

# Vpliv registracije obraznih področij na učinkovitost samodejnega razpoznavanja obrazov: študija z OpenBR

Sebastjan Fabijan, Vitomir Štruc

Univerza v Ljubljani, Fakulteta za Elektrotehniko,

Tržaška cesta 25, 1000 Ljubljana, Slovenia

E-mail: sebastjan.fabijan@gmail.com, vitomir.struc@fe.uni-lj.si

## Povzetek

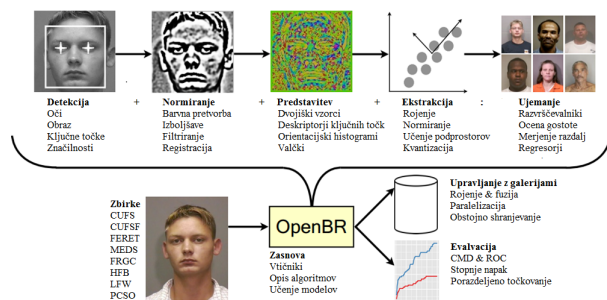
Razpoznavanje obrazov je v zadnjih letih postalo eno najuspešnejših področij samodejne, računalniško podprte analize slik, ki se lahko pohvali z različnimi primeri uporabe v praksi. Enega ključnih korakov za uspešno razpoznavanje predstavlja poravnava obrazov na slikah. S poravnavo poskušamo zagotoviti neodvisnost razpoznavanja od sprememb zornih kotov pri zajemu slike, ki v slikovne podatke vnašajo visoko stopnjo variabilnosti. V prispevku predstavimo tri postopke poravnavanja obrazov (iz literature) in proučimo njihov vpliv na uspešnost razpoznavanja s postopki, udejanjenimi v odprtokodnem programskem ogrodju Open Source Biometric Recognition (OpenBR). Vse poizkuse izvedemo na podatkovni zbirki Labeled Faces in the Wild (LFW).

## 1 Uvod

Razpoznavnik obrazov predstavlja vrsto sistema računalniškega vida, ki na podlagi slik obrazov izvede avtentikacijo ali identifikacijo oseb. Pri postopku avtentikacije poskuša sistem na podlagi vhodne slike uporabnika potrditi, če je oseba na sliki res ta, za katero se izdaja. Pri postopku identifikacije pa se obraz na vhodni sliki primerja z obrazi v določeni zbirki, kot rezultat pa se vrne bodisi identiteta osebe na sliki bodisi odgovor, da je oseba na vhodni sliki sistemu neznana. Najsodobnejši komercialni sistemi so zmožni primerjati slike obrazov s potnih listov, vozniških dovoljenj in osumljencev, katerih slike pridobiva policija, z natančnostjo, ki je že izredno blizu človeškim sposobnostim oz. človeške sposobnosti v določenih primerih že presega [1]. Do težav pa prihaja zaradi vplivov raznih dejavnikov, kot so osvetljenost okolice, razlike v pozah obrazov in zornem kotu, staranja in izrazih na obrazu. V tem prispevku se bomo osredotočili na problem spremenljivih zornih kotov. Predstavili in evalvirali bomo tri različne pristope k poravnavanju obrazov na slikah, ki pripomorejo k boljšemu razpoznavanju.

Glavni cilji prispevka so:

- predstaviti odprtokodno programsko ogrodje *Open Source Biometric Recognition* (OpenBR [2]) in proučiti njegovo učinkovitost pri razpoznavanju obrazov na podatkovni zbirki *Labeled Faces in the Wild* (LFW) [3],



Slika 1: Pregled funkcionalnosti OpenBR. Odprtokodno programsko ogrodje OpenBR vsebuje programske vmesnike za obdelavo slik obrazov in izvedbo različnih postopkov detekcije, normiranja, luščenja značilk, izračuna ujemanja, ipd. Vir slike: [2].

- predstaviti tri metode poravnavanja obrazov in preveriti njihov učinek na uspešnost razpoznavanja z odprtokodnim ogrodjem OpenBR, ter
- podati ugotovitve in sklepe v zvezi z ovrednotenimi postopki poravnave.

Preostanek prispevka je strukturiran v več razdelkov. V drugem razdelku na kratko predstavimo odprtokodno ogrodje OpenBR. V tretjem razdelku opišemo vse tri postopke poravnave, ki jih nato evalviramo v četrtem razdelku. Prispevek zaključimo s krajšim sklepom v petem razdelku.

## 2 OpenBR

Open Source Biometric Recognition (okrajšano OpenBR [2]) je odprtokodno programsko ogrodje, ki omogoča razpoznavanje obrazov ter oceno starosti in spola na podlagi slik oseb. OpenBR je bil razvit znotraj korporacije *The Mitre Corporation* zaradi potrebe po racionalizaciji postopka izdelave prototipov in vrednotenja algoritmov. Kot je ilustrirano na Sliki 1 vsebuje OpenBR funkcije za učenje modelov, upravljanje galerij, evalvacije algoritmov, prečna preverjanja in jedrnato sintakso za opis algoritmov [2]. Najpomembnejši dosežek projekta je jezik uporabljen za opis algoritmov. Vsaka beseda v algoritmu ustreza določenemu vtičniku, ki izvede določeno transformacijo na sliki, s čimer omogoča ločen razvoj posameznih korakov algoritma.

Dodatna posebnost OpenBR je učljiv algoritem z angleškim imenom *Spectrally Sampled Structural Subspace Features* [2] (okrajšano 4SF), ki za izvedbo razpoznavanja uporablja več diskriminativnih podprostorov. Sledi opis postopka razpoznavanja z algoritmom 4SF po korakih:

- **Detekcija obrazov:** Za detekcijo obrazov je uporabljen Viola-Jones detektor obrazov knjižnice OpenCV [2]. Za detekcijo oči se uporabljajo posebni korelacijski filtri (tj. ASEF filtri [4]), naučeni za zaznavanje lege oči znotraj detektiranih obraznih področij.
- **Normiranje/poravnava:** Obraze se registrira s pomočjo zaznanih lokacij obeh oči, na podlagi katerih se izvedeta rotacija in skaliranje. Velikost obraza na sliki po normiranju je  $128 \times 128$  slikovnih elementov. Po geometrijskem normiranju se izvede še popravljanje osvetlitve s pomočjo korekcijskega filtra, podrobneje predstavljene v [5].
- **Predstavitev:** Algoritem uporablja krajevne dvojiške vzorce (LBP) [6] in SIFT deskriptorje [7] vzorčene preko slikovnega področja obraza. Histogrami LBP so izluščeni na podlagi drsečega okna (angl. sliding window) velikosti  $8 \times 8$  slikovnih elementov s korakom drsenja 6 slikovnih elementov. 100 SIFT deskriptorjev je vzorčenih z mreže velikosti  $10 \times 10$  slikovnih elementov, z radijem deskriptorja 12 slikovnih elementov. Na izračunanih predstavitev se izvede še analiza glavnih komponent (PCA [8]), tako da se zadrži 95 odstotkov variance učnih podatkov. Projicirane predstavitve slike se na koncu normirajo še na enotsko dolžino v  $L_2$  [2].
- **Izračun značilnik in računanje ujemanja:** Kljub projekciji v PCA podprostor, je razsežnost učnih podatkov še zmeraj precejšnja. Za namene dodatnega zmanjšanja dimenzije se posamezni vektorji značilnik (v PCA podprostoru) iz učne množice podvzorčijo, tako da se ohrani zgolj 5 odstotkov značilnik vseh učnih PCA vektorjev. Na podvzorčenih vektorjih se nato izvede linearna diskriminantna analiza (LDA [8]), s katero se določi 12-dimenzionalne vektorje. Celotni postopek podvzorčenja se večkrat ponovi, s čimer se pridela večje število LDA podprostorov. Podobnost med dvema slikama se določi na podlagi vsote Evklidskih razdalj med vektorji v vsakem od LDA podprostorov.

Za dano testno sliko se vsi zgoraj predstavljeni koraki ponovijo; na sliki se detektira obraz, obrazno področje se poravna in svetlobno normira, izračuna se predstavitev obraznega področja, ki se nato projicira v vsakega od določenih LDA podprostorov, kjer se na koncu izvede izračun podobnosti/različnosti. Seveda je pri tem mogoče testne slike predhodno obdelati (npr. poravnati) in s tem zagotoviti višjo uspešnost algoritma 4SF.

Opisana različica algoritma je 1.0, ki je natančno dokumentirana v [2]. Za naše poizkuse smo uporabili različico 1.1, ki predstavlja nekoliko modificirano in nadgrajeno različico zgoraj predstavljenega algoritma, a trenutno še ni ustrezno dokumentirana. Glede na dosegljivo literaturo je različica algoritma 1.1 v osnovi

podobna algoritmu 1.0, razlikuje pa se v dejstvu, da v nadgrajeni različici kombinira holistične opise slik z opisi delov slik.

### 3 Poravnavanje obrazov

Cevovod razpoznavanja obrazov je v grobem sestavljen iz treh korakov: *i*) zaznavanja obrazov v slikah, *ii*) poravnave obraznih področij in *iii*) razvrščanja. Boljši kot je vsak od korakov, boljši bodo rezultati razpoznavanja. V našem prispevku se posvetimo študiju vpliva poravnave obrazov na uspešnost razpoznavanja.

Za namene študije, se omejimo na evalvacijo enega komercialnega postopka poravnavanja ter dveh postopkov poravnavanja obrazov, ki temeljita na tako-imenovanem slikovnem lijaku (angl. image funnel [9]). Slikovni lijak kot vhod sprejme grobo poravnano sliko obraza iz izbrane zbirke in kot rezultat vrne dobro poravnano sliko. Prednost tega pristopa je samodejno delovanje, pri katerem ni potrebno označiti delov objektov (obrazov) na slikah, niti ni potrebno identificirati začetne poze. Pomembno je le, da so objekti na začetku vsaj grobo poravnani [9]. Komericalna rešitev, na drugi strani, na vhodni sliki najprej detektira več značilnih točk obraza ter nato na sliki izvede perspektivno poravnano tako, da se detektirane točke na različnih slikah obrazov med seboj poravnajo.

#### 3.1 Slikovni lijak in strjevanje

Prvi postopek, ki nas bo v prispevku zanimal, je poravnava s slikovnim lijakom (angl. image funnel) in strjevanjem (angl. congealing) [9]. Pri tem postopku se nad izbrano učno množico slik, ki jo želimo poravnati, najprej izvede postopek strjevanja. Le-ta kot rezultat vrne slikovni lijak, ki ga je nato mogoče uporabiti za poravnano nevidenih slik z enakimi objekti (npr. obrazi).

Delovanje postopka strjevanja lahko opišemo z naslednjimi koraki: *i*) za dano zbirko učnih slik grobo poravnanih obrazov se za vsako lokacijo oz. slikovni element izračuna porazdelitveno polje, ki opisuje porazdelitev sivih nivojev (oz. značilnik) na tej lokaciji preko celotne učne množice slik, *ii*) poišče se afina transformacija vsake slike iz učne množice slik, tako da se zmanjša entropija porazdelitvenega polja oz. poveča verjetje, da je sliko generiralo porazdelitveno polje, *iii*) nato se ponovno izračuna porazdelitveno polje glede na transformirane slike in izvajanje se ponavlja do konvergence [9]. Omeniti velja, da celotni postopek temelji na SIFT značilnikah.

Rezultat strjevanja je zaporedje porazdelitvenih polj, ki mu v strokovni literaturi pravimo slikovni lijak. Z njim je mogoče poravnati dodatne, nevidene slike (npr. iz testne množice) tako, da v vsaki iteraciji na sliki izvedemo afino transformacijo v skladu s trenutno veljavnim porazdelitvenim poljem. Z opisanim postopkom lijakaste poravnave, se izognemo ponovni uporabi algoritma za strjevanje, saj lahko testne slike poravnamo brez neposredne uporabe učne množice. Ljakasta poravnava tako predstavlja učinkovitejši in hitrejši postopek poravnave dodatnih slik [9].

### 3.2 Globoka lijakasta poravnava

Drugi postopek, ki nas v naši študiji zanima, je postopek globoke lijakaste poravnave. Ideja poravnavanja je podobna kot pri postopku, predstavljenem v prejšnjem razdelku, pri čemer delovanje temelji na globokem učenju.

Tokrat se v fazi strjevanja ne uporabljajo hevristični deskriptorji kot je SIFT, ampak se značilke, ki služijo kot podlaga za oceno porazdelitvenega polja, naučijo neposredno iz učnih slik s pomočjo globokega učenja. Dodatno je algoritem za učenje modificiran tako, da vključuje t.i. kazen razpršenosti (angl. sparsity penalty), kar povzroči topografsko ureditev naučenih filtrov omrežja in izboljšanje nadaljnjih rezultatov poravnave [10].

Ostali koraki poravnave so enaki poravnavi s slikovnim lijakom: *i*) v učnem delu se s postopkom strjevanja izračuna globoki slikovni lijak, *ii*) nevidena slika se poravna s pomočjo afinih transformacij v skladu z lastnostmi globokega slikovnega lijaka.

### 3.3 Komerzialna rešitev

Poravnava temelji na lokalizaciji sedmih izhodiščnih točk na obrazu. Točke določajo pozicije obeh oči, ust in nosu. Sledi transformacija, ki lokalizirane točke preslika na vnaprej določene pozicije na sliki. Te točke so samodejno zaznane z uporabo detektorja, naučenega na podatkovni zbirki z ročno označenimi izhodiščnimi točkami [11].



Slika 2: Primerjava poravnave: (a) originalne slike iz zbirke LFW, (b) poravnava s slikovnim lijakom, (c) poravnava z globokim slikovnim lijakom, (d) poravnava s komercialno rešitvijo.

Slika 2 prikazuje primere slik štirih oseb iz štirih različic zbirke LFW. V stolpcu (a) so prikazane slike iz originalne zbirke LFW, v stolpcu (b) slike, poravnane s slikovnim lijakom, v stolpcu (c) slike, poravnane z globokim slikovnim lijakom in v stolpcu (d) slike, poravnane s komercialno rešitvijo.

## 4 Eksperimenti in rezultati

Eksperimente smo izvedli na štirih različicah podatkovne zbirke Labeled Faces in the Wild (v nadaljevanju LFW). LFW [3] je podatkovna zbirka slik oseb, zgrajena za proučevanje področja prepoznavanja obrazov. Vsebuje 13233 zbranih slik s spleta, 5749 različnih oseb. Vse slike so velikosti  $250 \times 250$  slikovnih točk. Slike zbirke so bile pridobljene samodejno s pomočjo zaznavanja z Viola-Jones obraznim detektorjem [12]. Poleg originalne zbirke LFW, smo za eksperimente uporabili še tri različice zbirke, na katerih so bile uporabljene metode, opisane v razdelkih 3.1, 3.2 in 3.3. Slike v zbirki *LFW (funneled)* [9] so poravnane s postopkom lijakaste poravnave, slike v zbirki *LFW (deep funneled)* [10] s postopkom globoke lijakaste poravnave, slike v zbirki *LFW-a* [11] pa s komercialno programsko opremo.

### 4.1 Eksperimentalni protokol

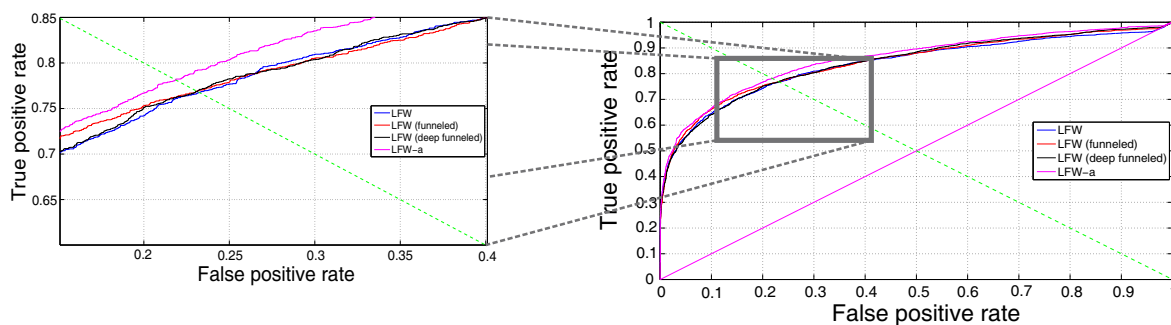
V eksperimentalnem delu primerjamo uspešnost razvrščanja z uporabo modela 4SF-Test, ki smo ga s pomočjo ogrodja OpenBR naučili sami z uporabo zbirk PUT [13] in CFD [14]. Kot merilo za primerjavo pri vsaki različici zbirke uporabimo 10-kratno prečno preverjanje. Pri  $k$ -kratnem prečnem preverjanju (ang.  $k$ -fold cross validation) zbirko podatkov naključno razdelimo na  $k$  podmnožic približno enakih velikosti. Vsako od podmnožic uporabimo enkrat za testiranje, preostale podmnožice pa za učenje. Ta postopek ponovimo  $k$ -krat. Ocena točnosti prečnega preverjanja je skupno število pravih razvrstitev, deljeno s številom preverjanj v podatkovni zbirki [15]. Vsako od desetih podmnožic v našem primeru sestavlja 600 primerjav (300 primerjav dveh slik iste osebe in 300 primerjav dveh slik različnih oseb).

Učinek poravnave preverjamo na problemu verifikacije obrazov. Rezultate zato podamo grafično v obliki ROC (angl. Receiver Operating Characteristics) krivulj [16, 17], ki ponazarjajo odnos med stopnjo pravilno potrjenih dogodkov (ang. True Positive Rate - TPR) in stopnjo napačno potrjenih dogodkov (ang. False Positive Rate - FPR) pri različnih vrednostih odločitvenega praga.

Kot kvantitativne mere podajamo površino pod ROC krivuljo (angl. Area Under the ROC Curve - AUC) ter vrednost napak na ROC krivulji kjer je  $1-TPR$  enak FPR, slednjo mero označimo kot ERR (angl. equal error rate). Vrednost površine pod krivuljo, AUC, je vedno med 0 in 1. Ker pa naključno ugibanje pri dvo-razrednih problemih kot je verifikacija izriše črto med točkama (0,0) in (1,1), pričakujemo za vse realne razvrščevalnike uspešnost razvrščanja nad 0.5. Vrednost AUC lahko tolmačimo kot verjetnost, s katero bo razvrščevalnik pravilno razvrstil naključen vzorec [16].

### 4.2 Rezultati

Rezultati poizkusov so v obliki ROC krivulj prikazani na Sliki 3 in v obliki kvantitavnih mer v Tabeli 1. Opazimo, da algoritem 4SF (ki smo ga naučili sami) doseže razmeroma visoko uspešnost razpoznavanja, saj AUC na



Slika 3: Uspešnost delovanja modela 4SF na podatkovni zbirki LFW, poravnani na več različnih načinov. Desna slika prikazuje celotno območje krivulje, leva povečan izsek v okolici operacijske točke EER.

Zbirka	AUC ( $\overline{AUC} \pm SE$ )	EER ( $\overline{EER} \pm SE$ )
LFW	$0.8367 \pm 0.0060$	$0.2323 \pm 0.0047$
LFW (funneled)	$0.8425 \pm 0.0058$	$0.2310 \pm 0.0057$
LFW (deep funneled)	$0.8419 \pm 0.0055$	$0.2313 \pm 0.0055$
LFW-a	$0.8573 \pm 0.0056$	$0.2187 \pm 0.0063$

Tabela 1: Tabela povprečnih vrednosti AUC in EER preko vseh 10-prečnih preverjanj. Poleg povprečnih vrednosti so podane tudi standardne napake (SE).

izvirni, neporavnani zbirki zavzame vrednost 0.8367. Nekoliko presenetljivi so rezultati na slikah, poravnanih s slikovnim lijakom oz. globokim slikovnim lijakom. V teh primerih je srednja vrednost AUC po vseh 10-tih prečnih preverjanjih nekoliko večja od vrednosti pri neporavnanih slikah, a je razlika v vrednostih AUC razmeroma majhna. Do podobnih ugotovitev pridemo tudi z analizo ROC krivulj na Sliki 3, kjer se krivulje, določene na podlagi neporavnanih slik in slik, poravnanih s slikovnim lijakom, v večjem delu prekrivajo. Največja razlika v vrednostih AUC je opazna pri uporabi slik, poravnanih s komercialno rešitvijo, kjer je razlika do vrednosti AUC, doseženi na neporavnanih slikah, okrog 2%. Opazimo lahko, da tudi v primeru, ko upoštevamo standardne napake, ni prekrivanja med ocenjenimi povprečnimi vrednostmi AUC in EER med tema zbirama slik.

## 5 Zaključek

V prispevku smo predstavili odprtokodno programsko ogrodje OpenBR in algoritem 4SF za razpoznavanje obrazov. Opisali in ovrednotili smo tri metode poravnavanja obrazov. Ugotovili smo, da je razpoznavanje na eni od zbirk, ki je predhodno poravnana, v povprečju uspešnejše za dve odstotni točki (v smislu AUC) v primerjavi z zbirko, ki vsebuje neporavnane slike. Za najuspešnejšo poravnavo se je izkazala poravnava z uporabo komercialne rešitve, ki temelji na lokalizaciji sedmih referenčnih točk na obrazu. Študija je bila izvedena zgolj z uporabo OpenBR. Kot nadaljnje raziskave bi bilo smiselno izvesti evalvacije še z uporabo drugih algoritmov za razpoznavanje obrazov, kar bi omogočilo natančnejšo določitev vpliva poravnave. Omeniti še velja, da je vpliv poravnave na zbirki LFW pričakovano omejen,

saj je bila zbirka pridobljena samodejno z uporabo detektorja obrazov Viote in Jonesa, ki je sposoben zaznati predvsem obraze v frontalni legi, pri odstopanjih od frontalne lege pa pogosto odpove.

## Literatura

- [1] P.J. Phillips, A.J. O'Toole. Comparison of human and computer performance across face recognition experiments. V *Image and Vision Computing*, 74–85, 2014.
- [2] J. C. Klontz, B. F. Klare in ostali. Open Source Biometric Recognition. V *IEEE Biometrics: Theory, Applications and Systems*, 2013.
- [3] G. B. Huang, M. Remesh in ostali. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49, 2007.
- [4] D.S. Bolme, B.A. Draper, and J.R. Beveridge. Average of Synthetic Exact Filters. V: *CVPR*, June 2009.
- [5] X. Tan in B. Triggs. Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. V *IEEE Transactions on Image Processing*, vol. 19, strani 1635-1650, 2010.
- [6] T. Ojala, M. Pietikäinen in T. Mäenpää. Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns. V: *LNCS*, vol. 1842, strani 404-420, 2000.
- [7] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. V *International Journal of Computer Vision*, vol. 60, strani 91-110, 2004.
- [8] A. M. Martínez in A. C. Kak. PCA versus LDA. V *IEEE TPAMI*, vol. 23, 2001.
- [9] G. B. Huang, V. Jain in E. Learned-Miller. Unsupervised Joint Alignment of Complex Images. *International Conference on Computer Vision (ICCV)*, 2007.
- [10] G. B. Huang, M. Mattar in ostali. Learning to Align from Scratch. *Advances in Neural Information Processing Systems (NIPS)*, 2012.
- [11] L. Wolf, T. Hassner in Y. Taigman. Effective Unconstrained Face Recognition by Combining Multiple Descriptors and Learned Background Statistics. V *IEEE TPAMI*, vol. 30, 1978-1990, 2011.
- [12] P. Viola in M. J. Jones. Robust Real-time Face Detection. *International Journal of Computer Vision*, 57(2):137-154, 2004.
- [13] A. Kasiński, A. Florek in A. Schmidt. The PUT Face Database. *Image Processing & Communications*, vol. 13, strani 59-64, 2008.
- [14] D. S. Ma, J. Correll in B. Wittenbrink. The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47, strani 1122-1135, 2015.
- [15] R. Kohavi. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence (IJCAI)*, 1995.
- [16] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27, strani 861-874, 2006.
- [17] R. Gajsek, V. Struc, S. Dobrišek, F. Mihelič. Emotion recognition using linear transformations in combination with video. *Interspeech*, 2009.