

PREPOZNAVA TEKAČEV NA FOTOGRAFIJAH

Miha Grgič Jelen, dr. Borut Batagelj

Fakulteta za računalništvo in informatiko

Univerza v Ljubljani

E-pošta: miha.grgic.jelen@gmail.com

POVZETEK *V članku je predstavljen postopek, ki smo ga razvili v sklopu diplomske naloge na fakulteti za računalništvo in informatiko, kot rešitev za problem identifikacije tekačev na fotografijah športnih prireditev. Postopek najprej izolira posamezne osebe, nato za vsako osebo definira tarčno regijo v kateri se najverjetneje nahaja štartna številka. V naslednjem koraku izvede segmentacijo tarčne regije in izolira posamezne številke. Nato vsak izsek v katerem se potencialno nahaja številka, poskušamo prepoznati z optičnem branjem znakov, z uporabo orodja TesseractOCR. Na koncu prepoznane številke dodatno filtriramo in združimo v štartno številko. Razvit je bil tudi uporabniški vmesnik, ki nam močno olajša pregledovanje fotografij in vizualizacijo rezultatov. Na koncu predstavimo nerešene probleme ter nekaj idej za nadaljnji razvoj.*

1. UVOD

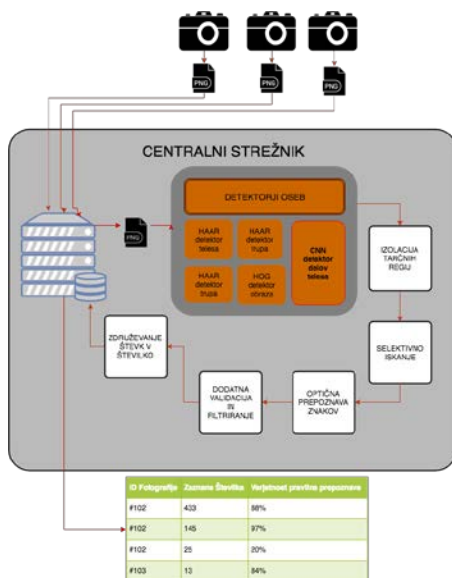
Kot študent sem več let delal kot fotograf tekaških prireditev v Sežani. Nastale fotografije so končale v obsežni spletni galeriji. Skoraj vsak udeleženec po prihodu domov pregleda svoje rezultate ter galerijo, v upanju da bo našel kakšno svojo fotografijo. V tem trenutku, pa je njegova vztrajnost na resnem preizkusu, saj je v spletni galeriji zelo veliko fotografij. V večini primerov z iskanjem odneha po nekaj desetih fotografijah.

Najprej smo pregledali, če takšna tehnološka rešitev že obstaja. Izkazalo se je, da je na internetu dostopnih ogromno aplikacij ter rešitev, namenjenih prepoznavi teksta ter klasifikaciji fotografij, kot so *TesseractOCR* [1] in *Google Vision* [3], vendar nobena od njih ni specializirana za prepoznavo štartnih števil. Še najbližje rešitve so tiste, ki služijo za programsko branje registrskih tablic avtomobilov, npr. knjižnica *OpenALPR* [2]. Vendar, ko smo jih poskusili prirediti in uporabiti na našem primeru, se je izkazalo, da je naš problem kompleksnejši in nismo prišli do uporabnih rezultatov. Tako je nastala ideja o razvoju prototipa za računalniško identifikacijo oseb na fotografijah, z namenom da bi udeleženci prireditev lahko enostavno dostopali do svojih fotografij.

Shema na sliki 1 prikazuje celoten proces prepoznave štartne številke. Fotografiji, ki so delovali na terenu, fotografije najprej pretočijo na centralni strežnik, kjer se shranijo v primerno strukturo. Nato na vsaki fotografiji iz zbirke izvedemo postopke, ki iz slike prepoznajo štartne številke.

Začnemo z izolacijo tarčnih regij (podrobneje opisano v sekciji 2.1), v katerih pričakujemo, da se nahajajo štartne številke. Za doseg tega najprej uporabimo detektorje različnih delov telesa. Na podlagi dobljenih rezultatov lahko nato izračunamo in izoliramo tarčne regije.

V naslednjem koraku se ukvarjamo z detekcijo posameznih števk, ki se nahajajo v tarčni regiji (podrobneje opisano v sekciji 2.2). Na vsaki izmed regij izvedemo algoritem selektivnega iskanja. Tako dobimo nove izseke slike, na kateri se potencialno nahaja posamezna številka. Na vsakem od njih izvedemo postopek optične prepoznave znakov. Na pridobljenih rezultatih izvedemo dodatno validacijo in filtriranje glede na verjetnost pravilne optične prepoznave, razmerja, lokacije in velikosti znaka. Na koncu rezultate združimo v številko.



Slika 1: Shema celotnega postopka detekcije štartnih številke

2. POSTOPEK DETEKCIJE ŠTARTNE ŠTEVILKE

2.1 Izolacija tarčne regije

Cilj v prvem koraku postopka je, da izoliramo regije, ki potencialno vsebujejo štartno številko. To lahko storimo na več načinov oz. uporabimo lahko različne detektorje. Nekateri od njih detektirajo obraz, drugi trup, tretji pa posamezne dele telesa. Na podlagi rezultatov omenjenih detektorjev, lahko izračunamo tarčno regijo, kjer je največja verjetnost, da se štartna številka nahaja. Na njej bomo nato izvajali nadaljnje operacije, ki nas bodo pripeljale do končnih rezultatov. Med preizkušenimi detektorji se je najbolje izkazal detektor, imenovan *Realtime Multi-Person 2D Pose Estimation using Part*

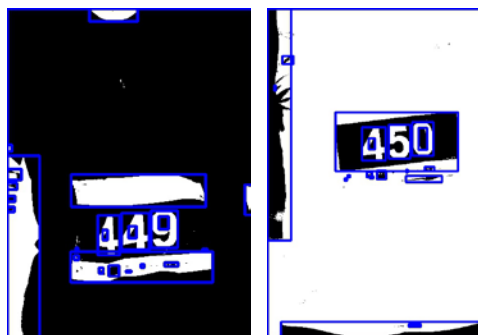
Affinity Fields [4]. Posebnost tega detektorja je, da v realnem času (ob uporabi GPU enote) detektira različne dele telesa (obraz, rame, boki, kolena, stopala...). V primerjavi z drugimi detektorji, kot je npr. detektor obraza, je pri uporabljenem detektorju tarčna regija veliko bolj natančno definirana in ne vsebuje delov drugih tekačev (slika 2).



Slika 2: Primer detekcije oseb in delov telesa z uporabo detektorja *Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*[4]. Skelet in točke, ki predstavljajo dele telesa so označene z rdečo in rumeno barvo, izolirana tarčna regija pa z zeleno.

2.2 Detekcija števk

V tem delu postopka, znotraj že definiranih tarčnih regij, najprej poiščemo t.i. kandidate, na katerih bomo kasneje izvedli algoritem, ki bo detektiral ali gre za števkico ali ne. Na binariziranem delu fotografije smo uporabili algoritem selektivnega iskanja [6]. Uporabili smo metodo *find_candidate_object_locations*, ki je del knjižnice *Dlib* [5]. Med rezultati ostanejo nekatere nepravilnosti. Lahko se bi jih znebili z uporabo *strožjega filtriranja*, vendar bi s tem tudi izpustili kakšno števkico. Na tej točki jih ignoriramo, saj bomo zanje poskrbeli kasneje v postopku. Tako so rezultati v veliki večini primerov res izolirane števkice (slika 3).



Slika 3: Primer rezultatov z uporabo algoritma *Selective Search for Object Recognition* [6]



Slika 4: Primer filtriranih rezultatov z uporabo algoritma *Selective Search for Object Recognition* [6]

Na predstavljenem primeru (slika 2) so na fotografiji številke bele barve, ozadje pa rdeče. Tako po izvedbi postopka binarizacije dobimo slike števk, ki imajo črno ozadje, sama številka pa je bele barve (slika 4). Na sliki, ki je rezultat binarizacije izvedemo inverz in dobimo posamezne številke zapisane črne na belem ozadju.

Na vsakem rezultatu algoritma *selektivno iskanje*, izvedemo optično prepoznavo znakov. To storimo z orodjem *TesseractOCR* [1]. Rezultati vsebujejo zaznano številko ter verjetnost pravilne detekcije.

Tabela 1: Način oblikovanja posameznih delov prispevka

<i>Detektirane številke</i>	<i>Verjetnost pravilne detekcije</i>
1	71%
0	93%
5	94%
4	90%

Na tej točki, imamo za vsako detekcijo dovolj podatkov, da lahko izvedemo *nekoliko pametnejše* filtriranje rezultatov. To storimo predvsem z razlikovanjem rezultatov, ki imajo visoko verjetnost pravilne detekcije (v nadaljevanju VVPD), v kombinaciji s preostalimi lastnostmi. V prvem koraku definiramo številke z VVPD (tabela 1). Med njimi so tudi anomalije, ki jih bomo poskušali odstraniti s pomočjo dodatnega filtriranja glede na razliko v višini, horizontalno oddaljenost od ostalih števk VVPD in vertikalno oddaljenost od števk VVPD.

2.3 Združevanje števk v številko

Na koncu preostale številke (rezultate, ki so prestali validacijo) le še združimo v celotno štartno številko, na podlagi horizontalnega položaja posameznega rezultata.

3. REZULTATI

Postopek smo preizkusili na različnih zbirkah fotografij, naključno izbranih iz različnih tekaških dogodkov v Sežani (*Mali kraški maraton*, *Tekaški pozdrav jeseni*). Na večini

izmed njih so 2 ali 3 osebe, na nekaterih je tudi 10 oseb, na drugih pa samo 1. Večina je posneta v ležeči kompoziciji, nekatere izmed njih pa v pokončni. V prvi zbirki (bele številke na rdečem ozadju), se na fotografijah pojavi 145 štartnih števil, metoda jih je pravilno prepoznala 85 kar pomeni 58,6% uspešnost (eng. *true positive rate*). V 53 primerih je metoda detektirala napačno številko (eng. *false positive*), izpustila pa je 60 štartnih števil, to je 41,4% (eng. *false negative*).

V drugi zbirki (črne številke na belem ozadju), se na fotografijah pojavi 78 štartnih števil, metoda jih je pravilno prepoznala 34, kar pomeni 43,5% uspešnost. V 41 primerih je metoda detektirala napačno številko, izpustila pa je 44 štartnih števil, to je 56,4%.

V trenutni fazi rezultati niso zadovoljivi. Tipične napake so:

- Detektor oseb ne detektira posamezne osebe, torej se za dano osebo tarčna regija sploh ne izračuna ter optična prepoznavna znakov sploh ne izvede,.
- Zaradi različnih barv tekaških majic pride do nepravilne binarizacije. Posledica tega je, da algoritem selektivnega iskanja ne izolira pravilno posameznih števk za kasnejšo prepoznavo.
- V primerih, ko je na sliki oseba delno zakrita (pred njo stoji druga oseba) niso vidne vse številke, ki sestavljajo štartno številko. Tako sistem prepozna samo nekaj izmed njih.

Za rešitev problema oseb, ki jih nismo našli, bi lahko uporabili kombinacijo več različnih detektorjev za definicijo tarčne regije. Detektor obraza, bi v primeru delno prekrita telesa še vedno zaznal osebo, v kolikor je na fotografiji viden obraz. Tako bi zmanjšali število manjkajočih detekcij.

Težave z binarizacijo bi lahko odpravili, če bi tarčno regijo zmanjšati na pravokotnik v katerem se nahaja štartna številka. Binarizacija po metodi *Otsui* [7] bi tako zelo uspešno binarizirala sliko in ločila številke od ozadja. Poskusili bi lahko tudi z drugimi algoritmi za izolacijo teksta iz ozadja [8] in naravnih scen [9].

4.1 NEREŠEN PROBLEM PREKRITIH ŠTEVK

Eden izmed težjih nerešenih problemov, so prekrite številke. Kot lahko vidimo na sliki 5 se velikokrat zgodi, da eno izmed števk, ki sestavlja štartno številko, prekriva drugi tekač ali del telesa. Štartna številka na sliki 5, bi bila prepoznana kot 37. V danem primeru je rezultat napačen, saj se sistem ne zaveda skrite številke, ki se skriva pod tekačevo roko.

Ena od možnih rešitev bi bili markerji. Vsak štartni listek bi vseboval 4 markerje - vsakega v enem od kotov štartne številke. Markerji so grafike, ki jih enostavno prepoznamo s programskim orodjem. V kolikor na prepoznani številki ne moremo prešteti vseh 4 markerjev, rezultat označimo kot neveljaven. Problem tega pristopa je, da bi najverjetneje zavrgli tudi veliko pravih rezultatov. Hkrati bi morali z organizatorjem uskladiti obliko štartne številke, da bi vsebovala markerje.

Naslednja možna rešitev bi bilo preverjanje razmerja stranic pravokotnika, ki zaobjema številke. V kolikor pravokotnik ne ustreza pravemu razmerju, rezultat zavržemo, saj je velika verjetnost da je vsaj ena izmed števk prekrita. Tak pristop daje dobre rezultate samo v primeru, da so številke na unikatnem ozadju (kot na primer rdeče ozadnje na sliki 5). V kolikor številke nimajo enakomernega oz. kontrastnega ozadja, pristop ne deluje več. Tudi v tem primeru bi bilo potrebno usklajevanje z organizatorjem glede oblike štartne številke.



Slika 5: Primer fotografije z delno prekrito štartno številko

5. ZAKLJUČEK

Predstavili smo problem identifikacije tekačev na fotografijah ter motivacijo za implementacijo rešitve. Ugotovili smo, da je najprimernejša metoda identifikacije osebe, da prepoznamo štartno številko, ki jo vsak tekač nosi na dresu. To smo dosegli tako, da smo najprej s pomočjo različnih detektorjev objektov iz fotografije izolirali posamezne osebe. Ugotovili smo, da se je najbolj izkazal detektor posameznih delov telesa, ki uporablja konvolucijske nevronske mreže. Na podlagi dobljenih rezultatov, smo zlahka definirali tarčno regijo v kateri se nahaja štartna številka. Nato smo na njej izvedli algoritem selektivnega iskanja, ki nam je definiral izseke slik, na katerih se morda nahaja posamezna številka. Na vsakem izseku smo s pomočjo orodja *TesseractOCR* izvedli optično detekcijo znakov. Vse rezultate smo dodatno validirali in filtrirali. Dobljene rezultate smo na koncu združili v štartno številko.

Razvili smo tudi sodoben uporabniški vmesnik, ki služi pregledovanju fotografij in vizualizaciji rezultatov. Z njegovo pomočjo smo algoritme in različne korake postopka veliko lažje prilagajali, saj smo lahko na pregleden način hitro opazili razlike ter shranjevali vmesne rezultate.

Problem identifikacije tekačev, nam je predstavljala velik izziv. Vseskozi smo imeli v mislih kako bi razvili funkcionalen sistem in ne zgolj prototipa. Tekom razvoja smo se

veliko naučili, še ogromno pa ostaja odprtega. Želja je da bi sistem dodatno izpopolniti in preizkusiti v produkcijskem okolju.

6. LITERATURA

1. Tesseract Open Source OCR Engine (main repository), dostopno na: <https://github.com/tesseract-ocr/tesseract>.
2. OpenALPR Technology, Inc, Automatic License Plate Recognition library, dostopno na: <https://github.com/openalpr/openalpr>.
3. Google, Inc, Cloud Vision API, dostopno na: <https://cloud.google.com/vision/>.
4. Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," 2017 IEEE Conference on Computer Vision and Pattern Recognition, str. 1302–1310, 2017.
5. Davisking, A toolkit for making real world machine learning and data analysis applications in C++, dostopno na: <https://github.com/davisking/dlib>.
6. J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," International Journal of Computer Vision, vol. 104, št. 2, str. 154–171, 2013.
7. Chen Yu, Chen Dian-ren, Li Yang and Chen Lei, "Otsu's thresholding method based on gray level-gradient two-dimensional histogram," 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010), Wuhan, 2010, str. 282-285.
8. Neumann L., Matas J.: Real-Time Scene Text Localization and Recognition, CVPR 2012. dostopno na: <http://cmp.felk.cvut.cz/~neumalu1/neumann-cvpr2012.pdf>.
9. B. Epshtein, E. Ofek and Y. Wexler, "Detecting text in natural scenes with stroke width transform," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 2010, str. 2963-2970.