gradientna particija z grupiranjem na podlagi tekstovnih črt (GT) ter (d) barvna particija z grupiranjem na podlagi tekstovnih črt (BT). Avtorji metode so vse štiri kombinacije preizkusili na zbirki ICDAR 2003 [44] in najboljše rezultate dosegli pri barvni particiji z grupiranjem sosednjih črk (BS), kar kaže na učinkovitost barvne redukcije pri detekciji teksta. Rezultati najboljše kombinacije (BS) na zbirki ICDAR 2003 so prikazani v tabeli 3.1.

3.3.4 Ostale regijske metode

Poleg metode SWT in metode barvne redukcije obstajajo številne druge regijske metode, ki se tipično razlikujejo predvsem po načinu izbire tekstovnih slikovnih elementov. Metoda, ki jo predlagata Neumann in Matas [36, 55], temelji na detekciji maksimalno stabilnih ekstremnih regij (angl. *maximally stable extremal regions*) MSER oziroma izpeljank MSER++ [55] in ER (angl. *extremal regions*) [36]. Z uporabo izčrpnega preiskovanja (angl. *exhaustive search*) vseh zaporedij detektiranih znakov Neumann in Matas detektirane regije grupirata v tekstovne črte (angl. *text lines*). Zaradi učinkovitega rezanja neobetavnih zaporedij znakov v preiskovalnem drevesu lahko metoda izvaja izčrpno iskanje v realnem času.

Na detekciji MSER temelji tudi metoda, ki jo predlagajo Chen in drugi [37]. Zaradi občutljivosti detektorja MSER na zamegljenost (angl. *blur*) slik se pogosto dogaja, da se posamezna regija MSER razteza preko večjega števila črk v sliki. Da bi detektirane regije MSER ustrezale le posameznim črkam v sliki, Chen detektor MSER nadgrajuje s Cannyjevim detektorjem robov [58]. Metoda detektirane robne točke v sliki poveže

v konture in iz posameznih regij MSER odstrani vse tiste slikovne elemente, ki ležijo na zunanjih straneh pripadajočih kontur. Končne regije MSER so dodatno filtrirane na podlagi geometrijskih pravil in distančne transformacije (angl. *distance transform*), ki deluje podobno kot filter SWT [33].

Metoda grupiranja mejnih točk (angl. *boundary clustering*) [56] temelji na predpostavki, da so tako črke kot podlage[5] barvno uniformne regije, ki jih medsebojno razmejujejo robovi črk oziroma t. i. mejne točke. Glede na barve regij, ki jih posamezne mejne točke razmejujejo, metoda z uporabo Gaussovih modelov (angl. *Gaussian mixture models*) vse mejne točke v sliki grupira v gruče. Centri gruč predstavljajo dominantne barve teksta v sliki vključno z dominantnimi barvami pripadajočih podlag. V fazi segmentacije metoda slikovnim elementom, ki ležijo v okolici mejnih točk, priredi barvo barvno najbližje gruče in jih poveže v povezane komponente. Grupiranje povezanih komponent v nize poteka na podoben način kot v [34] (glej poglavje 3.3.3). Na koncu metoda z uporabo klasifikatorja SVM (angl. *Support Vector Machine*) in Gaborjevih značilk dodatno preveri, ali detektirani nizi resnično ustrezajo tekstu.

3.4 Hibridne metode

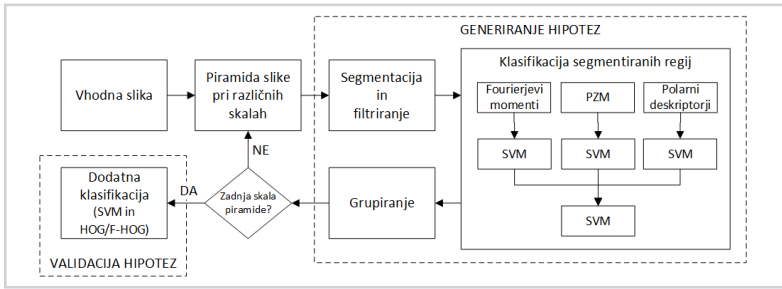
Hibridne metode pri detekciji teksta združujejo regijski in teksturni pristop, pri čemer izkoriščajo prednosti obeh. V osnovi delujejo podobno kot regijske metode, vendar zaradi prisotnosti poljubnih velikosti teksta v slikah regijski pristop izvajajo pri različnih skalah piramide. V nadaljevanju podrobneje opisujemo dve hibridni metodi, ki na zbirki ICDAR 2003 dosegata zelo dobre rezultate.

3.4.1 SnooperText

Metoda detekcije teksta, ki jo predlagajo Minetto in drugi [39, 40], je tipičen predstavnik hibridnih metod. Diagram poteka metode je prikazan na sliki 3.10. Metoda pri vsaki skali piramide posebej z uporabo morfološkega operatorja preklopa (angl. *toggle mapping*) [59] segmentira kandidatne regije in jih klasificira s hierarhičnim klasifikatorjem SVM. Za klasifikacijo se uporabljajo trije tipi značilk: Fourierjevi momenti, pseudo Zernikovi momenti (PZM) in polarni deskriptorji. Regije, ki so klasificirane kot tekst, se geometrijsko filtrirajo in integrirajo v osnovno skalo piramide. Zaradi dodatne robustnosti se detektirane regije po končanem postopku dodatno klasificirajo

[5] Z izrazom "podlaga" označujemo površino, na kateri je tekst. Primer podlage je reklamna tabla.

s klasifikatorjem SVM in z značilkami HOG (angl. *Histogram of Oriented Gradients*) [39] oz. F-HOG (angl. *Fuzzy HOG*) [40]. Namen dodatne klasifikacije je odstranitev šumnatih regij s periodičnimi teksturami (ograje, opeka ipd.).



Slika 3.10

Diagram poteka hibridne metode, ki jo predlagajo Minetto in drugi [39, 40]. Razlika med [39] in [40] je v fazi validacije hipotez. Validacija v [39] temelji na deskriptorjih HOG, medtem ko validacija v [40] temelji na deskriptorjih F-HOG.

3.4.2 *Metoda CRF*

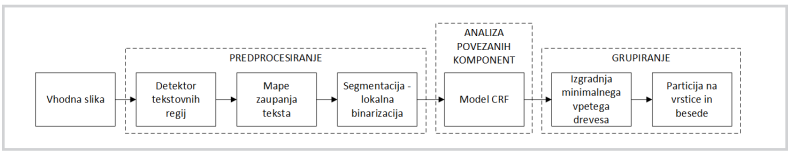
Hibridna metoda CRF (angl. *Conditional Random Field*) [38] problem filtriranja in grupiranja teksta rešuje na drugačen način kot ostale opisane metode. Tipično se filtriranje in grupiranje povezanih komponent izvajata na podlagi vnaprej določenih geometrijskih pravil, medtem ko metoda CRF filtriranje in grupiranje izvaja na podlagi učno usmerjene minimizacije energijske funkcije. Podoben način grupiranja uporabljamo tudi v [60].

Diagram poteka metode CRF je prikazan na sliki 3.11. Metoda v fazi predprocesiranja pri različnih skalah piramide klasificira regije pod preiskovalnim oknom, velikosti 16×16 slikovnih elementov. Klasifikacija temelji na klasifikatorju WaldBoost in značilkah HOG. Verjetnost, da določena regija pod preiskovalnim oknom vsebuje tekst, se skupaj z informacijo o skali uporabi pri izgradnji t. i. mape zaupanja teksta (angl. *text confidence map*). Za binarizacijo slike se uporablja Niblackova lokalna binarizacijska metoda [61] z variabilnim jedrom, pri čemer se velikost jedra določa dinamično na podlagi vrednosti mape zaupanja.

V fazi analize povezanih komponent se za filtriranje uporablja model CRF, ki na podlagi unarnih in binarnih značilk povezanih komponent izloči netekstovne regije (parametri modela CRF se določijo v fazi iterativnega učenja, pri čemer se ob vsaki iteraciji z rezanjem grafa minimizira energijska funkcija). Unarne značilke označujejo lastnosti posameznih poveznih komponent (velikost, razmerje med višino in širino,

kompaktnost ipd.), medtem ko binarne značilke zajemajo relacije med dvema povezanima komponentama. Mednje sodijo razlika v velikosti, medsebojna razdalja, stopnja prekrivanja ipd.

V fazi grupiranja metoda preostale povezane komponente poveže v polni graf[6], na katerem s Kruskalovim algoritmom poišče minimalno vpeto drevo. Z minimizacijo energijske funkcije metoda poišče optimalno razbitje drevesa na vrstice. Na podoben način se posamezne vrstice dodatno razbijejo na besede.



Slika 3.11

Diagram poteka metode CRF [38].

3.4.3 Ostale hibridne metode

Med hibridne metode se uvrščajo tudi metode [41–43]. Metoda [41], ki jo predlagajo Jung, Liu in Kim, robne točke slike filtrira s podobnim filtrom kot metoda [17].[7] Slikovni elementi z dovolj visoko odzivnostjo filtra se morfološko filtrirajo in povežejo v povezane komponente. Regije, ki obdajajo posamezne povezane komponente, se normalizirajo na višino 15 slikovnih elementov in ustrezno dolžino, ki ohranja razmerje velikosti originalne regije. Metoda vsako izmed normaliziranih regij analizira s preiskovalnim oknom, velikosti 15×15 slikovnih elementov, pri čemer za klasifikacijo uporablja klasifikator SVM. Število premikov preiskovalnega okna znotraj normalizirane regije je odvisno od njene dolžine. Rezultati klasifikacije v posameznih korakih se povprečijo in če povprečni klasifikacijski rezultat normalizirane regije dosega vrednost nad določenim pragom, je regija klasificirana kot tekst. Metoda detektirane regije na koncu še dodatno skrči in na ta način izloči netekstovne dele na njihovi skrajno levi in skrajno desni strani. Kljub temu, da metoda ne preiskuje vhodne slike pri različnih skalah, se zaradi normalizacije detektiranih regij (ki ima podoben učinek kot preiskovanje različnih skal slike) uvršča med hibridne metode.

[6] Cena povezave med dvema povezanima komponentama je enaka linearni kombinaciji njunih značilk.

[7] Na tem mestu je treba pojasniti, da kljub temu, da metodi [17] in [41] uporabljata podoben filter, ne spadata v isto kategorijo metod. Metoda [17] filtrirano sliko namreč preiskuje s preiskovalnim oknom, kar jo uvršča med teksturne metode, medtem ko metoda [41] filtrirane slikovne elemente povezuje v povezane komponente, ki jih nato dodatno analizira pri normaliziranih velikostih, kar jo uvršča med hibridne metode.

Podoben koncept kot metoda [41] uporablja tudi metoda [43]. Metoda v fazi lokalizacije teksta (ki sovpada z regijskim vidikom hibridnega pristopa) v sliki detektira horizontalne in vertikalne robne točke ter jih s kombinacijo morfoloških operatorjev grupira v regije (horizontalne posebej in vertikalne posebej). Izmed vseh regij metoda izbere le tiste, ki vsebujejo tako horizontalne kot vertikalne robove. Regije lahko vsebujejo večje dele teksta, zato jih metoda z iterativnim postopkom razmejevanja (angl. *splitting*) razbije na manjše regije, pri čemer razmejitvene črte določi na podlagi projekcijskih profilov (angl. *projection profiles*). V fazi verifikacije teksta se posamezne regije klasificirajo s klasifikatorjem SVM. Pri tem je vsaka regija predhodno normalizirana na višino 16 slikovnih elementov in analizirana s preiskovalnim oknom, velikosti 16×16 slikovnih elementov (kar sovpada s teksturnim vidikom hibridnega pristopa). Metoda rezultate klasifikacije posameznih korakov (preiskovalno okno se premika preko posamezne regije v več korakih) združi v mero zaupanja, ki je definirana kot obtežena vsota klasifikacijskih rezultatov posameznih korakov. Za obtežitev se uporablja funkcija Gaussove porazdelitve glede na odmik preiskovalnega okna od centra regije.

3.5 Diskusija

Metode detekcije teksta se v splošnem delijo na regijske in teksturne. Regijske metode delujejo na nivoju slikovnih elementov, ki jih povezujejo v regije. So hitre, robove črk detektirajo natančno in so primerne za detekcijo nehorizontalnih tekstov, vendar so zaradi narave lokalnih značilk občutljive na šum. Večina novejših metod se uvršča med regijske metode. Teksturne metode so zaradi piramidnega pristopa časovno kompleksnejše. Pri različnih skalah piramide preiskujejo sliko in klasificirajo regije pod preiskovalnim oknom. Teksta ne detektirajo natančno, prav tako niso primerne za detekcijo nehorizontalnih tekstov. Poleg regijskih in teksturnih obstajajo tudi hibridne metode, ki izkoriščajo prednosti obeh pristopov. Tipično so kompleksnejše za implementacijo.

Grupiranje teksta v vrstice in besede se najpogosteje izvaja na dva načina: s povezovanjem povezanih komponent na podlagi geometrijskih pravil in z iskanjem particije, ki minimizira energijsko funkcijo grafa. Slednji način je sicer bolj robusten, vendar je močno odvisen od izbire pravih učnih primerov in protiprimerov v fazi učenja.

V doktorski disertaciji se osredotočamo na regijski metodi SWT in TOCR, saj na njima temelji predlagana metoda. Metoda SWT je zasnovana na predpostavki o paralelnosti robov črk, medtem ko je ideja metode TOCR redukcija barv v sliki na manjše

število dominantnih barv. Metoda SWT je robustna, vendar pogosto spušča dele črk. Metoda TOCR črke detektira v celoti, vendar ima probleme z detekcijo nedominantnih barv teksta.

4

Predlagana metoda

4.1 Uvod

V tem poglavju opisujemo metodi barvne redukcije na podlagi glasovanja SWT (angl. *SWT voting-based color reduction*) [49] in detekcije smeri SWT [49]. Obe metodi predstavljata izvirni znanstveni prispevek. Metoda barvne redukcije na podlagi glasovanja SWT je detekciji teksta prilagojena segmentacijska metoda, ki z integracijo strukturne informacije SWT usmerja postopek barvne redukcije. V primerjavi z metodo TOCR [48] dosega boljše rezultate. Barve, ki so bogate s slikovnimi elementi SWT, po vsej verjetnosti pripadajo tekstu, zato jih metoda blokira in jim ne dovoli konvergiranja proti barvam ozadja. Zaradi integracije tako strukturne kot barvne informacije se razlikuje od večine ostalih metod detekcije teksta, ki v fazi segmentacije tipično upoštevajo bodisi strukturno [32, 33, 36–40, 55] bodisi barvno informacijo teksta [34, 48]. Metoda je primerna za uporabo v prvih dveh korakih regijskega pristopa (slika 3.2).

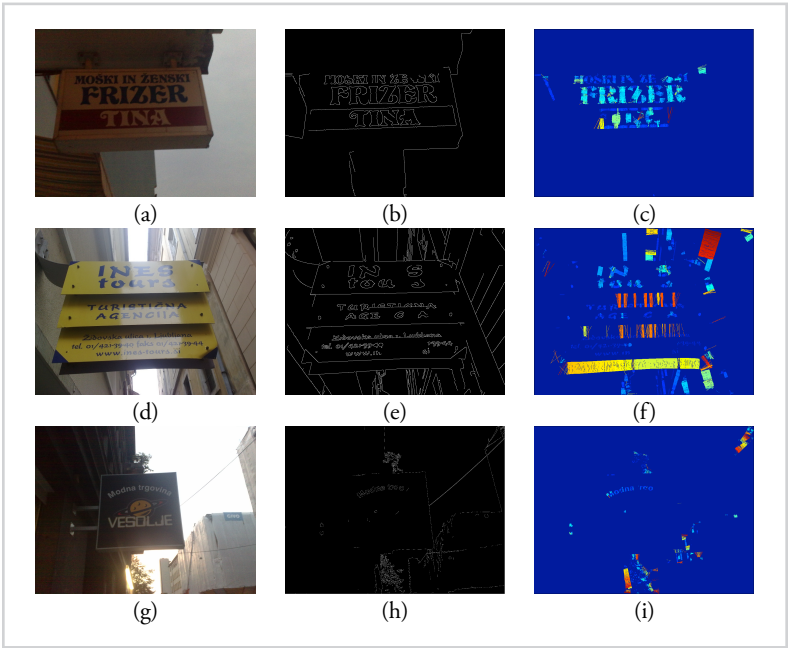
Vzrokov za izbiro metod SWT [33] in TOCR [48] pri zasnovi lastne metode je več. Prvič, metodi se uvrščata med regijske[1], ki so zaradi hitrosti primernejše za uporabo v realnem času (npr. pri integraciji detekcije teksta v mobilno aplikacijo, ki je naš dolgoročni cilj). Prav tako so regijske metode zaradi natančnosti detekcije robov črk primernejše za neposredno povezovanje s klasifikatorji OCR. Drugič, izmed regijskih metod tako metoda SWT kot metoda na podlagi segmentacije TOCR [34] trenutno dosegata ene najboljših rezultatov (tabela 3.1). Še več, metoda [34] dokazuje, da uporaba barvne informacije izboljša detekcijo teksta v primerjavi z detekcijo na podlagi strukturne informacije.

Metoda barvne redukcije na podlagi glasovanja SWT za pravilno delovanje potrebuje točne podatke SWT. Če zaradi napačne smeri iskanja metoda SWT detektira netekstovne regije, barvna redukcija ne favorizira barv teksta. Metoda detekcije smeri SWT, ki jo predlagamo, problem pravilne smeri iskanja rešuje z razbitjem slike na bloke in primerjavo histogramov blokov gradientnih ter protigradientnih slik SWT.

4.2 Slabosti metode SWT

Metoda SWT [33] je močno odvisna od uspešnosti detekcije robov. V primeru šumnatih slik z nekontrastnimi teksti so robovi črk pogosto slabo detektirani, kar povzroča nepravilno delovanje metode SWT. Slika 4.1 prikazuje tri primere problematičnih slik

[1] Metoda TOCR sicer ni regijska, jo je pa kljub temu možno uporabiti v fazi segmentacije regijskega pristopa.



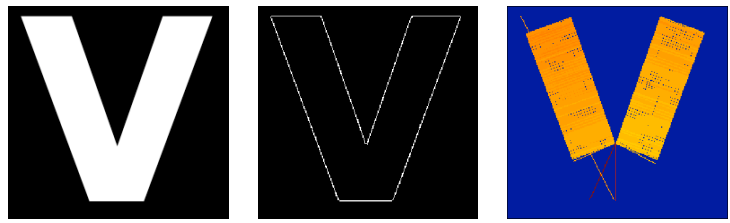
Slika 4.1

Primeri delovanja filtra SWT na svetlobno problematičnih slikah zbirke CVL OCR DB. Detekcija robov je izvedena s Cannyjevim detektorjem robov (spodnji prag: 0,1, zgornji prag: 0,2). Levi stolpec prikazuje originalne slike, zajete pri slabi/problematični osvetlitvi, srednji stolpec pripadajoče slike detektiranih robov, desni stolpec pa pripadajoče slike SWT.

zbirke CVL OCR DB s pripadajočimi slikami robov in slikami SWT. Robovi so detektirani z uporabo Cannyjevega detektorja robov [58] (spodnji in zgornji prag sta enaka 0,1 in 0,2). V primeru (a) so robovi črke "M" v besedi "MOŠKI" in črk "N", "S", "K" v besedi "ŽENSKI" detektirani nepopolno, zato metoda SWT črk ne segmentira pravilno. V primeru (b) zaradi centralnega odbleska na sliki detektor robov spušča celotne črke: "E" v besedi "INES", "r" v besedi "Tours" itd. Podobno situacijo prikazuje primer (c), kjer so robovi črk v besedi "trgovina" detektirani nepopolno, beseda "VESOLJE" pa je v celoti izpuščena.

Dodatna pomanjkljivost metode SWT je predpostavka o absolutni paralelnosti robov črk, saj v praksi niso vedno vzporedni. Tipičen primer je črka "V", ki je prikazana na sliki 4.2. Zaradi neparalelnosti spodnjega dela metoda SWT črko detektira le parcialno.

Kljub vsemu je metoda SWT precej robustna. Tudi kadar detektor robov ne detektira vseh robov pravilno ali ko robovi črke niso popolnoma paralelni, metoda SWT še



(a) (b) (c)

Slika 4.2

Problem neparalelnosti robov črke "V". (a) Slika črke "V". (b) Slika robov. (c) Slika SWT. Spodnji del črke "V" je zaradi neparalelnosti robov nedetektiran.

vedno uspe detektirati posamezne dele črk oziroma besed. Ker v metodi barvne redukcije na podlagi glasovanja SWT informacije SWT ne uporabljamo absolutno, temveč nanjo gledamo kot verjetnost, da se v nekem delu slike pojavlja tekst, je metoda SWT za naše potrebe več kot primerna.

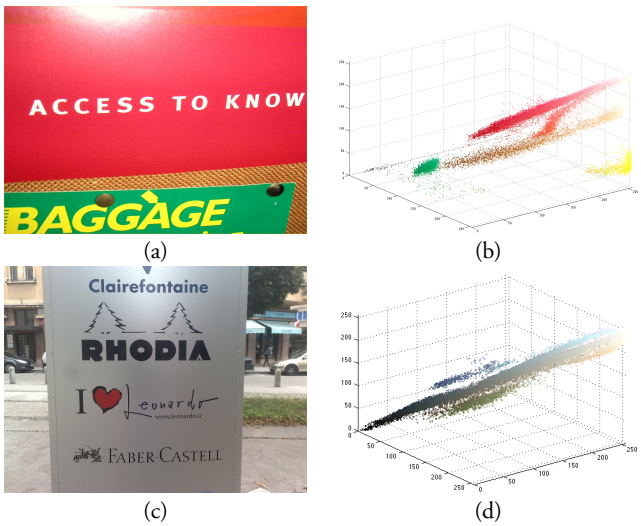
4.3 Slabosti barvne redukcije

Če so barve slike v barvnem prostoru skoncentrirane okoli dominantnih centrov, ki ležijo dovolj narazen, barvna redukcija uspešno reducira barve v sliki. Takšen primer prikazujeta sliki 4.3a in 4.3b. V splošnem slike naravnih scen vsebujejo veliko večje število barv, ki so razpršene po celotnem barvnem prostoru, kot je prikazano na slikah 4.3c in 4.3d. V takšnih primerih se pogosto dogaja, da nedominantne barve teksta konvergirajo proti bolj dominantnim barvam ozadja, zato jih barvna redukcija izpušča. Primer takšne anomalije je prikazan na sliki 4.4. Barvna redukcije rdečo črko "S" na sliki izpusti, saj rdeča barva leži preblizu temno sive barve in konvergira proti njej.

4.4 Barvna redukcija na podlagi glasovanja SWT

Pri zasnovi predlagane metode izhajamo iz predpostavke o paralelnosti robov in barvni uniformnosti teksta. Tekst se namreč od ostalih struktur v sliki razlikuje po specifični paralelni obliki robov in enotni barvi posameznih črk. Tudi kadar imamo opravka s pisanimi oziroma večbarvnimi teksti (slika 1.2b), so posamezne črke tipično še vedno uniformne barve. Na podlagi predpostavke se pojavlja korelacija z metodo SWT[2] [33]

[2]V poglavju 4.2 smo sicer pokazali, da predpostavka o paralelnosti robov teksta ne drži vedno, vendar na tem mestu zaradi lažjega razumevanja metode glasovanja SWT predpostavljamo, da so robovi teksta paralelni.



Slika 4.3

Primera slik teksta v naravnih scenah in pripadajoča barvna histograma.

Slika 4.4

Izpuščanje barv pri barvni redukciji. Originalna slika (a) in barvno reducirana slika (b). Barvna redukcija izpusti rdečo črko "S" v besedi "Segafredo".

in z metodo TOCR [48], ki ju želimo smiselno povezati. Da bi izkoristili prednosti in se v čim večji meri izognili njunim slabostim, smo obe metodi natančno analizirali in identificirali njune prednosti ter slabosti, kar prikazuje tabela 4.1. Metoda SWT sicer pogosto spušča dele črk in jih ne detektira v celoti (problem parcialne detekcije črk), vendar je robustna v smislu, da teksta tipično ne zgreši in ga detektira vsaj delno. Po drugi strani metoda TOCR z razdrobljenostjo teksta nima težav, saj so segmentirane črke večinoma kompaktne, dogaja pa se, da spušča manjše tekste nedominantnih barv.

Tabela 4.1

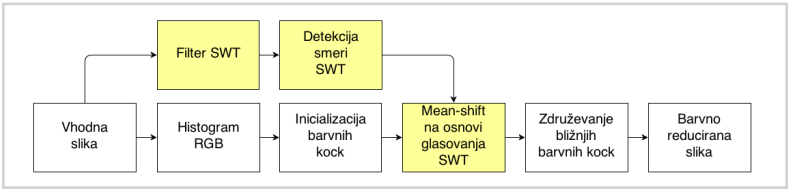
Prednosti in slabosti metod SWT in TOCR.

Metoda	+	-
SWT [33]	robustnost	razdrobljenost črk
TOCR [48]	kompaktnost detektiranih črk	problem nedominantnih barv teksta

Metoda TOCR predstavlja zelo dobro osnovo za segmentacijo teksta, vendar je kljub temu še vedno premalo tekstovno usmerjena, saj deluje na celotni sliki in ne upošteva lokalnih značilnosti teksta. Zato v predlagani metodi, tj. metodi barvne redukcije na podlagi glasovanja SWT (angl. *SWT voting-based color reduction*) [49], koncept metode TOCR nadgrajujemo s t. i. glasovanjem SWT. Diagram poteka metode prikazuje slika 4.5, pri čemer so faze nadgradnje označene z rumeno barvo. V fazi glasovanja SWT na podlagi števila pripadajočih slikovnih elementov SWT ugotovimo, ali določena barva pripada tekstu ali ne. Če barva pripada tekstu, jo blokiramo in ji ne dovolimo, da konvergira proti netekstovnim barvam. Ker slikovni elementi SWT glasujejo v korist barvi, ki ji pripadajo, postopek imenujemo glasovanje SWT.

Kot smo omenili v poglavju 3.3.1, je pomanjkljivost metode SWT problem določanja pravilne smeri iskanja. Metoda SWT sicer problem rešuje z dvakratnim izvajanjem celotnega postopka (v gradientni in protigradientni smeri), vendar za naše potrebe takšna rešitev ne pride v poštev, saj faza glasovanja SWT potrebuje pravilno detektirane slikovne elemente SWT. Problem pravilne smeri iskanja rešujemo z metodo detekcije smeri SWT, ki jo podrobneje opisujemo v poglavju 4.4.3.

Kljub temu, da metoda SWT neparalelnih robov črk ne detektira pravilno, informacija SWT še vedno služi kot dovolj dobra aproksimacija, kje v sliki je tekst.



Slika 4.5

Diagram poteka metode barvne redukcije na podlagi glasovanja SWT [49]. Z rumeno barvo so obarvane faze, ki nadgrajujejo koncept metode TOCR [48].

Faze, ki na diagramu poteka na sliki 4.5 niso označene z rumeno barvo, so v splošnem enake kot pri metodi TOCR (slika 3.6). Razlikujejo se le v dveh točkah. Prvič, za razliko od metode TOCR, ki centre barvnih kock v fazi inicializacije izbira naključno (poglavje 3.3.2), predlagana metoda neobiskane celice barvnega histograma sortira v padajočem vrstnem redu glede na število elementov, ki jih vsebujejo, in centre barvnih kock jemlje z vrha seznama. Na ta način metoda daje prednost bolj polnim (in s tem bolj reprezentativnim) celicam histograma, hkrati pa zagotavlja deterministično delovanje, tj. generiranje istih rezultatov na enakih vhodnih podatkih. Drugič, v fazi generiranja barvno reducirane slike namesto barvnega modela RGB uporabljamo model HSL, saj smo empirično ugotovili, da deluje bolje od modela RGB. Pri tem za določanje, kateri barvni kocki je določena barva v sliki najbližja, uporabljamo metriko razdalj HSL, ki je opisana v [62].

4.4.1 Vpogledna tabela SWT

Za pravilno delovanje koncepta blokiranja barv teksta je treba vedeti, katere barve so barve teksta. Odgovor na vprašanje ponuja slika SWT. Slikovni elementi SWT večinoma ustrezajo tekstu, zato lahko za barvo, ki je prekrita z dovolj slikovnimi elementi SWT, trdimo, da pripada tekstu (slika 4.6). Naslednje vprašanje, ki se postavlja, je, kdaj je določena barva prekrita s slikovnimi elementi SWT oziroma, bolj splošno, kakšna je preslikava iz prostora SWT v barvni prostor. Kot rešitev predlagamo t. i. vpogledno tabelo SWT (angl. *SWT lookup table*).

Vpogledna tabela SWT je tridimenzionalna tabela iste velikosti kot barvni histogram RGB. V primeru 24-bitne implementacije barvnega prostora RGB, pri katerem je vsaka barvna komponenta predstavljena z osmimi biti, ima vpogledna tabela SWT velikost $256 \times 256 \times 256$. Vsaka celica tabele ustreza določeni barvi $C = (R_C, G_C, B_C)$ in vsebuje seznam vrednosti SWT vseh slikovnih elementov z barvo C v originalni sliki. Kot primer vzemimo celico $(120, 50, 70)$. Vsebino celice dobimo tako, da v originalni

(a) (b)

Slika 4.6

Primer slike SWT. (a) Originalna slika. (b) Pripadajoča slika SWT. Neničelni slikovni elementi SWT večinoma ustrezajo tekstu.

sliki poiščemo vse slikovne elemente te barve in vanjo shranimo pripadajoče vrednosti SWT.

Vpogledna tabela SWT predstavlja učinkovit način preslikave prostora SWT v barvni prostor RGB, saj lahko za vsako barvo natančno vidimo, koliko slikovnih elementov SWT ji pripada in kakšne so njihove vrednosti. Če ima določena barva v vpogledni tabeli SWT zelo malo slikovnih elementov SWT, lahko z določeno verjetnostjo trdimo, da barva ne pripada tekstu. Po drugi strani lahko za barvo, ki ima v vpogledni tabeli SWT veliko slikovnih elementov s podobnimi vrednostmi SWT, trdimo, da pripada tekstu.

4.4.2 Glasovanje SWT

Metoda TOCR [48] z zaporedjem premikov *mean-shift* vsako izmed začetnih barvnih kock postopoma premika do njene končne pozicije v barvnem prostoru RGB. Da bi preprečili konvergiranje manj dominantnih barv teksta proti bolj dominantnim barvam ozadja, pred vsakim posameznim premikom *mean-shift* preverimo, ali gre za tranzicijo barve teksta proti barvi ozadja, in če gre, *mean-shift* obstoječe barve ustavimo.

Označimo s CB_1 in CB_2 barvni kocki pred in po posameznem premiku *mean-shift*, kot prikazuje slika 4.7. Znotraj barvnih kock CB_1 in CB_2 generiramo koncentrični kocki SWT CB_{SWT_1} ter CB_{SWT_2} z dolžino stranice $L_{SWT} = \beta \cdot L_{CB}$, pri čemer je L_{CB} dolžina stranice barvne kocke CB_1 oziroma CB_2 in $0 < \beta \leq 1$. Za vsako izmed obeh kock SWT izračunamo naslednje tri parametre:

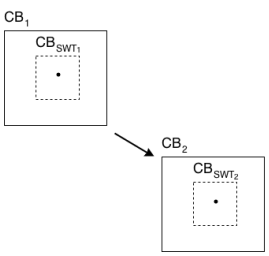
- Gostota SWT D_{SWT}. Gostota SWT je razmerje med številom slikovnih ele-

mentov SWT ter številom trojic (R, G, B), ki jih kocka SWT pokriva:

$$D_{SWT} = \frac{\#(LT_{SWT}(CB_{SWT}))}{\#(h_{RGB}(CB_{SWT}))}, \tag{4.1}$$

pri čemer LT_{SWT} označuje vpogledno tabelo SWT, h_{RGB} pa histogram RGB vhodne slike. Barvne kocke, ki pokrivajo barve teksta, imajo gostoto SWT tipično zelo visoko.

- Standardna deviacija dolžin SWT SD_{SWTL}. SD_{SWTL} meri varianco vrednosti SWT slikovnih elementov, ki jih pokriva kocka SWT. Nižja deviacija pomeni, da kocka SWT pokriva slikovne elemente SWT s podobnimi vrednostmi, kar je indikator, da po vsej verjetnosti pokriva območje barve teksta.
- Standardna deviacija odmikov SWT SD_{SWTO}. SD_{SWTO} je indikator razpršenosti slikovnih elementov SWT znotraj kocke SWT.



Slika 4.7

Koncept glasovanja SWT. CB_1 in CB_2 označujeta barvno kocko pred in po posameznem premiku *mean-shift*. CB_{SWT_1} in CB_{SWT_2} sta koncentrični barvni kocki SWT, ki sta v notranjosti barvnih kock CB_1 in CB_2.

Če za kocki SWT CB_{SWT_1} in CB_{SWT_2} pred določenim premikom *mean-shift* veljajo vsi naslednji štirje pogoji:

$$\begin{aligned} D_{SWT_2} &< D_{SWT_1} \cdot \tau_D \,, \\ D_{SWT_1} &\geq D_{min} \,, \\ SD_{SWTL_2} &> SD_{SWTL_1} \cdot \tau_L \,, \\ SD_{SWTO_2} &< SD_{SWTO_1} \cdot \tau_O \,, \end{aligned} \tag{4.2}$$

metoda premika *mean-shift* ne dovoli ter nadaljuje *mean-shift* naslednje barvne kocke. Pogoji v enačbi (4.2) imajo naslednji pomen:

- Prvi pogoj enačbe (4.2). Tipično imajo barvne kocke, ki prekrivajo barve teksta, zelo visoke gostote SWT v primerjavi z ostalimi barvnimi kockami. Kadar pride do nenadnega in signifikantnega padca gostote SWT, je to eden izmed indikatorjev tranzicije barve teksta proti barvi ozadja.

- Drugi pogoj enačbe (4.2). V praksi se lahko zgodi, da pride do signifikantnega padca gostote SWT tudi med dvema netekstovnima barvnima kockama z zelo nizko gostoto SWT, ki je posledica šuma[3]. V tem primeru drugi pogoj v enačbi (4.2) zagotavlja, da se padec gostote upošteva le, če ima barvna kocka CB_1 dovolj visoko gostoto SWT.

- Tretji pogoj enačbe (4.2). Kadar barvna kocka prekriva barvo teksta, so pripadajoče vrednosti SWT bolj ali manj enake: tekst je tipično uniformne debeline, kar pomeni, da imajo slikovni elementi SWT, ki ležijo na njem, približno enake vrednosti SWT. Posledično imajo takšne barvne kocke nižjo deviacijo dolžin SWT. Signifikantno povečanje deviacije dolžin SWT je prav tako eden izmed indikatorjev tranzicije barve teksta proti barvi ozadja.

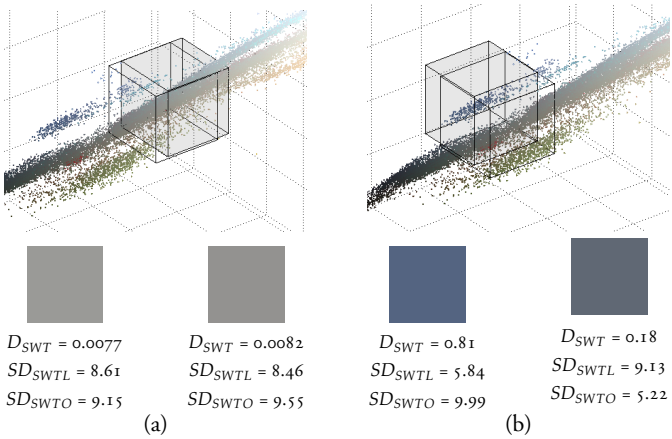
- Četrti pogoj enačbe (4.2). Tranzicija barve teksta proti barvi ozadja je tipično sestavljena iz več posameznih premikov. Bolj, ko se barvna kocka odmika od barve teksta, manj razpršeni so slikovni elementi SWT v njej, ker se nahajajo le še na robovih kocke. Da bi barvo teksta blokirali že pri prvem premiku, mora veljati zadnji pogoj v enačbi (4.2).

Parametri τ_D, τ_L in τ_O regulirajo občutljivost glasovanja SWT. S spreminjanjem njihovih vrednosti vplivamo na število barv, ki so blokirane. Več o parametrih v poglavju 5.

Slika 4.8 prikazuje dva primera premikov *mean-shift*. V primeru (a) premik *mean-shift* dovolimo, saj so tako gostota SWT kot spremembi deviacij zanemarljive. Po drugi strani premika *mean-shift* (b) ne dovolimo, saj je gostota prve barvne kocke zelo visoka, prav tako pa so signifikantni tako padec gostote kot spremembi obeh deviacij.

Primerjava delovanja metode TOCR [48] in metode barvne redukcije na podlagi glasovanja SWT [49] je prikazana na sliki 4.9. Slike 4.9a, 4.9b in 4.9c prikazujejo začetne barvne kocke, pripadajoče barve in začetno barvno reducirano sliko. Metoda

[3]Šum na nivoju zajete slike, šum, ki je posledica pomanjkljivega delovanja filtra SWT, in podobno.



Slika 4.8

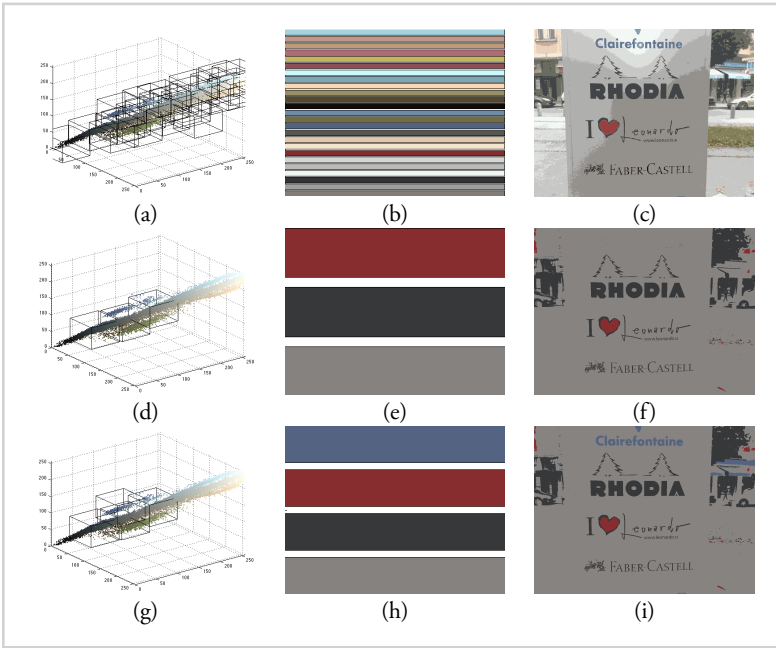
Primera premikov *mean-shift*. (a) Primer premika *mean-shift* svetlo sive proti temno sivi barvi. Pogoji za blokiranje premika niso izpolnjeni. (b) Primer premika *mean-shift* modre proti sivi barvi (modra barva ustreza modri barvi teksta "Clairefontaine" na sliki 4.3c). Pogoji za blokiranje premika so izpolnjeni.

TOCR (slika 4.9d, 4.9e in 4.9f) izpusti modro barvo teksta "Clairefontaine", medtem ko barvna redukcija na podlagi glasovanja SWT modro barvo pravilno segmentira (slika 4.9g, 4.9h in 4.9i).

4.4.3 Detekcija smeri SWT

Metoda SWT [33] detektira paralelne robove z iskanjem v smeri gradienta robnih točk. V primeru temnih tekstov na svetli podlagi (slika 4.10a) predpostavka drži, saj gradienti pravilno kažejo proti notranjosti črk (slika 4.10b). V nasprotnem primeru, ko imamo opravka s svetlimi teksti na temni podlagi (slika 4.10d), pa gradient kaže ven iz črk, kar povzroča nepravilno detekcijo (slika 4.10e). Metoda SWT problem rešuje z dvakratnim izvajanjem celotnega postopka detekcije teksta, tj. v gradientni in protigradientni smeri, in združevanjem rezultatov obeh smeri. Takšen način detekcije ni najbolj natančen, saj pogosto detektira medčrkovne regije. Poleg tega v predlagani metodi celotnega postopka detekcije ne moremo izvajati dvakrat, saj faza glasovanja SWT namreč že predvideva, da slikovni elementi SWT ustrezajo tekstu in ne napačnim regijam. Problem pravilne smeri iskanja rešujemo z metodo detekcije smeri SWT, ki jo opisujemo v nadaljevanju.

Označimo s SWT^+ sliko, generirano z iskanjem v gradientni smeri, in s SWT^-

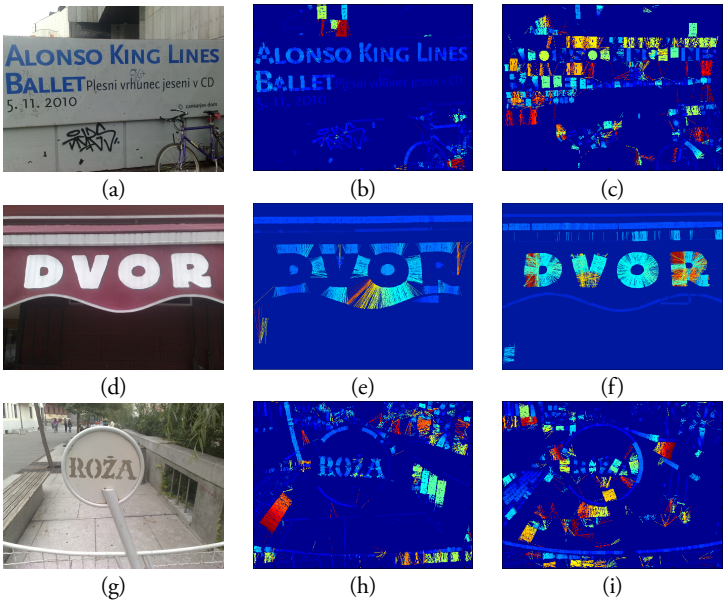


Slika 4.9

Primerjava delovanja metod TOCR in barvne redukcije na podlagi glasovanja SWT. Začetne barvne kocke (a), pripadajoče barve (b) in začetna barvno reducirana slika pred fazo *mean-shift* (c). Končne barvne kocke, generirane z metodo TOCR (d), pripadajoče barve (e) ter pripadajoča končna barvno reducirana slika (f). Končne barvne kocke, generirane z barvno redukcijo na podlagi glasovanja SWT (g), pripadajoče barve (h) ter pripadajoča končna barvno reducirana slika (i). Metoda barvne redukcije na podlagi glasovanja SWT pravilno segmentira modri tekst "Clairefontaine", medtem ko ga metoda TOCR izpusti.

sliko, generirano z iskanjem v protigradientni smeri. Če pozorno pogledamo prvo vrstico slike 4.10, opazimo, da je razlika med slikama SWT^+ in SWT^- zelo očitna. Slika SWT^+, ki ustreza dejanskemu tekstu, je veliko bolj barvno uniformna, saj je tekst v sliki približno enake debeline. Po drugi strani slika SWT^- ustreza netekstovnim regijam zelo različnih debelin, zato je barvno precej heterogena. Razlika med slikama SWT^+ in SWT^- v tretji vrstici slike 4.10 ni tako očitna, vendar predpostavka o barvni homogenosti oz. heterogenosti kljub temu še vedno drži za centralni del slike z napisom "ROŽA".

Na podlagi povedanega je zasnovana metoda detekcije smeri SWT. Najprej na podlagi vhodne slike I izračunamo sliki SWT^+ in SWT^- ter generiramo sliko SWT^* tako, da sliko SWT^+ superponiramo na sliko SWT^-. Ker sta sliki SWT^+ in SWT^- komplementarni (neničelni slikovni elementi SWT^+ se ne prekrivajo z neničelnimi slikovnimi elementi slike SWT^-) v postopku superponiranja sliko SWT^+ vzamemo kot osnovo, in nanjo na pripadajoče lokacije položimo neničelne slikovne elemente slike



(a) (b) (c)
(d) (e) (f)
(g) (h) (i)

Slika 4.10

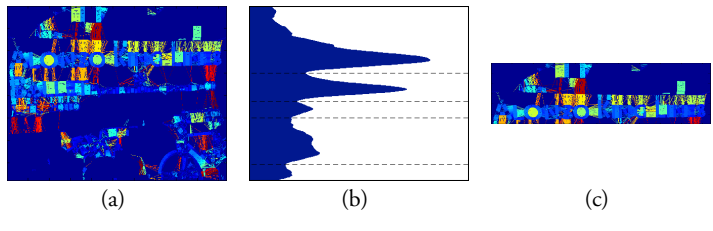
Primeri slik SWT pri različnih smereh iskanja. (a) Slika s temnim tekstom na svetli podlagi in pripadajoči sliki SWT^+ (b) in SWT^- (c). (d) Primer slike s svetlim tekstom na temni podlagi s pripadajočima slikama SWT^+ (e) in SWT^- (f). (g) Primer slike s pripadajočima slikama SWT^+ (h) in SWT^- (i), pri kateri distinkcija med SWT^+ in SWT^- ni več tako očitna.

SWT^-. Z drugimi besedami, slika SWT^* je unija slik SWT^+ in SWT^-. Primer slike SWT^*, generirane iz slike 4.10a, je prikazan na sliki 4.11a. Nato sliko SWT^* razbijemo na bloke. Podobno kot v [26] in [63], kjer se za iskanje tekstovnih regij uporabljajo profili robov (angl. *edge profiles*), za razbitje slike na bloke uporabljamo profile SWT[4] (angl. *SWT profiles*). Najprej sliko SWT^* z uporabo horizontalnih profilov SWT razbijemo na vertikalne bloke[5] (slika 4.11b) in nato vsak vertikalni blok posebej (slika 4.11c) z uporabo vertikalnih profilov SWT dodatno razbijemo na horizontalne bloke.

Po razbitju slike SWT^* na posamezne bloke za vsak blok posebej detektiramo pravilno smer SWT. Postopek detekcije je prikazan na sliki 4.12. Označimo s SWT^*_{SB} blok slike SWT^*, ki ga analiziramo, ter z W in H njegovo dolžino in višino. Iz slik

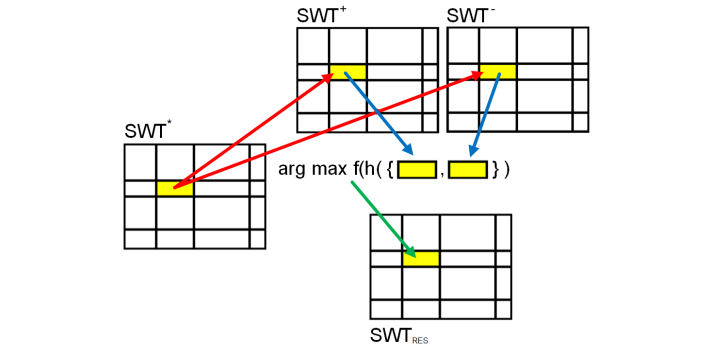
[4]Izraz profil SWT uporabljamo, ker namesto štetja robov štejemo slikovne elemente SWT.

[5]Razbitje na bloke poteka tako, da poiščemo lokalne minimume in maksimume profila SWT ter kot razmejitvene točke med bloki vzamemo tiste lokalne minimume, ki po vrednosti dovolj odstopajo od najbližjega lokalnega maksimuma.



Slika 4.11

Profili SWT. (a) Slika SWT^*, generirana iz slik SWT^+ na sliki 4.10b in SWT^- na sliki 4.10c. (b) Horizontalni profil SWT slike (a). Vertikalni bloki so označeni s črtkanimi črtami. (c) Zgornji vertikalni blok slike (a).



Slika 4.12

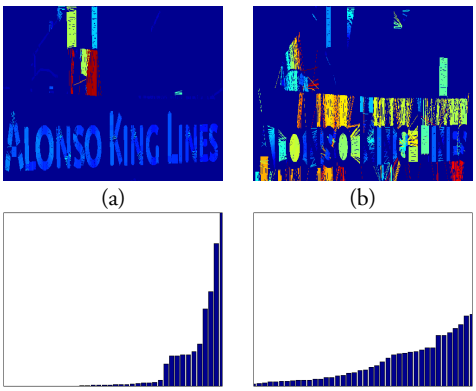
Analiza smeri SWT posameznega bloka. V diagramu je za potrebe prikaza uporabljena histogramska mera f. Namesto mere f se lahko uporabi tudi mera g.

SWT^+ in SWT^- ekstrahiramo bloka SWT^+_{SB} in SWT^-_{SB}, ki sta na isti lokaciji kot blok SWT^*_{SB} in sta velikosti $W \times H$ (rdeči puščici na sliki 4.12), ter generiramo njuna histograma SWT[6] h^+ in h^-:

$$h^+(i) = \#\Big\{ p \in SWT^+_{SB} \mid SWT^+_{SB}(p) > 0 \wedge (i-1)\frac{max_{SWT}}{N_b} < SWT^+_{SB}(p) \leq i\frac{max_{SWT}}{N_b} \Big\}, \tag{4.3}$$

$$h^-(i) = \#\Big\{ p \in SWT^-_{SB} \mid SWT^-_{SB}(p) > 0 \wedge (i-1)\frac{max_{SWT}}{N_b} < SWT^-_{SB}(p) \leq i\frac{max_{SWT}}{N_b} \Big\}, \tag{4.4}$$

[6]Vsaka celica histograma SWT vsebuje število slikovnih elementov v bloku, ki imajo vrednost SWT znotraj intervala celice.



Slika 4.13

Primer blokov in pripadajočih histogramov. (a) Blok SWT^{+}_{SB}, ki ustreza bloku na sliki 4.11c. (b) Blok SWT^{-}_{SB}, ki ustreza bloku na sliki 4.11c. (c) Histogram h^{+} bloka (a) z merama $f = 425{,}7$ in $g = 177{,}1$. (d) Histogram h^{-} bloka (b) z merama $f = 111{,}5$ in $g = 29{,}6$.

pri čemer je p slikovni element, N_b število celic histograma, i indeks posamezne celice in max_{SWT} maksimalna vrednost SWT obeh blokov SWT^{+}_{SB} in SWT^{-}_{SB}. Z drugimi besedami, histograma h^{+} in h^{-} predstavljata porazdelitev vrednosti SWT slikovnih elementov blokov SWT^{+}_{SB} in SWT^{-}_{SB}.

Histograma h^{+} in h^{-} nato sortiramo v naraščajočem vrstnem redu. Primera h^{+} in h^{-} sta prikazana na sliki 4.13. Histogram h+, ki ustreza dejanskemu tekstu, je veliko bolj strm, bolj kompakten in nezvezen, za razliko od histograma h^{-}, ki je širši in narašča bolj zvezno. Povedano se zdi smiselno, saj imajo slikovni elementi, ki pripadajo tekstu, uniformne vrednosti SWT, ki so v histogramu h^{+} močno skoncentrirane. Po drugi strani netekstovne strukture predstavljajo šum, ki je v histogramu h^{-} bolj enakomerno porazdeljen preko celotnega spektra vrednosti SWT.

Da bi favorizirali histograme, ki ustrezajo dejanskemu tekstu, predlagamo meri f in g:

$$f(h) = \frac{1}{N_{nz}} \left(\max_i h(i) - \min_i h(i) \right), \tag{4.5}$$

$$g(h) = \frac{1}{N_{nz}} \sqrt{\sum_{i=2}^{N_b} \left(h(i) - h(i-1) \right)^2}, \tag{4.6}$$

pri čemer je h histogram SWT in N_{nz} število neničelnih celic histograma. Deljenje s

številom neničelnih celic N_{nz} je ključno, saj favorizira ožje histograme. Izmed blokov SWT^+ in SWT^- izberemo tistega, ki ima večjo vrednost mere f oziroma g (modri puščici na sliki 4.12)[7]:

$$SWT_{SB} = \Big(\arg\max f(h(B)) \mid B \in \{SWT^+_{SB}, SWT^-_{SB}\}\Big), \tag{4.7}$$

$$SWT_{SB} = \Big(\arg\max g(h(B)) \mid B \in \{SWT^+_{SB}, SWT^-_{SB}\}\Big). \tag{4.8}$$

Blok SWT_{SB} zlepimo na pripadajoče mesto v končno sliko SWT_{RES} (zelena puščica na sliki 4.12), ki je enake velikosti kot vhodna slika I.

4.4.4 *Zgornja meja SWT*

Metoda SWT pri iskanju v napačni smeri pogosto trči ob ravne strukture v sliki (robovi tabel, zidovi ipd.), pri čemer generira paralelne regije konstantnih debelin, ki lahko dosegajo zelo visoke vrednosti f in g. Primer takšne anomalije je prikazan na sliki 4.14. Zaradi kompaktne regije rumene barve na sliki 4.14b ima pripadajoči histogram na sliki 4.14f veliko večji vrednosti f in g kot histogram na sliki 4.14g, ki ustreza dejanskemu tekstu na sliki 4.14c.

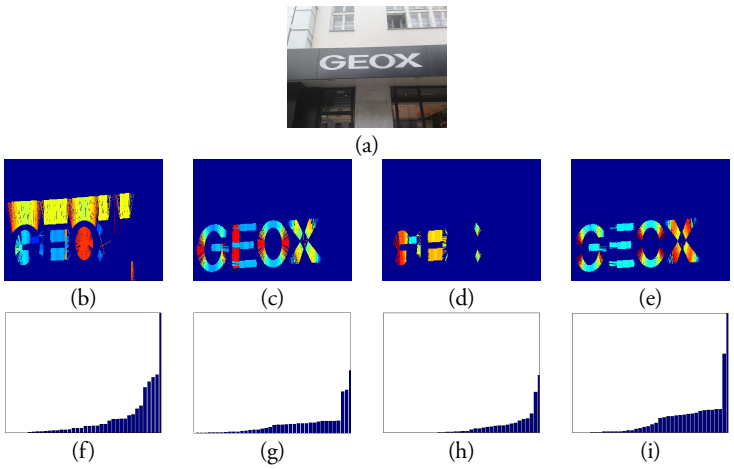
Da bi problem motečih struktur čim bolj omejili, vpeljujemo koncept zgornje meje SWT (angl. *upper SWT boundary*) in vse slikovne elemente SWT, ki mejo presegajo, postavimo na nič. Zgornja meja SWT ub_{SWT} je definirana kot obteženo povprečje povprečnih vrednosti blokov SWT^+_{SB} in SWT^-_{SB}:

$$\begin{aligned} ub_{SWT} = \; & \alpha \cdot \min\Big(\mu(SWT^+_{SB}), \mu(SWT^-_{SB})\Big) + \\ & (1-\alpha) \cdot \max\Big(\mu(SWT^+_{SB}), \mu(SWT^-_{SB})\Big), \end{aligned} \tag{4.9}$$

pri čemer je $\mu()$ povprečna vrednost bloka in $0 < \alpha < 1$. Več o parametru α v poglavju 5. Glede na to, da je razdalja med robovi posamezne črke tipično krajša od razdalje med črkami in ostalimi strukturami, zgornja meja SWT poreže daljše (tipično netekstovne) strukture. Seveda predpostavka o krajših razdaljah med samimi robovi črke ne drži vedno, zato se zgornja meja ne uporablja kot absolutno merilo, temveč le kot dodaten pripomoček pri detekciji smeri SWT.

Filtriranje z zgornjo mejo SWT se izvaja na vsakem bloku posebej. Pred generiranjem histogramov h^+ in h^- ter izračunavanjem vrednosti f oziroma g sta bloka SWT^+_{SB}

[7] Pri izbiri bloka se meri f in g uporabljata opcijsko – vedno se uporablja le ena izmed njiju. Učinkovitost obeh mer je razvidna pri eksperimentalnih rezultatih v poglavju 5.



Slika 4.14

Primer blokov SWT brez in z uporabo zgornje meje SWT. (a) Originalna slika. (b) Blok SWT^{+}_{SB} in (c) blok SWT^{-}_{SB} brez uporabe zgornje meje SWT. (d) Blok SWT^{+}_{SB} in (c) blok SWT_SB^{-} z uporabo zgornje meje SWT, pri čemer je α = 0,33. Pripadajoči histogrami so prikazani v spodnji vrstici slike: (f) f = 263,2 in g = 144,0, (g) f = 119,1 in g = 67,1, (h) f = 85,9 in g = 42,3, (i) f = 133,1 in g = 76,9.

in SWT^{-}_{SB} predhodno filtrirana z zgornjo mejo SWT. Primer filtriranja je prikazan na sliki 4.14. Histogram na sliki 4.14i, ki ustreza dejanskemu tekstu na sliki 4.14e, tokrat dosega veliko večji vrednosti f in g kot netekstovni histogram na sliki 4.14h.

4.5 Diskusija

V poglavju smo opisali prednosti in slabosti metod SWT [33] in TOCR [48] ter predstavili metodi barvne redukcije na podlagi glasovanja SWT [49] in detekcije smeri SWT [49].

Za razliko od večine metod detekcije teksta, ki tekst tipično detektirajo bodisi na podlagi strukture bodisi na podlagi barve, metoda barvne redukcije na podlagi glasovanja SWT temelji na integraciji obeh. Metoda osnovni koncept barvne redukcije, ki je premalo tekstovno usmerjen, nadgrajuje z iskanjem korespondence med barvo in geometrijo struktur v sliki, ki jih barva pokriva. Barve, bogate s slikovnimi elementi SWT, po vsej verjetnosti pripadajo tekstu, zato so v fazi *mean-shift* blokirane. Izbira SWT kot strukturne značilke je smiselna, saj poleg intuitivnosti hipoteze o paralelnosti robov teksta metoda SWT dosega zelo dobre rezultate. Če metoda SWT določenih slikovnih elementov, ki sicer pripadajo tekstu, ne detektira pravilno, to drastično ne vpliva

na delovanje predlagane metode. Preostali pravilno najdeni slikovni elementi SWT še vedno nosijo dovolj strukturne informacije, da pravilno usmerjajo postopek barvne redukcije. Če metoda SWT ne detektira nobenega slikovnega elementa, ki pripada tekstu, predlagana metoda deluje praktično identično kot metoda TOCR, saj nima na voljo nobene strukturne informacije o tekstu v sliki. Tudi kadar metoda SWT detektira slikovne elemente, ki ne pripadajo tekstu, to na delovanje metode ne vpliva preveč, saj nepravilno detektirani slikovni elementi prekrivajo le manjše dele ozadij in gostote SWT barvam ozadja ne dvignejo premočno. V najslabšem možnem primeru metoda barvne redukcije na podlagi glasovanja SWT deluje enakovredno metodi TOCR, v vseh ostalih primerih pa jo izboljšuje. Omenjena dejstva potrjujemo v naslednjem poglavju, v katerem podajamo rezultate posameznih metod.

Metoda detekcije smeri SWT, ki je opisana na koncu poglavja, je pomemben dosežek. Za razliko od originalne metode SWT, ki celoten postopek detekcije teksta izvaja dvakrat – v gradientni in protigradientni smeri – predlagana metoda omogoča takojšnje generiranje pravilne slike SWT ne glede na barvo teksta in ozadij v sliki.

5

Eksperimentalni rezultati

5.1 Uvod

Za evalvacijo metod smo uporabili testno množico, sestavljeno iz prvih 60 binarno anotiranih slik normalne kategorije zbirke CVL OCR DB (v nadaljevanju CVL OCR BIN DB[1]) [47]. Pri izbiri velikosti testne množice smo se odločili za kompromis med zadostnim številom slik in naporno binarizacijo slik teksta v naravnih scenah. Avtorji metode TOCR [48] so metodo testirali na privatni evalvacijski zbirki, ki vključuje 50 slik [48], zato smo se odločili, da število slik v svoji testni množici še rahlo povečamo. Evalvacijska zbirka [48] je sestavljena iz slik naslovnic knjig z zelo kontrastnimi barvami, ki v prostoru RGB ležijo precej narazen (slika 5.1). Zbirka CVL OCR BIN DB s tega stališča predstavlja veliko trši oreh, saj vsebuje zelo kompleksne, barvno nekontrastne slike raznovrstnih naravnih scen.

Slika 5.1

Primeri slik evalvacijske zbirke [48].

5.2 Detekcija smeri SWT

Metodo detekcije smeri SWT smo evalvirali na zbirki CVL OCR BIN DB. Ker metoda detekcije smeri SWT vhodno sliko dinamično razdeli na določeno število blokov in analizira vsakega posebej, smo pri evalvaciji upoštevali le tiste, ki so vsebovali tekst. Izmed njih smo prešteli takšne, na katerih je metoda smer SWT pravilno detektirala in izračunali stopnjo detekcije. Za objektivnejše ovrednotenje učinkovitosti mer f in g smo za določanje smeri SWT uporabili tudi dodatne splošno razširjene histogramske mere: entropijo, standardno deviacijo, nagnjenost (angl. *skew*) in kurtozo (angl. *kurtosis*). Enako kot pri f in g je bil pri vseh dodatnih merah (razen pri entropiji)

[1]Beseda BIN označuje, da gre za binarno anotirane slike.

izmed blokov SWT^{+}_{SB} in SWT^{-}_{SB} detektiran tisti z višjo vrednostjo mere. V primeru entropije je bil zaradi njene narave detektiran blok z nižjo vrednostjo. Pri evalvaciji smo uporabili naslednje empirično določene vrednosti parametrov: $N_B = 40$ in $\alpha = 0{,}33$ (glej enačbe (4.3), (4.4) in (4.9)). Rezultati evalvacije so prikazani v tabeli 5.1.[2]

Pri detekciji smeri SWT brez uporabe zgornje meje SWT dosega entropija najvišjo stopnjo detekcije (86,74%), sledita pa ji meri f (82,65%) in g (82,6%). Stopnja detekcije ostalih mer je precej nižja. Razlog za slabše delovanje mer f in g glede na entropijo je naslednji: meri f in g sta zasnovani, da favorizirata ozke in stopničaste histograme, ki so značilni za tekst. V primeru slabe particije na bloke ta hipoteza ne drži vedno, saj lahko določen blok pokriva le manjši del teksta in večji del homogenih okoliških struktur. Izkaže se, da je v takšnem primeru bolje izbrati histogram, ki je manj neurejen, torej takšen, ki ima nižjo entropijo. Uporaba zgornje meje SWT popolnoma spremeni situacijo. Zaradi tekstu prijaznega filtriranja predolgih struktur je slabih particij na bloke občutno manj, kar precej dvigne stopnjo detekcije mer f in g. Mera f pri uporabi zgornje meje SWT dosega najvišjo stopnjo detekcije (91,85%), medtem ko mera g za odtenek nižjo (90,37%). Ker so zaradi boljše particije na bloke razlike med tekstovnimi in netekstovnimi histogrami bolj očitne, koncept neurejenosti izgubi pomen, kar je razlog za slabšo stopnjo detekcije pri uporabi entropije (83,70%). V splošnem velja, da je nepravilna detekcija smeri SWT najpogosteje posledica slabe particije slike na bloke, tj. kadar posamezni blok vsebuje veliko netekstovnih struktur in ko tekst predstavlja le manjši del njegove površine.

Primeri kvalitativne evalvacije detekcije smeri SWT na posameznih blokih so prikazani na slikah 5.2, 5.3 in 5.4. Slika 5.2a predstavlja primer kompleksnega teksta na vertikalni črtasti podlagi. Kljub motečim dejavnikom metoda pravilno detektira smer SWT. Histogram bloka, ki pripada dejanskemu tekstu (sliki 5.2c in 5.2e), dosega višji vrednosti mer f in g kot histogram netekstovnega bloka (sliki 5.2b in 5.2d).

Slika 5.3a prikazuje primer zelo kompleksnega teksta. Zaradi zavitosti teksta je iz pripadajočih blokov SWT^{+}_{SB} (slika 5.3b) in SWT^{-}_{SB} (slika 5.3c) zelo težko določiti, kateri blok je pravi – vsaj nekateri deli bloka delujejo zelo podobno. Kljub temu metoda pravilno detektira smer SWT, vendar z minimalno razliko. Pri razlikovanju blokov mera f deluje malenkost bolje, saj je razlika med histogramoma (sliki 5.3d in 5.3e) z uporabo mere f enaka 11,1, medtem ko je razlika med njima z uporabo mere g enaka

[2] Podobni rezultati so objavljeni v [49], vendar je tam uporabljena slabša particijska metoda, zato se število blokov v tabeli 5.1 in [49] razlikuje.

Tabela 5.1

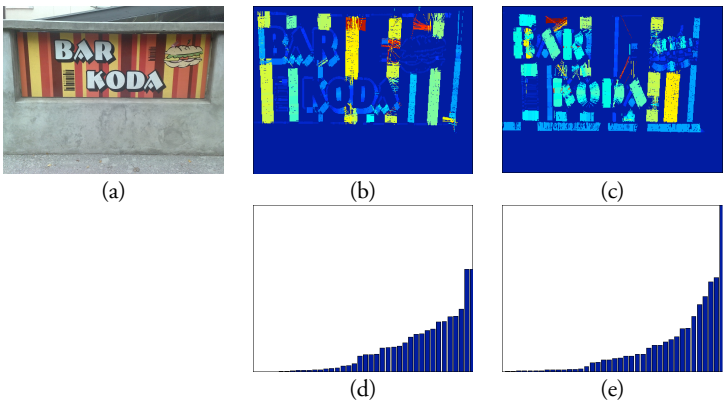
Rezultati detekcije smeri SWT. Stolpec *Pravilni bloki* označuje število tekstovnih blokov s pravilno detektirano smerjo SWT, stolpec *Vsi bloki* pa število vseh tekstovnih blokov, ki smo jih evalvirali. Število vseh tekstovnih blokov pri neuporabi (98) in uporabi zgornje meje SWT (135) je različno, saj filtriranje z zgornjo mejo SWT spremeni videz slike SWT, zaradi česar je particija na bloke drugačna kot v primeru nefiltriranja.

Metoda	*Mera*	*Pravilni bloki*	*Vsi bloki*	*Stopnja detekcije*
Brez zgornje meje SWT	entropija	85	98	86,74 %
	f	81	98	82,65 %
	g	81	98	82,65 %
	nagnjenost (angl. *skew*)	73	98	74,49 %
	standardna deviacija	70	98	71,43 %
	kurtoza (angl. *kurtosis*)	68	98	69,39 %
Z zgornjo mejo SWT	f	124	135	91,85 %
	g	122	135	90,37 %
	standardna deviacija	120	135	88,89 %
	entropija	113	135	83,70 %
	nagnjenost (angl. *skew*)	100	135	74,07 %
	kurtoza (angl. *kurtosis*)	97	135	71,85 %

1,1.

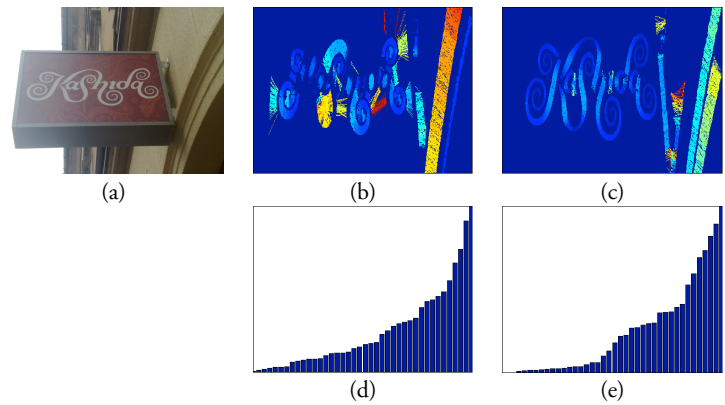
Primer napačne detekcije smeri SWT je prikazan na sliki 5.4. V tem primeru niti mera f niti mera g ne favorizirata tekstovnega histograma (sliki 5.4c in 5.4e). Neteksovni blok (slika 5.4b in 5.4d) vsebuje dve zelo kompaktni rumeni regiji s paralelnimi robovi, ki povzročita močno koncentracijo slikovnih elementov v histogramu.

Rezultate detekcije smeri SWT na celotnih slikah (po analizi vseh blokov v sliki) prikazuje slika 5.5. Sliki (a) in (b) predstavljata primera pravilne detekcije smeri SWT, tudi kadar sta v sliki prisotna tako temen tekst na svetli podlagi kot svetel tekst na temni podlagi. Slika (c) predstavlja primer, ko metoda zaradi kompaktnega ozadja na-



Slika 5.2

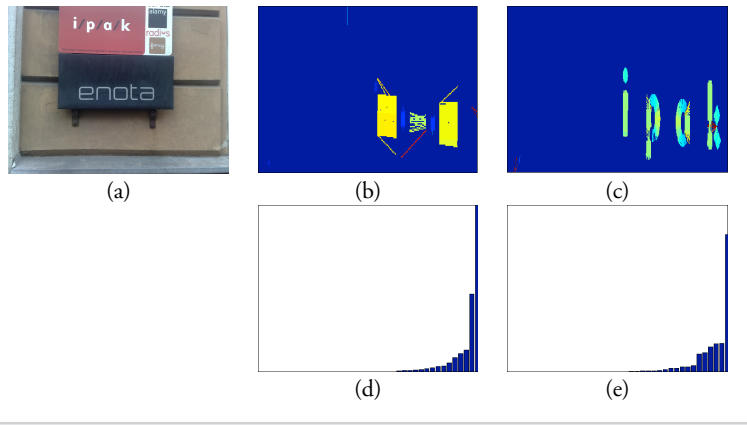
Pravilna detekcija smeri SWT kljub črtasti vertikalni podlagi. (a) Originalna slika. (b) Blok SWT^{+}_{SB} in pripadajoči histogram SWT (d). (c) Blok SWT^{-}_{SB} in pripadajoči histogram SWT (e). (d) f = 284,9 in g = 120,6. (e) f = 426,7 in g = 199,2.



Slika 5.3

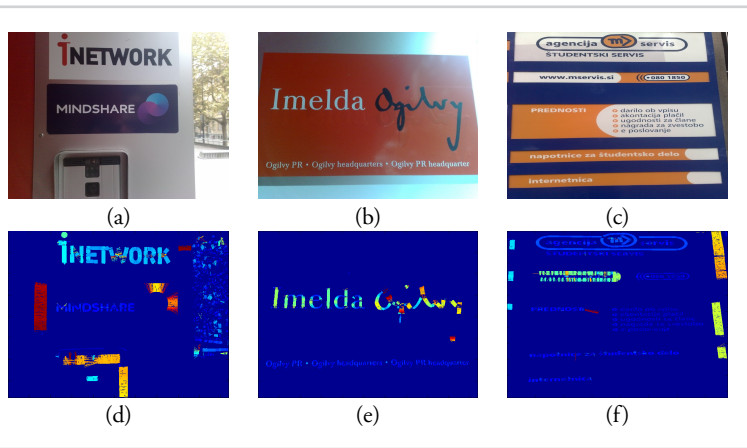
Mejno delovanje metode. (a) Originalna slika. (b) Blok SWT^{+}_{SB} in pripadajoči histogram SWT (d). (c) Blok SWT^{-}_{SB} in pripadajoči histogram SWT (e). (d) f = 132,7 in g = 36,2. (e) f = 143,8 in g = 37,3.

pačno detektira smer teksta ”www.mservis.si”. Smeri ostalega teksta v sliki (c) metoda detektira pravilno.



Slika 5.4

Nepravilno delovanje metode zaradi medčrkovnih regij rumene barve. Tako mera f kot g favorizirata napačen (netekstovni) histogram. (a) Originalna slika. (b) Blok SWT^{+}_{SB} in pripadajoči histogram SWT (d). (c) Blok SWT^{-}_{SB} in pripadajoči histogram SWT (e). (d) f = 147,3 in g = 93,4. (e) f = 100,9 in g = 80,9.



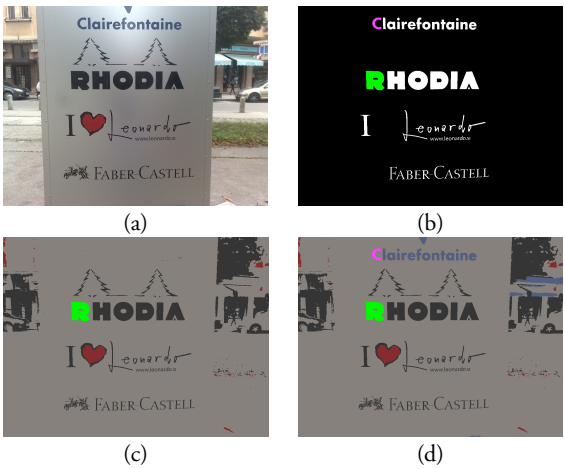
Slika 5.5

Detekcija smeri SWT na celotnih slikah. Zgornja vrstica prikazuje originalne slike, spodnja pa pripadajoče slike SWT na podlagi detektirane smeri SWT.

5.3 *Barvna redukcija na podlagi SWT*

Metodi barvne redukcije na podlagi glasovanja SWT [49] in TOCR [48] smo evalvirali na zbirki CVL OCR BIN DB. Postopek evalvacije je podoben [48] in je prikazan na sliki 5.6. Za vsako sliko v zbirki (slika 5.6a) generiramo barvno reducirani sliki I_{CR} (slika 5.6c) in I_{CRSWT} (slika 5.6d), pri čemer I_{CR} označuje barvno reducirano sliko, generira-

no z metodo TOCR, medtem ko I_{CRSWT} označuje barvno reducirano sliko, generirano z metodo barvne redukcije na podlagi glasovanja SWT. Na slikah I_{CR}, I_{CRSWT} in pripadajoči binarno anotirani sliki (slika 5.6b) poiščemo vse povezane komponente.[3] Za vsako povezano komponento binarno anotirane slike (označimo jo s CC_{GT}) poiščemo povezano komponento barvno reducirane slike (označimo jo s CC_{DET}), ki se ji najbolje prilega. Če sta razmerje površin CC_{GT} in CC_{DET} ter razmerje površin $CC_{DET} \cap CC_{GT}$ in CC_{GT} večja od praga T_R, zabeležimo zadetek. V nasprotnem primeru zadetka ni. Na sliki 5.6b sta z zeleno in vijolično barvo označeni povezani komponenti, ki ustrezata črkama "R" v besedi "RHODIA" in "C" v besedi "Clairefontaine". Črka "R" ima zadetek tako na sliki I_{CR} (slika 5.6c) kot I_{CRSWT} (slika 5.6d), medtem ko ima črka "C" zadetek le na sliki I_{CRSWT} (slika 5.6d).



Slika 5.6

Generiranje povezanih komponent za potrebe evalvacije. Z zeleno in vijolično barvo sta označena primera povezanih komponent, ki ustrezata črkama "R" in "C". (a) Originalna slika. (b) Pripadajoča binarno anotirana slika. (c) Barvno reducirana slika I_{CR}. (d) Barvno reducirana slika I_{CRSWT}.

Da bi uspešnost obeh metod določili čim bolj objektivno, smo evalvacijo izvajali pri različnih pragih T_R. Pri tem smo za vsak prag posebej določili absolutno stopnjo detekcije in povprečno stopnjo detekcije posamezne metode.[4] Pri evalvaciji smo upo-

[3] Sosednja slikovna elementa, ki sta enake barve, pripadata isti povezani komponenti.

[4] Absolutna stopnja detekcije označuje stopnjo detekcije vseh povezanih komponent na vseh slikah zbirke, medtem ko povprečna stopnja detekcije, ki jo uporabljata Nikolaou in Papamarkos [48], označuje povprečje stopenj detekcije na posameznih slikah. Obe meri sta podrobno opisani v poglavju 2.6.3

rabili naslednje empirično določene vrednosti parametrov: β = 0.63, D_{min} = 0.18, τ_D = 0.70, τ_L = 0.80 in τ_O = 0.80 (enačba (4.2)). Vse ostale vrednosti parametrov barvne redukcije so bile enake kot v [48]. Rezultati evalvacije so prikazani v tabelah 5.2 in 5.3. Metoda barvne redukcije na podlagi glasovanja SWT dosega boljše rezultate od metode TOCR pri vseh pragih, kar je vizualno predstavljeno tudi na sliki 5.7.

Tabela 5.2

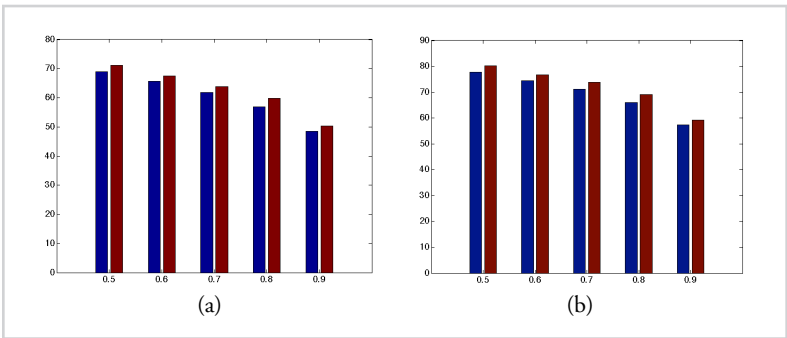
Absolutna stopnja detekcije.

	Prag T_R				
	0,5	0,6	0,7	0,8	0,9
TOCR [48]	69,02 %	65,75 %	61,86 %	56,95 %	48,53 %
Barvna redukcija na podlagi glasovanja SWT	71,01 %	67,48 %	63,93 %	59,84 %	50,41 %

Tabela 5.3

Povprečna stopnja detekcije.

	Prag T_R				
	0,5	0,6	0,7	0,8	0,9
TOCR [48]	77,71 %	74,40 %	71,06 %	66,02 %	57,40 %
Barvna redukcija na podlagi glasovanja SWT	80,24 %	76,72 %	73,77 %	69,12 %	59,27 %



Slika 5.7

Absolutna stopnja detekcije (a) in povprečna stopnja detekcije (b) pri različnih pragih T_R (os x). Modri stolpci ustrezajo metodi TOCR [48], rdeči pa metodi barvne redukcije na podlagi glasovanja SWT.

Slika 5.8

Primerjava delovanja metode TOCR [48] in metode barvne redukcije na podlagi glasovanja SWT. Prva kolona (a) ustreza originalnim slikam, medtem ko druga (b) in tretja kolona (c) ustrezata pripadajočim barvno reduciranim slikam, dobljenim z metodo TOCR in metodo barvne redukcije na podlagi glasovanja SWT.

Na sliki 5.8 je prikazanih nekaj primerov delovanja obeh metod v praksi. V prvih treh vrsticah slike metoda barvne redukcije na podlagi glasovanja SWT pravilno detektira dele teksta (“S” v besedi “Segafredo”, “INES tours”, “PECIVO” itd.), ki jih metoda TOCR ne detektira. Četrta vrstica slike prikazuje primer, ko metoda TOCR sicer detektira črko “ϵ”, vendar po sreči. Barvna redukcija namreč rdeče barve ne ohra-

ni, zato se v fazi generiranja barvno reducirane slike rdeči barvi priredi najbližja končna barva – v tem primeru siva barva. Predlagana metoda rdečo barvo črke "ϵ" pravilno ohrani. Zadnja vrstica slike prikazuje primer, ko obe metodi teksta ne detektirata pravilno. Kljub temu predlagana metoda za razliko od metode TOCR detektira vsaj del prve črke "C".

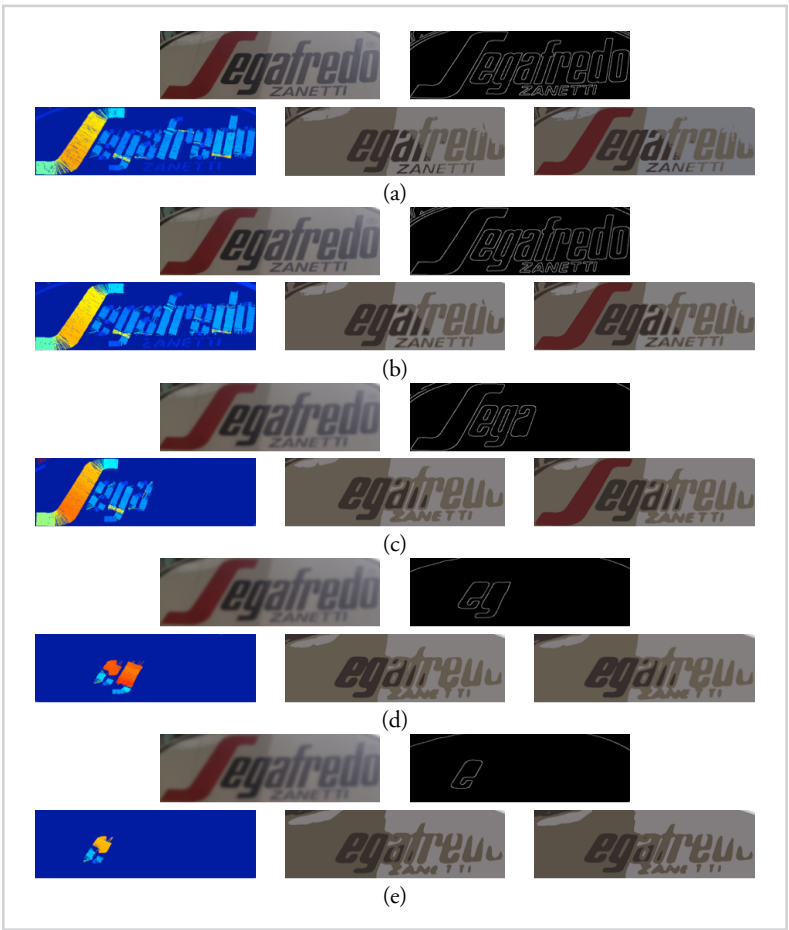
5.4 Vpliv kakovosti slik na delovanje metod

V računalniškem vidu igra kakovost vhodnih slik zelo pomembno vlogo, zato smo se odločili, da natančneje analiziramo njen vpliv na delovanje metod detekcije teksta. S to tematiko smo se v doktorski nalogi sicer delno že ukvarjali (v poglavjih 4.2 in 4.3 smo prikazali nekaj primerov delovanja metod SWT [33] in TOCR [48] na slikah, zajetih pri slabših svetlobnih pogojih), vendar se je na tem mestu lotevamo podrobneje in bolj sistematično. Kakovost slik tipično ocenjujemo na podlagi različnih kriterijev, med katerimi najbolj izstopata ostrina slike in količina šuma v sliki, zato se v nadaljevanju osredotočamo nanju. Da bi analizirali vpliv kakovosti slik na delovanje metod detekcije teksta, smo izbrali manjšo testno množico slik iz zbirke CVL OCR BIN DB in opazovali obnašanje vseh treh metod (metode SWT, metode TOCR in metode barvne redukcije na podlagi glasovanja SWT) pri postopnem zmanjševanju ostrine slike (test ostrine) in postopnem povečevanju količine šuma v sliki (test šuma).

5.4.1 Test ostrine

Pri testu ostrine smo originalni sliki z uporabo Gaussovega zamegljevanja (angl. *Gaussian blur*) iterativno zmanjševali ostrino in v vsaki iteraciji opazovali obnašanje metod SWT, TOCR in metode barvne redukcije na podlagi glasovanja SWT. Za Gaussovo zamegljevanje smo uporabili filter z velikostjo jedra 5×5 in $\sigma = 5$. Vrednosti obeh parametrov smo določili empirično, pri čemer smo pazili, da je bila degradacija slike ravno pravšnja. V primeru prehitre/prepočasne degradacije slike bi namreč težje nazorno grafično prikazali postopno poslabševanje delovanja metod.

Slika 5.9 prikazuje obnašanje metod pri različnih ostrinah slike z napisom "Segafredo ZANETTI". Zaradi natančnosti prikaza je prikazan le centralni del slike z napisom. Slika 5.9 je sestavljena iz petih sklopov, označenih od (a) do (e). Vsak sklop vsebuje pet slik. Zgornja leva slika sklopa predstavlja vhodno sliko z določeno stopnjo ostrine, medtem ko desna zgornja slika sklopa predstavlja pripadajočo sliko robov, detektiranih s Cannyjevim detektorjem robov [58]. Spodnje tri slike sklopa ustrezajo sliki SWT (na



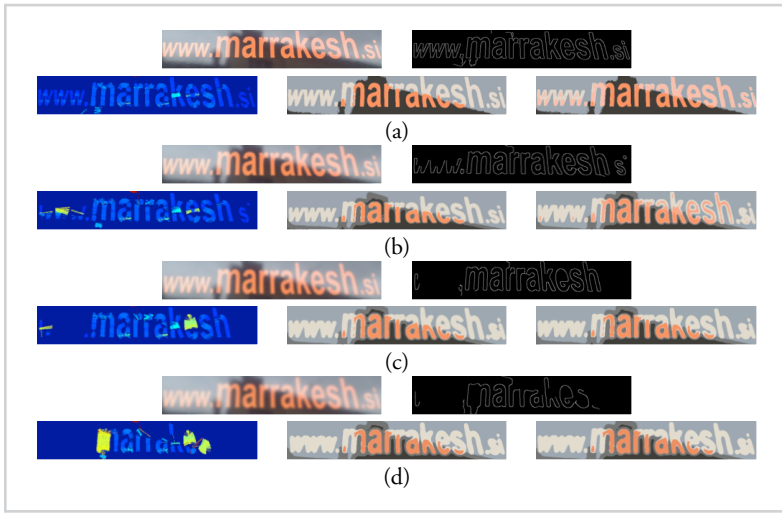
Slika 5.9

Odvisnost delovanja metod od ostrine slike (prvi primer). Vsak sklop od (a) do (e) je sestavljen iz petih slik. Leva zgornja slika sklopa je originalna slika, ki ji postopno zmanjšujemo ostrino, desna zgornja slika sklopa je pripadajoča slika robov, spodnje tri slike sklopa od leve proti desni predstavljajo sliko SWT, sliko, generirano z metodo TOCR, in sliko, generirano z barvno redukcijo na podlagi glasovanja SWT. (a) Originalna slika pred prvo iteracijo. (b) Prva iteracija. (c) Peta iteracija. (d) Osma iteracija. (e) Deseta iteracija.

levi strani), sliki, generirani z metodo TOCR (na sredini), in sliki, generirani z barvno redukcijo na podlagi glasovanja SWT (na desni strani). Sklop (a) ustreza originalni sliki pred postopkom zamegljevanja, medtem ko sklopi (b), (c), (d) in (e) ustrezajo prvi, peti, osmi in deseti iteraciji zamegljevanja.[5] Opazimo lahko, da zamegljevanje na

[5]Zaradi jasnosti prikaza prikazujemo le najpomembnejše iteracije, tj. tiste, pri katerih pride do zanimivega

delovanje metode TOCR ne vpliva drastično, saj se skozi iteracije barvno reducirana slika ne spreminja premočno. Po drugi strani zamegljevanje predstavlja velik problem detektorju robov. Že pri peti iteraciji (c) zamegljevanja detektor robov detektira le še štiri črke napisa. Nepravilna detekcija robov posledično vpliva na nepravilno delovanje metode SWT. Ko metoda SWT v osmi iteraciji (d) ne detektira rdeče črke "S", tudi metoda barvne redukcije na podlagi glasovanja SWT rdeče barve ne blokira več in deluje praktično identično metodi TOCR.



Slika 5.10

Odvisnost delovanja metod od ostrine slike (drugi primer). Pomen posameznih slik in režim označevanja sta enaka kot na sliki 5.9. (a) Originalna slika pred prvo iteracijo. (b) Tretja iteracija. (c) Peta iteracija. (d) Deseta iteracija.

Podobno situacijo prikazuje slika 5.10. Tudi v tem primeru je veliko bolj kot barvna redukcija na zamegljevanje občutljiv detektor robov in s tem posledično metoda SWT. V peti iteraciji zamegljevanja (c) metoda SWT pravilno detektira malo več kot polovico črk, medtem ko v deseti iteraciji (d) metoda SWT skoraj odpove. Razlika med metodo TOCR in barvno redukcijo na podlagi glasovanja SWT ni velika, je pa zaznavna. Do tretje iteracije (b) barvna redukcija na podlagi glasovanja SWT še uspe blokirati svetlejši odtenek oranžne barve dela napisa "sh.si" (označimo ga s C_{or}), po tretji iteraciji pa ta že konvergira proti nesaturiranemu odtenku skoraj bele barve (c, d). Konvergiranje barve C_{or} po tretji iteraciji je razumljivo: v tretji iteraciji (b) metoda SWT del napisa "sh.si"

obnašanja metod.

še pravilno detektira, kar barvi C_{or} prinese dovolj glasov SWT. V kasnejših iteracijah tega dela napisa metoda SWT ne detektira več, zato barva C_{or} nima dovolj glasov SWT, da bi bila blokirana.

5.4.2 *Test šuma*

Pri testu šuma smo iterativno povečevali šum v slikah in v vsaki iteraciji opazovali obnašanje metod SWT, TOCR in metode barvne redukcije na podlagi glasovanja SWT. Slike smo zašumljali umetno z dodajanjem Gaussovega šuma s povprečjem 0 in varianco 0,001. Podobno kot pri testu ostrine smo tudi v tem primeru oba parametra določili empirično in pazili, da je bila degradacija slike primerna za dovolj nazoren grafični prikaz poslabšanja delovanja posameznih metod.

Slika 5.11 prikazuje obnašanje metod na sliki s tekstom "Segafredo ZANETTI" pri postopnem dodajanju šuma. Sklop (a) predstavlja originalno sliko brez umetno dodanega šuma, medtem ko sklopi (b), (c), (d) in (e) ustrezajo prvi, drugi, tretji in četrti iteraciji. Opazimo lahko, da je občutljivost metod na količino šuma ravno obratna kot pri testu ostrine. Detektor robov in metoda SWT delujeta stabilno v vseh prikazanih iteracijah, medtem ko ima barvna redukcija težave že pri prvi iteraciji. Barvna redukcija na podlagi glasovanja SWT v prvih treh iteracijah uspe blokirati rdečo barvo črke "S", v četrti iteraciji pa rdeča barva konvergira proti barvi ozadja – kljub temu, da je metoda SWT črko "S" detektirala popolnoma pravilno. Razlog za takšno obnašanje metode se skriva v razraščanju barvnega histograma slike. Slika 5.12 prikazuje barvne histograme vhodne slike v posameznih iteracijah. Z vsako iteracijo se barvni histogram vse bolj razrašča in močno otežuje proces barvne redukcije. V četrti iteraciji je zaradi razraščanja barvnega histograma rdeča barva preveč razpršena po barvnem prostoru, da bi dobila dovolj glasov SWT.

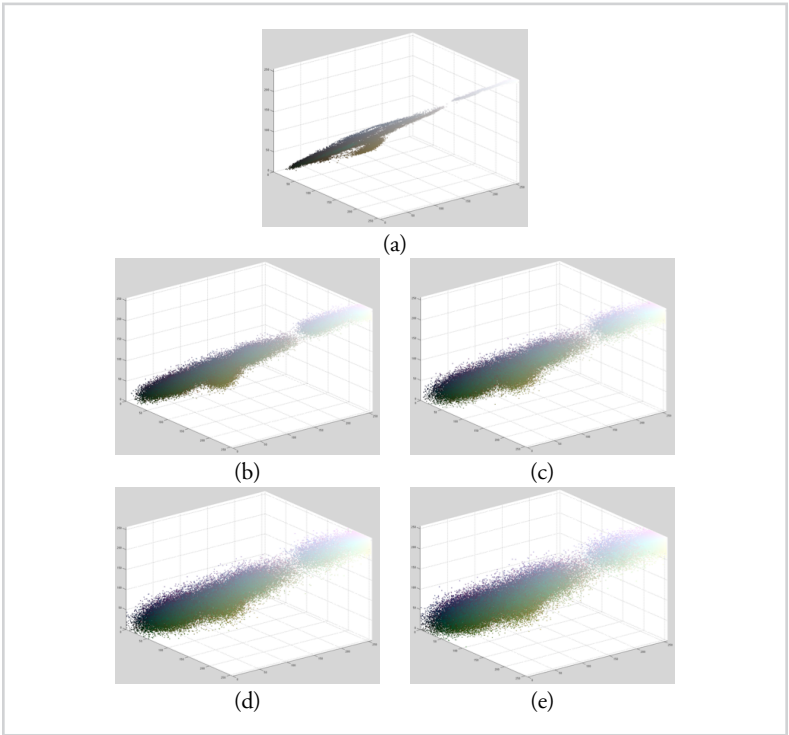
Na sliki 5.11 vidimo, da detektor robov in metoda SWT v prvih štirih iteracijah delujeta pravilno. Da bi ugotovili, pri kolikšni količini šuma metoda SWT odpove, smo z iteracijami nadaljevali in opazovali njeno obnašanje. Rezultati so prikazani na sliki 5.13. Pri dvajseti iteraciji (slika 5.13b) začne detektor robov zaradi prevelike količine šuma detektirati napačne slikovne elemente, vendar metoda SWT še vedno pravilno detektira besedo "Segafredo". Pri trideseti in štirideseti iteraciji (slika 5.12c in 5.12d) pa je vpliv šuma že tako velik, da metoda SWT črk ne detektira več pravilno.

Slika 5.14 prikazuje še en primer testa šuma. Tudi v tem primeru metoda SWT deluje dovolj stabilno, medtem ko metoda TOCR zaradi prisotnosti šuma generira

Slika 5.11

Odvisnost delovanja metod od količine šuma (prvi primer). Pomen posameznih slik in režim označevanja sta enaka kot na sliki 5.9, le da se v vsaki iteraciji namesto ostrine spreminja količina šuma. (a) Originalna slika pred prvo iteracijo. (b) Prva iteracija. (c) Druga iteracija. (d) Tretja iteracija. (e) Četrta iteracija.

rahlo bolj zrnate regije. Barvna redukcija na podlagi glasovanja SWT uspe blokirati oranžno barvo dela napisa "sh.si" do druge iteracije (slika 5.14c). V tretji iteraciji (slika 5.14d) oranžna barva zaradi razraščanja barvnega histograma konvergira proti svetlo oranžni. Delovanje metode SWT pri dodatnem zašumljanju je prikazano na sliki 5.15. Podobno kot na sliki 5.13 se težave metode SWT začnejo pojavljati po trideseti iteraciji (slika 5.15c).



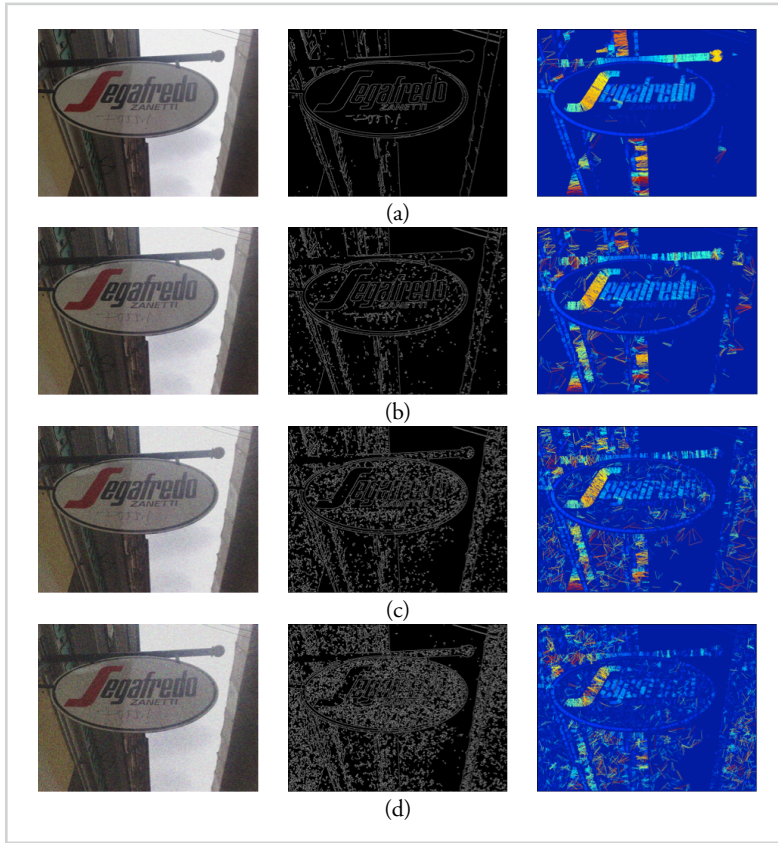
Slika 5.12

Razraščanje barvnega histograma. Barvni histogrami od (a) do (e) ustrezajo vhodnim slikam na sliki 5.11.

5.5 Diskusija

V poglavju smo podali rezultate evalvacije metode detekcije smeri SWT in metode barvne redukcije na podlagi glasovanja SWT. Obe metodi smo evalvirali na zbirki CVL OCR BIN DB.

Rezultati evalvacije metode detekcije smeri SWT potrjujejo dejstvo, da obstajajo statistične zakonitosti, na podlagi katerih je možno pravilno detektirati smer SWT. Metoda deluje presenetljivo dobro in z uporabo mere f in zgornje meje SWT dosega skoraj 92% stopnjo detekcije. Kljub primerom, ko metoda smeri SWT ne detektira pravilno, to na fazo glasovanja SWT ne vpliva drastično. Tipično so smeri SWT nepravilno detektirane le na manjših delih teksta, zato pravilno detektirani deli še vedno



Slika 5.13

Vpliv količine šuma na delovanje metode SWT (prvi primer). (a) 10. iteracija. (b) 20. iteracija. (c) 30. iteracija. (d) 40. iteracija.

vsebujejo dovolj informacije SWT, ki pravilno usmerja postopek barvne redukcije.

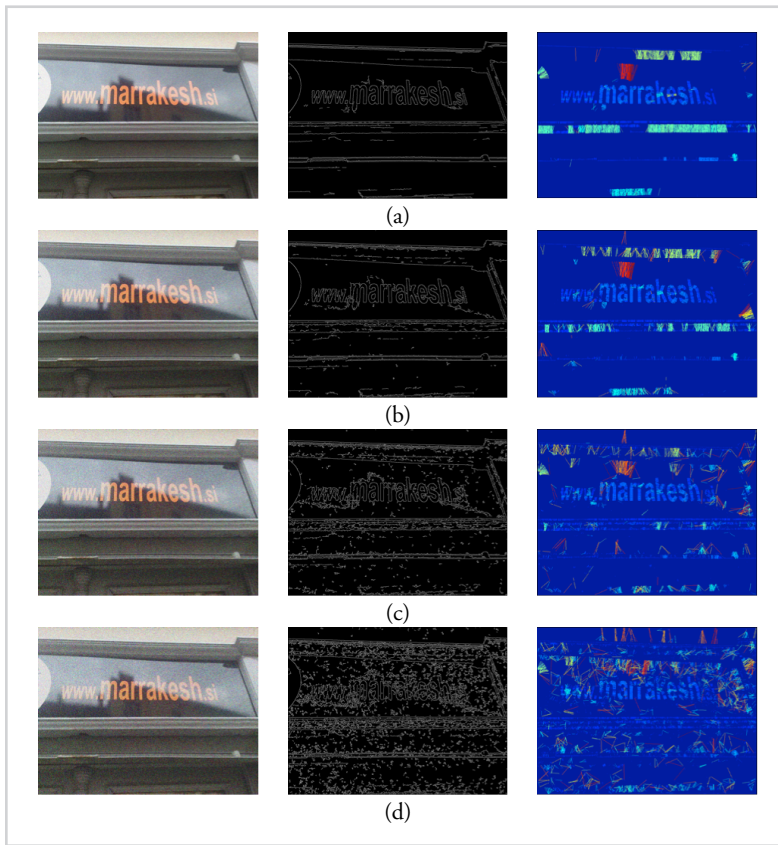
Barvna redukcija na podlagi glasovanja SWT uspešno združuje strukturno in barvno informacijo teksta. Z integracijo značilk SWT predlagana metoda nadzira postopek barvne redukcije in potencialnim barvam teksta ne dovoli, da bi konvergirale proti bolj dominantnim barvam ozadja. Metoda dosega do 2,89% višjo absolutno stopnjo detekcije in do 3,10% višjo povprečno stopnjo detekcije kot metoda TOCR [48].

Dodatno testiranje metod glede na kakovost vhodnih slik je pokazalo, da je metoda SWT robustnejša od metode TOCR, kadar je v sliki prisoten šum. Metoda postane

Slika 5.14

Odvisnost delovanja metod od količine šuma (drugi primer). Pomen posameznih slik in režim označevanja sta enaka kot na sliki 5.11. (a) Originalna slika pred prvo iteracijo. (b) Prva iteracija. (c) Druga iteracija. (d) Tretja iteracija. (e) Četrta iteracija.

neuporabna šele takrat, ko količina šuma močno presega nivoje, ki so tipični za slike, zajete s povprečno opremo za zajem slik. Po drugi strani pa ima metoda SWT za razliko od metode TOCR probleme pri neostrih slikah, saj je močno odvisna od kakovosti detekcije robov. Na podlagi rezultatov lahko sklepamo, da je metoda barvne redukcije na podlagi glasovanja SWT robustnejša od metode TOCR. Kljub pomanjkanju ostrine oziroma prisotnemu šumu večinoma pravilno deluje večje število iteracij kot metoda TOCR. Seveda pa v splošnem velja, da ko začne pešati metoda SWT, se to odraža tudi na nepravilnem delovanju glasovanja SWT.



Slika 5.15

Vpliv količine šuma na delovanje metode SWT (drugi primer). (a) 10. iteracija. (b) 20. iteracija. (c) 30. iteracija. (d) 40. iteracija.

6

Zaključek

V doktorski nalogi smo poleg pregleda trenutnega stanja na področju detekcije teksta v slikah naravnih scen opisali predlagani metodi barvne redukcije na podlagi glasovanja SWT in detekcije smeri SWT. Metodi smo zasnovali, implementirali, testirali in evalvirali na zbirki CVL OCR BIN DB.

Po natančni in izčrpni analizi področja smo v ožji izbor uvrstili najbolj obetavne metode in identificirali njihove prednosti ter slabosti. Za razliko od večine ostalih metod smo se pri zasnovi lastne metode osredotočili na integracijo tako strukturne kot barvne informacije v postopek segmentacije teksta.

Brez dodatnih informacij metoda TOCR vse barve v sliki obravnava enakovredno, ne glede na to, kakšni geometrijski strukturi pripadajo. Zato v metodi barvne redukcije na podlagi glasovanja SWT postopek barvne redukcije usmerjamo z dodatno strukturno informacijo SWT. Barve, ki so bogate s slikovnimi elementi SWT, najverjetneje pripadajo tekstu, zato jim metoda ne dovoli, da bi konvergirale proti barvam ozadja. Da bi lahko metodo TOCR in metodo barvne redukcije na podlagi glasovanja SWT neposredno in objektivno primerjali, smo uporabili binarno anotirane slike zbirke CVL OCR DB, ki omogočajo evalvacijo na nivoju posameznih črk. Ostale zbirke, kot sta npr. ICDAR in SVT, zaradi anotacije s pravokotniki niso primerne za takšen tip evalvacije.

S problemom takojšnje detekcije smeri SWT (tj. detekcije na podlagi gradientne in protigradientne slike SWT), ki je nujno potrebna za pravilno delovanje glasovanja SWT, se v literaturi po naših podatkih ni ukvarjal še nihče. Metoda detekcije smeri SWT, ki smo jo opisali, na podlagi analize histogramov posameznih blokov slik SWT z relativno visoko stopnjo natančnosti detektira pravilne smeri iskanja.

6.1 Sklepi

V tem poglavju povzemamo glavne sklepe posameznih poglavij.

Evalvacijske zbirke (poglavje 2):

- Pomanjkljivost zbirk ICDAR 2003, ICDAR 2011 in SVT je anotacija s pravokotniki, ki močno otežuje anotacijo nehorizontalnih tekstov. Ker anotirani pravokotniki ustrezajo posameznim besedam, mora metoda detekcije teksta sama poskrbeti za pravilno grupiranje teksta v besede, kar zmanjša objektivnost evalvacije metod detekcije teksta.

- Zbirka ICDAR 2003 je na določenih mestih nenatančno in nekonsistentno anotirana. Zbirka ICDAR 2011 ta problem odpravlja.
- Evalvacijska shema ICDAR 2003 primere *ena proti mnogo* in *mnogo proti ena* obravnava kot napačne, medtem ko shema ICDAR 2011 takšne primere pravilno detektira, vendar jih še vedno kaznuje.
- Problem zbirke SVT je anotacija le določenega dela besed v sliki, tj. besed, ki so v pripadajočem leksikonu.
- Določene pomanjkljivosti zbirk ICDAR 2003, ICDAR 2011 in SVT odpravlja zbirka CVL OCR DB. Anotacija z n-poligoni omogoča anotacijo nehorizontalnih tekstov. Binarna anotacija odpravlja tako problem nehorizontalnih tekstov kot problem grupiranja teksta v besede, saj omogoča evalvacijo na nivoju posameznih črk.
- Binarna evalvacijska shema CVL OCR DB deluje na nivoju povezanih komponent in uporablja dve meri natančnosti: absolutno stopnjo detekcije in povprečno stopnjo detekcije.

Obstoječe metode detekcije teksta (poglavje 3):

- Metode detekcije teksta se delijo na regijske in teksturne. Obstajajo tudi hibridne metode, ki izkoriščajo prednosti obeh.
- Metode se dodatno delijo na omejitveno in učno usmerjene. Omejitveno usmerjene metode temeljijo na uporabi množice geometrijskih pravil, medtem ko učno usmerjene temeljijo na uporabi klasifikatorjev.
- Rangiranje metod glede na ICDAR 2003 je problematično, saj so avtorji pri poročanju rezultatov metod nekonsistentni (uporaba različnih delov zbirke, različno računanje mer p in r ter podobno). Zbirki SVT in ICDAR 2011 sta relativno novi in ju za evalvacijo uporablja mnogo manj metod.
- Teksturne metode v zadnjem času niso pogoste. Tipično so časovno kompleksnejše od regijskih metod, prav tako niso primerne za detekcijo nehorizontalnih tekstov.

- Večina novejših metod temelji na regijskem pristopu, ki je časovno ugodnejši in bolj fleksibilen (ni omejen na horizontalne tekste).
- Metoda SWT temelji na predpostavki o paralelnosti robov črke. Deluje tako, da slikovnim elementom priredi debelino paralelne strukture, na kateri ležijo. Metoda je robustna, vendar pogosto spušča dele črk, ki nimajo popolnoma paralelnih robov.
- Metoda TOCR originalno število barv v sliki, ki lahko sega tudi do nekaj sto tisoč, reducira na N dominantnih barv, pri čemer je N tipično manjše od 10. Z detekcijo parcialnih delov črk nima težav, dogaja pa se, da spušča nedominantne barve teksta.
- Hibridne metode so tipično kompleksnejše za implementacijo.

Predlagana metoda (poglavje 4):

- Metoda barvne redukcije na podlagi glasovanja SWT postopek barvne redukcije usmerja z informacijo SWT. Barve, ki so bogate s slikovnimi elementi SWT, so blokirane in ne konvergirajo proti barvam ozadja.
- Preslikavo iz prostora SWT v barvni prostor omogoča vpogledna tabela SWT.
- Strukturne značilke, ki se uporabljajo v procesu glasovanja SWT, so: gostota SWT, standardna deviacija vrednosti SWT in standardna deviacija odmikov slikovnih elementov SWT od središča kocke SWT.
- Metoda detekcije smeri SWT za določitev prave smeri iskanja uporablja sliki SWT^+ in SWT^-. Sliki z uporabo profilov SWT razbije na bloke in analizira njihove histograme. Za primerjavo histogramov se uporabljata meri f in g.
- Uporaba zgornje meje SWT močno izboljša delovanje metode detekcije smeri SWT.

Eksperimentalni rezultati (poglavje 5):

- Metoda detekcije smeri SWT z uporabo mere f in zgornje meje SWT na zbirki CVL OCR BIN DB dosega skoraj 92% stopnjo detekcije. Metoda pravilno detektira smeri SWT tudi, kadar slika vsebuje temne tekste na svetli podlagi in obratno.

- Metoda barvne redukcije na podlagi glasovanja SWT na zbirki CVL OCR BIN DB dosega do 3,10% višjo stopnjo detekcije od metode TOCR.
- Metoda SWT je robustnejša na šum kot metoda TOCR.
- Metoda TOCR je robustnejša na zamegljevanje kot metoda SWT.
- Metoda barvne redukcije na podlagi glasovanja SWT je robustnejša od metode TOCR. V primeru manjše količine šuma oziroma manjše zameglitve slike še vedno deluje pravilno. Pri večjih količinah šuma oziroma večji zameglitvi slike je robustnost metode odvisna od robustnosti metode SWT.

6.2 Prispevki k znanosti

Znanstveni prispevki doktorske disertacije so naslednji:

- Predlagana metoda barvne redukcije na podlagi glasovanja SWT dosega višjo stopnjo detekcije od metode TOCR [48]. Na metodi TOCR temelji metoda strukturne particije in grupiranja [34], ki trenutno dosega enega najboljših rezultatov (glede na rangiranje ICDAR 2003). Koncept usmerjanja barvne redukcije s strukturno informacijo v obliki SWT je novost in odpira možnosti za nadgradnje ter izboljšave.
- Literatura problema detekcije smeri SWT ne omenja eksplicitno, zato predlagana metoda detekcije smeri SWT predstavlja pomemben doprinos na področju detekcije teksta. Metoda dosega visoko stopnjo detekcije na slikah zbirke CVL OCR BIN DB.
- Koncept histogramov SWT, ki jih uporabljamo pri detekciji smeri SWT, predhodno še ni bil uporabljen, zato predstavlja novost. Histogrami SWT imajo širšo uporabno vrednosti in niso omejeni le na detekcijo smeri SWT. Glede na to, da je z njimi možno uspešno razlikovati med tekstovnimi in netekstovnimi bloki, jih je možno uporabiti kot dodatne značilke pri filtriranju netekstovnih regij.
- Javno dostopna zbirka slik teksta v naravnih scenah CVL OCR DB, ki smo jo postavili za potrebe evalvacije, za razliko od ostalih javno dostopnih zbirk namesto anotacije s pravokotniki uporablja anotacijo z n-poligoni in binarno anotacijo. Binarna anotacija zbirke CVL OCR DB odpravlja potrebo po pravilnem

grupiranju detektiranega teksta v besede. Omogoča natančno in bolj objektivno evalvacijo metod detekcije teksta na nivoju posameznih črk.

- Ne nazadnje, doktorska naloga predstavlja pregled področja detekcije teksta v slikah naravnih scen.

6.3 Nadaljnje delo

Dosedanje raziskovalno delo odpira precej možnosti za nadaljnji razvoj. Metoda barvne redukcije na podlagi glasovanja SWT določeno barvo blokira, ko so izpolnjeni pogoji v enačbi (4.2). Smiselna nadgradnja bi bila uporaba klasifikatorja (npr. SVM), ki bi odločal, kdaj določeno barvo blokirati in kdaj ne. Podobno bi bila smiselna uporaba klasifikatorja v primeru detekcije smeri SWT. Namesto odločanja na osnovi mer f in g bi lahko uporabili klasifikator, ki bi na podlagi značilk histograma odločal, kateri izmed obeh blokov pripada dejanskemu tekstu.

Metodo detekcije smeri SWT bi bilo možno dodatno izpopolniti z izboljšavo razbitja slike. Nepravilna detekcija smeri je namreč najpogosteje posledica slabega razbitja. Kadar tekst pokriva le manjši del bloka, je vpliv šumnatih struktur v bloku prevelik, da bi detekcija delovala pravilno. Namesto na bloke bi lahko sliko razbili na regije poljubnih oblik okoli posameznih povezanih komponent. Na takšen način bi izločili večje število primerov, ko v bloku nastopajo tako tekstovne kot netekstovne strukture, in s tem zmanjšali vpliv šumnatih struktur.

Slikovni elementi SWT, ki ne ustrezajo tekstu, drastično ne vplivajo na postopek glasovanja SWT. Kljub temu bi lahko njihov vpliv še dodatno omejili. Z uporabo t. i. map zaupanja SWT (angl. *SWT confidence maps*) bi se omejili samo na slikovne elemente SWT, ki ležijo na območjih z dovolj veliko mero zaupanja SWT. Mera zaupanja bi se izračunala na podlagi lastnosti slikovnih elementov v okolici posameznega slikovnega elementa SWT. Histogrami SWT so se izkazali za zelo uporabne pri razlikovanju med tekstovnimi in netekstovnimi bloki, zato bi pri izgradnji map zaupanja SWT lahko služili kot dodatna značilka.

Namesto glasovanja SWT bi lahko postopek barvne redukcije, ki poteka v treh dimenzijah (R, G in B), razširili s četrto dimenzijo, tj. dimenzijo SWT. Seveda rešitev ni trivialna, saj dimenzija SWT ni enakovredna ostalim trem dimenzijam. Kljub temu bi z ustrezno transformacijo prostor iskanja morda lahko razširili z dodatno dimenzijo.

Ne nazadnje, zelo zanimivi bi bili binarizacija zbirke ICDAR in evalvacija metode barvne redukcije na podlagi glasovanja SWT na njej.

LITERATURA

[1] N. Stamatopoulos, B. Gatos, and S. J. Perantonis. A method for combining complementary techniques for document image segmentation. *Pattern Recognition*, 42 (12):3158–3168, 2009.

[2] F. Chang, S.-Y. Chu, and C.-Y. Chen. Chinese document layout analysis using an adaptive regrouping strategy. *Pattern Recognition*, 38(2):261–271, 2005.

[3] Y. Li, Y. Zheng, D. Doermann, S. Jaeger, and Y. Li. Script-idependent text line segmentation in freestyle handwritten documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1313–1329, 2008. doi: 10.1109/TPAMI.2007.70792.

[4] Y.-L. Chen and B.-F. Wu. A multi-plane approach for text segmentation of complex document images. *Pattern Recognition*, 42(7):1419–1444, 2009. doi: 10.1016/j.patcog.2008.10.032.

[5] K. Zagoris, S. A. Chatzichristofis, and N. Papamarkos. Text localization using standard deviation analysis of structure elements and support vector machines. *EURASIP Journal on Advances in Signal Processing*, 2011 (1):47, 2011.

[6] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi. Text extraction and document image sgmentation using matched wavelets and mrf model. *Image Processing, IEEE Transactions on*, 16(8): 2117–2128, 2007. doi: 10.1109/TIP.2007.900098.

[7] W. Kim and C. Kim. A new approach for overlay text detection and extraction from complex video scene. *IEEE Transactions on Image Processing*, 18(2):401–411, 2009. doi: 10.1109/TIP.2008.2008225.

[8] P. Shivakumara, W. Huang, and C. L. Tan. Efficient video text detection using edge features. In *Proc. of 19th International Conference on Pattern Recognition*, pages 1–4, 2008. doi: 10.1109/ICPR.2008.4761415.

[9] P. Shivakumara, T. Q. Phan, and C. L. Tan. Video text detection based on filters and edge features. In *Proc. of IEEE International Conference on Multimedia and Expo*, pages 514–517, 2009. doi: 10.1109/ICME.2009.5202546.

[10] D. Chen, K. Shearer, and H. Bourlard. Text enhancement with asymmetric filter for video ocr. In *Proc. of International Conference on Image Analysis and Processing*, pages 192–197, 2001. doi: 10.1109/ICIAP.2001.957007.

[11] Q. Ye, W. Gao, W. Wang, and W. Zeng. A robust text detection algorithm in images and video frames. In *Proc. of the Joint Conference of ICICS and PCM*, volume 2, pages 802–806, 2003. doi: 10.1109/ICICS.2003.1292567.

[12] Y. Hao, Z. Yi, H. Z. Guang, and T. Min. Automatic text detection in video frames based on bootstrap artificial neural network and ced. In *Proc. of International Conference on Computer Graphics, Visualization and Computer Vision*, 2003.

[13] C. Liu, C. Wang, and R. Dai. Text detection in images based on unsupervised classification of edge-based features. In *Proc. of 8th International Conference on Document Analysis and Recognition*, pages 610–614, 2005. doi: 10.1109/ICDAR.2005.228.

[14] X. Zhao, K.-H. Lin, Y. Fu, Y. Hu, Y. Liu, and T. S. Huang. Text from corners: A novel approach to detect text and caption in videos. *IEEE Transactions on Image Processing*, 20(3):790–799, 2011. doi: 10.1109/TIP.2010.2068553.

[15] P. Shivakumara and C. L. Tan. Novel edge features for text frame classification in video. In *Proc. of 20th International Conference on Pattern Recognition*, pages 3191–3194, 2010. doi: 10.1109/ICPR.2010.781.

[16] X. Huang and H. Ma. Automatic detection and localization of natural scene text in video. In *Proc. of 20th International Conference on Pattern Recognition*, pages 3216–3219, 2010. doi: 10.1109/ICPR.2010.786.

[17] X. Li, W. Wang, S. Jiang, Q.Huang, and W. Gao. Fast and effective text detection. In *Proc. of 15th IEEE International Conference on Image Processing*, pages 969–972, 2008. doi: 10.1109/ICIP.2008.4711918.

[18] M. Anthimopoulos, B. Gatos, and I. Pratikakis. A two-stage scheme for text detection in video images. *Image and Vision Computing*, 28(9):1413–1426, 2010. doi: 10.1016/j.imavis.2010.03.004.

[19] P. Shivakumara, W. Huang, T. Q. Phan, and C. L. Tan. Accurate video text detection through classification of low and high contrast images. *Pattern Recognition*, 43(6):2165–2185, 2010. doi: http://dx.doi.org/10.1016/j.patcog.2010.01.009.

[20] P. Shivakumara, A. Dutta, T. Q. Phan, C. L. Tan, and U. Pal. A novel mutual nearest neighbor based symmetry for text frame classification in video. *Pattern Recognition*, 44(8):1671–1683, 2011. doi: http://dx.doi.org/10.1016/j.patcog.2011.02.008.

[21] M. R. Lyu, J. Song, and M. Cai. A comprehensive method for multilingual video text detection, localization and extraction. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(2):243–255, 2005. doi: 10.1109/TCSVT.2004.841653.

[22] N. Ezaki, M. Bulacu, and L. Schomaker. Text detection from natural scene images: towards a system for visually impaired persons. In *Proc. of International Conference on Pattern Recognition*, volume 2, pages 683–686, 2004. doi: 10.1109/ICPR.2004.1334351.

[23] N. Ezaki, K. Kiyota, B. T. Minh, M. Bulacu, and L. Schomaker. Improved text-detection methods for a camera-based text reading system for blind persons. In *Proc. of 8th International Conference on Document Analysis and Recognition*, pages 257–261, 2005. doi: 10.1109/ICDAR.2005.137.

[24] X. Chen, J. Yang, J. Zhang, and A. Waibel. Automatic detection and recognition of signs from natural scenes. *IEEE Transactions on Image Processing*, 13(1):87–99, 2004. doi: 10.1109/TIP.2003.819223.

[25] T. N. Dinh, J. Park, and G. Lee. Korean text detection and binarization in color signboards. In *Proc. of International Conference on Advanced Language Processing and Web Information Technology*, pages 235–240, 2008. doi: 10.1109/ALPIT.2008.41.

[26] J. Park, G. Lee, E. Kim, J. Lim, S. Kim, H.Yang, M. Lee, and S. Hwang. Automatic detection and recognition of korean text in outdoor signboard images. *Pattern Recognition Letters*, 31(12):1728–1739, 2010. doi: 10.1016/j.patrec.2010.05.024.

[27] K. Jung, K. I. Kim, and A. K. Jain. Text information extraction in images and video: a survey. *Pattern Recognition*, 37(5):977–997, 2004. doi: 10.1016/j.patcog.2003.10.012.

[28] J. Liang, D. Doermann, and H. Li. Camera-based analysis of text and documents: a survey. *International Journal of Document Analysis and Recognition*, 7(2-3): 84–104, 2005. doi: 10.1007/s10032-004-0138-z.

[29] J. Zhang and R. Kasturi. Extraction of text objects in video documents: Recent progress. In *8th IAPR International Workshop on Document Analysis Systems*, pages 5–17, 2008. doi: 10.1109/DAS.2008.49.

[30] Q. Ye, Q. Huang, W. Gao, and D. Zhao. Fast and robust text detection in images and video frames. *Image and Vision Computing*, 23(6):565–576, 2005. doi: 10.1016/j.imavis.2005.01.004.

[31] X. Chen and A. L. Yuille. Detecting and reading text in natural scenes. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 366–373, 2004. doi: 10.1109/CVPR.2004.1315187.

[32] J.-J. Lee, P.-H. Lee, S.-W. Lee, and A. Yuilleand C. Koch. Adaboost for text detection in ntural scene. In *Proc. of International Conference on Document Analysis and Recognition*, pages 429–434, 2011. doi: 10.1109/ICDAR.2011.93.

[33] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2963–2970, 2010. doi: 10.1109/CVPR.2010.5540041.

[34] C. Yi and Y. L. Tian. Text string detection from natural scenes by structure-based partition and grouping. *IEEE Transactions on Image Processing*, 20(9): 2594–2605, 2011. doi: 10.1109/TIP.2011.2126586.

[35] T. D. Nguyen, J. Park, and G. Lee. Tensor voting based text localization in natural scene images. *IEEE Signal Processing Letters*, 17(7):639–642, 2010. doi: 10.1109/LSP.2010.2049595.

[36] L. Neumann and J. Matas. Real-time scene text localization and recognition. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 3538–3545, 2012. doi: 10.1109/CVPR.2012.6248097.

[37] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. In *Proc. of 18th IEEE International Conference on Image Processing*, pages 2609–2612, 2011. doi: 10.1109/ICIP.2011.6116200.

[38] Y.-F. Pan, X. Hou, and C.-L. Liu. A hybrid approach to detect and localize texts in natural scene images. *IEEE Transactions on Image Processing*, 20(3):800–813, 2011. doi: 10.1109/TIP.2010.2070803.

[39] R. Minetto, N. Thome, M. Cord, J. Fabrizio, and B. Marcotegui. Snoopertext: A multiresolution system for text detection in complex visual scenes. In *Proc. of 17th IEEE International Conference on Image Processing*, pages 3861–3864, 2010. doi: 10.1109/ICIP.2010.5651761.

[40] R. Minetto, N. Thome, M. Cord, J. Stolfi, F. Precioso, J. Guyomard, and N. J. Leite. Text detection and recognition in urban scenes. In *Proc. of IEEE International Conference on Computer Vision*, pages 227–234, 2011. doi: 10.1109/ICCVW.2011.6130247.

[41] C. Jung, Q. Liu, and J. Kim. A stroke filter and its application to text localization. *Pattern Recognition Letters*, 30(2):114–122, 2009. doi: 10.1016/j.patrec.2008.05.014.

[42] A. Ekin. Local information based overlaid text detection by classifier fusion. In *Proc. of IEEE Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 2–2, 2006. doi: 10.1109/ICASSP.2006.1660452.

[43] D. Chen, J.-M. Odobez, and J.-P. Thiran. A localization/verification scheme for finding text in images and video frames based on contrast independent features and machine learning methods. *Signal Processing: Image Communication*, 19(3), 2004.

[44] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young. Icdar 2003 robust reading competitions. In *Proc. of 7th International Conference on Document Analysis and Recognition*, 2003.

[45] A. Shahab, F. Shafait, and A. Dengel. Icdar 2011 robust reading competition callenge 2: reading text in scene images. In *Proc. of International Conference on Document Analysis and Recognition*, pages 1491–1496, 2011. doi: 10.1109/ICDAR.2011.296.

[46] K. Wang, B. Babenko, and S. Belongie. End-to-end scene text recognition. In *Proc. of IEEE International Conference on Computer Vision*, pages 1457–1464, 2011. doi: 10.1109/ICCV.2011.6126402.

[47] A. Ikica and P. Peer. Cvl ocr db, an annotated image database of text in natural scenes, and its usability. *Info. MIDEM*, 41(2):150–154, 2011.

[48] N. Nikolaou and N. Papamarkos. Color reduction for complex document images. *International Journal of Imaging Systems and Technology*, 19(1):14–26, 2009. doi: 10.1002/ima.20174.

[49] A. Ikica and P. Peer. Swt voting-based color reduction for text detection in natural scene images. *EURASIP Journal on Advances in Signal Processing*, 2013(1):1–13, 2013. doi: 10.1186/1687-6180-2013-95.

[50] S. M. Lucas. Icdar 2005 text locating competition results. In *Proc. of 8th International Conference on Document Analysis and Recognition*, pages 80–84, 2005. doi: 10.1109/ICDAR.2005.231.

[51] C. Wolf and J.-M. Jolion. Object count/area graphs for the evaluation of object detection and segmentation algorithms. *International Journal of Document Analysis and Recognition*, 8(4):280–296, 2006. doi: 10.1007/s10032-006-0014-0.

[52] J. Liang, I. T. Phillips, and R. M. Haralick. Performance evaluation of document layout analysis algorithms on the uw data set. In *Proc. of SPIE, Document Recognition IV*, pages 149–160, 1997.

[53] C. Wolf and J.-M. Jolion. Extraction and recognition of artificial text in multimedia documents. *Pattern Analysis and Applications*, 6(4):309–326, 2003. doi: 10.1007/s10044-003-0197-7.

[54] C. Wolf. *Text detection in images taken from video sequences for semantic indexing*. PhD thesis, INSA de Lyon, 2003.

[55] L. Neumann and J. Matas. Text localization in real-world images using efficiently pruned exhaustive search. In *Proc. of International Conference on Document Analysis and Recognition*, pages 687–691, 2011. doi: 10.1109/ICDAR.2011.144.

[56] C. Yi and Y. Tian. Localizing text in scene images by boundary clustering, stroke segmentation, and string fragment classification. *IEEE Transactions on Image Processing*, 21(9):4256–4268, 2012. doi: 10.1109/TIP.2012.2199327.

[57] Y. Zhao, T. Lu, and W. Liao. A robust color-independent text detection method from complex videos. In *Proc. of International Conference on Document Analysis and Recognition*, pages 374–378, 2011. doi: 10.1109/ICDAR.2011.83.

[58] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986. doi: 10.1109/TPAMI.1986.4767851.

[59] J. Serra. *From pixels to features*, chapter Toggle mappings, pages 61–72. North-Holland, Elsevier, 1989.

[60] A. Ikica and P. Peer. Grupiranje teksta v slikah naravnih scen. In *Proc. of ROSUS*, pages 77–84, 2012.

[61] W. Niblack. *An Introduction to Digital Image Processing*. Prentice-Hall, 1986.

[62] R. B. Fisher. Change detection in color images, 1999. URL `http://homepages.inf.ed.ac.uk/rbf/PAPERS/iccv99.pdf`.

[63] A. Ikica and P. Peer. An improved edge profile based method for text detection in images of natural scenes. In *Proc. of IEEE Conference on Computer as a Tool (EUROCON)*, pages 1–4, 2011. doi: 10.1109/EUROCON.2011.5929289.