DOLOČANJE ODZIVA NA ZAUŽITO HRANO Z METODAMI ZA PREPOZNAVO OBRAZNEGA IZRAZA

Leon Roposa, Borut Batagelj, Franc Solina

Laboratorij za računalniški vid
Fakulteta za računalništvo in informatiko, Univerza v Ljubljani
leon.roposa@gmail.com, {borut.batagelj, franc.solina}@fri.uni-lj.si

POVZETEK: *V članku opisujemo metodo za prepoznavo obraznega izraza da bi določili odziv testirane osebe na okus zaužite hrane. Trenutno obstaja zgolj nekaj podobnih raziskav, kjer so z uporabo obstoječe programske opreme, ki zna na obrazih ločiti med šestimi osnovnimi čustvi, iskali korelacijo med oceno hrane, ki jo je podal preizkuševalec, in njegovo reakcijo izraženo na obrazu. V naši raziskavi smo izdelali svojo metodo za prepoznavo čustev na obrazih in jo preizkusili na lastni bazi posnetkov okušanja hrane. Problem smo definirali kot dvorazredni klasifikacijski, torej, ali lahko s pomočjo prepoznave obraznega izraza ugotovimo, ali je osebi okus zaužite hrane všeč ali ne. Ločeno smo obravnavali odzive na zaužito hrano in zaužito pijačo. Dosegli smo dobre rezultate pri analizi obraznih odzivov na hrano in zelo dobre rezultate pri analizi obraznih odzivov na pijačo.*

1. UVOD

Večina raziskav na področju prepoznave obraznih izrazov z računalniškim vidom, poteka na osnovi zaznavanja šestih osnovnih čustev: jeze, strahu, gnusa, presenečenja, veselja in žalosti. Teh šest čustev je v 70.-letih 20. stoletja psiholog Paul Ekman definiral kot univerzalna, kar pomeni, da so ta čustva prisotna v vseh kulturah in skozi celotno človeško zgodovino**.** V podatkovnih bazah in člankih se včasih pojavijo tudi še kakšna dodatna čustva. Recimo, ko ni zaznano nobeno od naštetih čustev, se včasih doda tudi nevtralni izraz. Poleg šestih osnovnih čustev se pri prepoznavi obraznih izrazov uporablja tudi dimenzija prijetnost/vzburjenje (ang. Valence/Arousal). V tem primeru prevedemo problem iz klasifikacijskega v regresijski, kjer vzburjenje pomeni stopnjo intenzitete čustva. Te metode je možno uporabiti tudi za povsem določene aplikacije, recimo za ugotavljanje bolečine iz obraznega izraza.

V raziskavah [1] so ugotovili da obstaja povezava med obraznim izrazom in okusom zaužite hrane, tako zaradi prijetnega okusa, kot intenzitete okusa. Obrazni izrazi so najbolj izraziti pri visokih koncentracijah neprijetnih okusov, kot sta kislo (izrazi na

ustnicah) in grenko (izrazi na očeh in čelu). Ugotovljeno je bilo tudi, da prijetnejši okusi sprožijo na obrazu najmanj opazne obrazne izraze.
Obstaja nekaj člankov kjer so z namenom ugotavljanja odzivov ljudi na zaužito hrano, uporabili metode za prepoznavo obraznega izraza [2, 3, 4, 5]. Avtorji člankov so sicer aktivni na področju raziskav povezanih s hrano, zato se članki ne osredotočajo na metode računalniška vida, temveč opisujejo kako so z uporabo obstoječe programske opreme analizirali odziv na zaužito hrano V vseh omenjenih člankih so za prepoznavo obraznih izrazov uporabili program FaceReader [6]. Program napove prisotnost enega izmed šestih osnovnih čustev in nevtralnega izraza. Za vsakega izmed šestih čustev in nevtralni izraz, nam pove tudi intenziteto čustva na obrazu.
Avtorji so v člankih primerjali kako so se spreminjale napovedi intenzitet čustev, kot jih je napovedal FaceReader. V člankih so uporabili različne vrste hrane in pijače. Pozitivni odziv na okus je imel največjo korelacijo ali z nevtralnim izrazom [3] ali pa z veseljem [4, 5]. Negativni odziv je imel korelacijo z gnusom, jezo in žalostjo [2, 3, 4, 5]. V [2] je imel negativni odziv na okus veliko večji vpliv na spremembo zaznanih intenzitet čustev. Nekateri izmed avtorjev [4, 5] so mnenja, da bi bile metode za prepoznavo obraznega izraza lahko primerne za določanje odziva na okus hrane. V člankih so tudi ugotovili, da je ob zaporednih okušanjih korelacija med obraznimi izrazi in prijetnostjo okusa vse manj izrazita [3] in da so opazne tudi razlike v intenzitetah med različnimi vrstami hrane [5].

2. METODA ZA PREPOZNAVANJE ČUSTEV NA OBRAZU

Našo metodo smo implementirali v C++, z uporabo knjižnice OpenCV in obsega: detekcijo obraza in lokacije oči s Haarovi značilkami, poravnavo obraza glede na zaznano lokacijo centra oči, lokalne binarne vzorce (ang. local binary pattern, LBP) za izločanje značilk, metodo glavnih komponent (ang. principal component analysis, PCA) za zmanjšanje dimenzionalnosti, pri čemer smo ohranili 95% variance in metodo podpornih vektorjev (ang. support vector machine, SVM) z uporabo RBF jedra in klasifikacijo po strategiji ”eden proti ostalim” Poravnava oči je bila izvedena z afino transformacijo tako, da sta centra oči na istih lokacijah. Uporabili smo LBP različico z uniformnimi vzorci, soseščino smo določili z 8 sosednimi elementi v radiju 1. Na sliki obraza smo izračunali histograme v mreži 10×10, tako da je končna dolžina vektorja značilk za sliko obraza znašala 5900. Pri klasifikaciji z metodo SVM smo za iskanje najprimernejših parametrov s prečnim preverjanjem uporabili metodo *auto_train* iz knjižnice OpenCV.

3. PREIZKUS METODE NA STANDARDNIH BAZAH

Predlagano metodo smo najprej preizkusili na dveh standardnih podatkovnih bazah, ki vsebujeta prikaz več različnih čustev: Cohn Kanade CK+ [7] in GEMEP-FERA 2011 [8].

3.1 Rezultati na bazi Cohn Kanade CK+

Baza Cohn Kanade CK+ vsebuje 327 posnetkov v katerih nastopa 118 oseb, od tega 82 ženskega in 36 moškega spola, osebe pripadajo različnim rasnim skupinam. Osebe

izražajo eno izmed sedmih čustev: jeza, žalost, veselje, strah, gnus, presenečenje in prezir. Posnetek je predstavljen kot zaporedje sličic in v njem nastopa zgolj ena oseba. Na začetku vsakega posnetka ima oseba na obrazu nevtralen izraz, ki nato postopoma preide v določeno čustvo, intenziteta izražanja čustva pa je največja na koncu posnetka. Osebe v bazi so posnete frontalno, v kontroliranem okolju z manjšimi premiki glave. Baza je zato za analizo manj zahtevna.

Baza nima točno določene testne in učne množice, zato smo ju določili sami. Za vsak posnetek smo izbrali tri sličice s konca posnetka, ki izražajo čustvo najbolj intenzivno. Tako sestavljene skupine treh slik smo naključno razdelili v učno in testno množico. V učni množici smo uporabili 876 posameznih sličic, v testni množici pa 84 sličic. Sličice iz posameznega posnetka so lahko zgolj v testni ali zgolj v učni množici. Izbrane slike, kjer naša metoda ni mogla zaznati oči, smo ročno zamenjali s slikami z zaznanimi očmi, ki nastopajo prej v posnetku in se na njih še zmeraj izraža čustvo. Težave z zaznavanjem oči smo imeli, ker osebe pri intenzivno izraženih čustvih pogosto imele oči zaprte. Na testni množici je naša metoda dosegla zelo dobre rezultate: 96.42% natančnost. Kar pomeni 81 pravilno napovedanih izrazov, pri 3 slikah pa ni bilo zaznano nobeno čustvo.

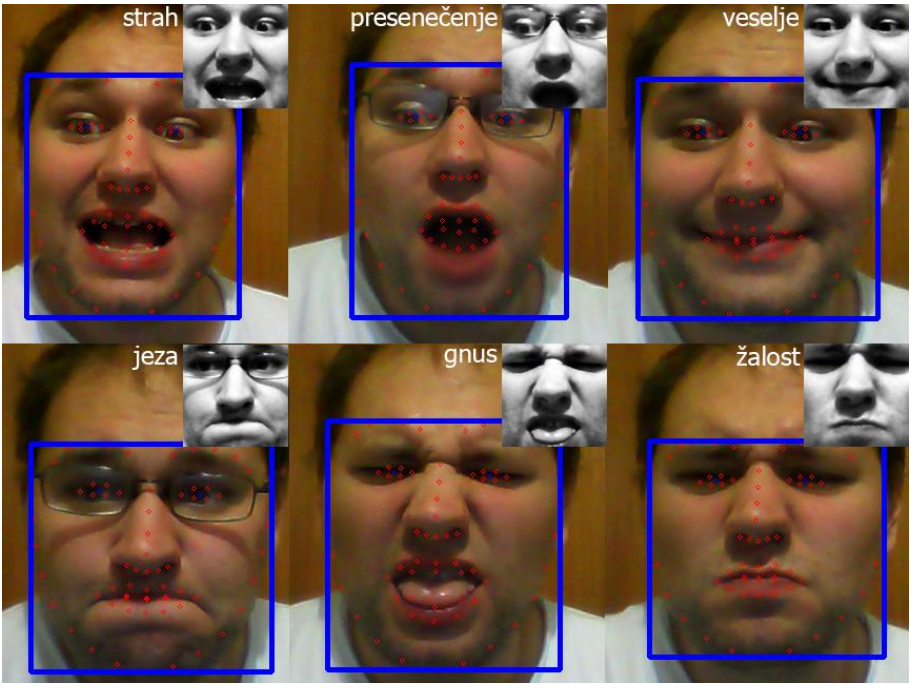
3.2 Rezultati na bazi GEMEP-FERA 2011

GEMEP-FERA je del baze GEMEP, ki se je uporabil na tekmovanju FERA-2011 (Facial Expression Recognition and Analysis challenge 2011) [8]. Baza GEMEP vsebuje posnetke 18 različnih čustev, katere so prispevali poklicni igralci, zato se posnetki smatrajo za spontane. Čustva v posnetkih so izražena z različno intenziteto, osebe v posnetkih se tudi veliko več premikajo, najbolj izrazito je premikanje glave. Zato se baza smatra za zahtevnejšo. V posnetkih vedno nastopa zgolj ena oseba, posnetki so zajeti v kontroliranem okolju. Baza GEMEP-FERA vsebuje 155 posnetkov v učni množici in 134 posnetkov v testni množici, v posnetkih nastopa skupno sedem različnih igralcev, obeh spolov in različnih starosti, v posnetkih je prisotnih 5 čustev (veselje, jeza strah, žalost in olajšanje). Posnetke smo pretvorili v zaporedje sličic, končna napoved za nek posnetek je bila čustvo z največjim številom napovedi po sličicah posnetka. Končni rezultat je torej število pravilno napovedanih posnetkov V 3 posnetkih testne množice ni bil nikoli zaznan obraz z obema očesoma, zato smo jih šteli kot nepravilne, sicer so bile končne slike obrazov dovolj natančne. Rezultat naše predlagane metode na tej bazi je: 79.85% natančnost, od tega 84.81% natančnost na osebah, ki so nastopile v učni množici in 72.72% natančnost na osebah, ki nastopajo samo v testni množici.

3.3 Preizkus s posnetki spletne kamere

Ker smo nameravali metodo določanja čustev na obrazu preizkusiti na lastnih posnetkih s spletno kamero, smo našo metodo preizkusili tudi na tovrstnih posnetkih, da bi ugotovili, ali je metoda dovolj natančna (Slika 1). Ugotovili smo, da je detekcija oči s Haarovi značilkami premalo natančna. Zato smo jo sprva nadomestili z omejenim lokalnim modelom (angl. constrained local model, CLM) [9], kasneje pa z detektorjem obraznih značilk iz knjižnice *dlib*, ki temelji na gradientemu boostingu [10]. Detektor obraznih značilk vsakič posebej zazna lokacije obraznih značilk na zaznani sliki obraza. Detekcija obraza v knjižnici *dlib* temelji na rabi histograma orientiranih gradientov (ang.

histogram of oriented gradients, HOG). Centra oči, ki se uporabita za poravnavo, smo določili kot povprečje napovedi obraznih značilk okoli očesa.



Slika 1: Prepoznavanje čustev na posnetkih spletne kamere. Za iskanje značilnic na obrazu je uporabljena metoda CLM [9]. Metoda za prepoznavo čustev je bila naučena na učni množici Cohn Kanade CK+. Zgoraj desno so normalizirane slike obrazov namenjene prepoznavi čustev.

4. PRIPRAVA PODATKOVNE BAZE

Ker nismo našli prosto dostopne baze posnetkov okušanja hrane smo jo pripravili sami. Naša baza vsebuje 135 posnetkov s petimi različnimi osebami. V posnetku vedno nastopa samo ena oseba. Tri osebe so moškega spola in dve ženskega, starost oseb v posnetkih se giblje od 28 do 71 let. Za vsako osebo je na voljo od 22 do 30 posnetkov. Snemali smo tako prizore okušanja hrane, kot okušanja pijače, ki smo jih kasneje ločeno testirali. Posnetke smo snemali pri enakomerni osvetlitvi, tako da je bil obraz okuševalcev frontalno dobro viden. Obraz okuševalca je bil v posnetku postavljen na sredino. Izbrano ozadje je bilo monotono brez motečih predmetov. Ker naš sistem ni sposoben avtomatsko ugotoviti, ali je obraz zakrit bodisi z roko, posodo ali hrano, smo zakrivanje obraza med okušanjem poskušali kar se da minimizirati.

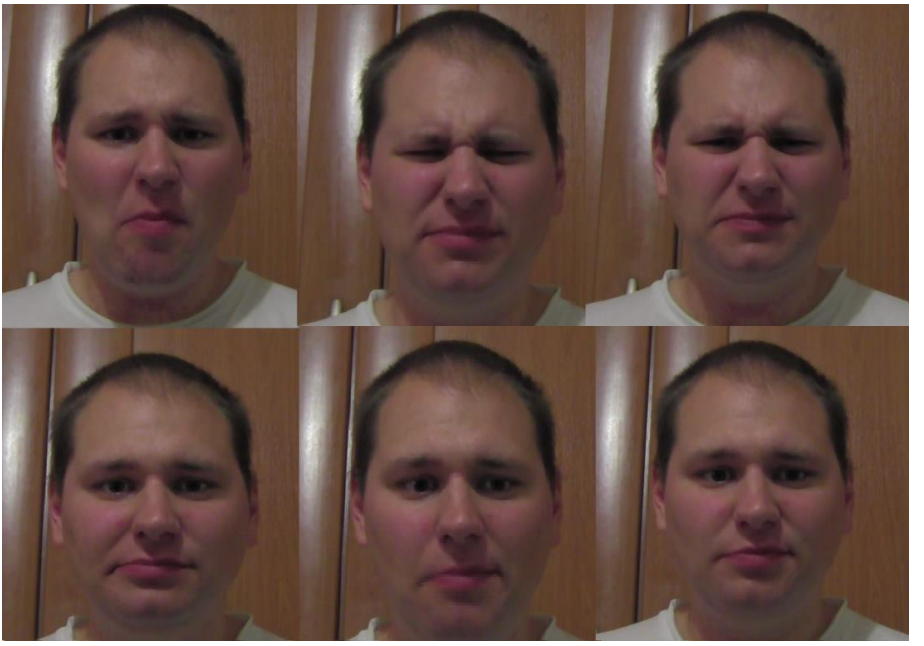
Vsak posnetek se začne pred okušanjem hrane oziroma pijače. Pri okušanju pijače so okuševalci pri pitju uporabili slamico zato, da bi bilo pri pitju čim manj prekrivanja obraza. Okuševalci so spili določeno količino pijače, umaknili slamico in pustili pijačo

par sekund v ustih zato, da bi se na obrazu jasno videl odziv na okus. Obraz so okuševalci imeli med snemanjem frontalno obrnjen proti kameri. Pazili smo, da se okuševalci pri vnosu pijače niso sklanjali naprej, temveč so si pijačo prinesli z roko bližje k obrazu in da so po zaužitju čim prej odmaknili slamico iz obraznega dela. Sicer smo se pri snemanju trudili, da so bili izrazi čim bolj spontani. Pri zaužitju hrane so okuševalci vnesli hrano večinoma z rokami. Tudi tukaj smo pazili, da je bilo pri zaužitju, prekrivanje obraza čim krajše. Pazili smo tudi, da so bili obrazi snemani frontalno in da se okuševalci pri zajemu hrane niso sklanjali naprej. Pri okušanju hrane smo okuševalcem dopustili, da so hrano žvečili poljubno dolgo. Posledično so bili posnetki okušanja hrane daljši in s tem množica slik pri testiranju večja. Zato, da posnetki okušanje hrane niso bili predolgi, smo uporabili manjše kose hrane. Pri snemanju smo uporabili različne vrste hrane in pijače. Ob koncu okušanja so okuševalci podali oceno okusa hrane oz. pijače na lestvici od 1 do 7, kjer 1 pomeni zelo neprijeten okus, 7 pa zelo prijeten, 4 pa nevtralen okus. Posnetke z ocenami od 1 do 3 smo potem obravnavali kot negativne odzive, posnetke z ocenami od 4 do 7 pa kot pozitivne odzive. Pri snemanju smo za vsako osebo posneli negativne in pozitivne odzive, tako za pitje pijače, kot tudi za okušanje hrane. V večini primerov smo pri posamezni osebi posneli dva posnetka z določeno vrsto hrane oziroma pijače. Bazo sestavlja 66 posnetkov okušanja pijače in 69 posnetkov okušanja hrane.

5. PREIZKUS NA PRIPRAVLJENI BAZI

Posnetke smo najprej pretvorili v zaporedje sličic velikosti 200×200 slikovnih elementov. Na dobljenih zaporedjih sličic smo potem ročno določili območja po zaužitju hrane, ki smo jih potem uporabili pri klasifikaciji. Bazo smo ovrednotili z večkratnim trikratnim prečnim preverjanjem, ločeno smo obravnavali množici okušanja pijače in hrane. Eno izmed treh skupin smo uporabili kot učno in ostali dve kot testni množici. Tako je bil vsak posnetek dvakrat v testni množici. V vsaki iteraciji prečnega preverjanja smo posnetke naključno razdelili v 3 skupine tako, da je bilo za vsako osebo število pozitivnih in negativnih odzivov v skupinah čimbolj enakomerno razporejeno. Za klasifikacijo smo uporabili en klasifikator SVM z jedrom RBF. Končen odziv posnetka je bil določen glede na prevladujočo napoved po sličicah. Končni rezultati prečnih preverjanj smo zaokrožili na celo število. Pri pijači je povprečna natančnost z desetimi prečnimi preverjanji 92%, najnižja dosežena natančnost je bila 87%, najvišja pa 94%. Pri hrani je povprečna natančnost z desetimi prečnimi preverjanji 81%, najnižja dosežena natančnost je bila 76%, najvišja pa 84%.

Podrobno smo analizirali tudi napake. Tako pri pijači kot pri hrani je velik delež napak predstavljajo zgolj nekaj posnetkov. Pri pijači je 44 posnetkov vedno brez napak, torej točno dve tretjini, medtem, ko je bilo pri hrani takih posnetkov slaba tretjina. Pri hrani prevladujejo narobe napovedani posnetki z negativnim odzivom. Za te posnetke je značilno, da je za razliko s posnetki negativnega odziva, ki so pravilno klasificirani, na teh posnetkih sploh težko zaznati negativen odziv na obrazu (Slika 2).



Slika 2: Posnetki negativnega odziva na hrano; zgornje tri slike prikazujejo izrazit negativni odziv, pri spodnjih treh slikah pa negativni odziv ni izrazit.

Pri pijači in pri hrani smo identificirali dva posnetka z negativnim odzivom, ki sta bila velikokrat narobe klasificirana. Oba posnetka sta bila druga oziroma tretja zaporedna posnetka snemanja okušanja iste pijače oziroma hrane. V primerjavi s predhodnimi posnetki, na teh posnetkih negativni odziv na obrazu ni zelo izrazit, saj se je okuševalec najbrž navadil na neprijetni okus. Tako pri pijači, kot tudi hrani, smo pri posnetkih z negativnim odzivom opazili še eno slabost naše metode. Posnetki so sicer prikazovali izrazito opazen negativni odziv na okus, vendar je ta odziv nastopil šele z zamikom. Zato je bil posnetek napačno označen kot pozitiven odziv. Pri pijači je bil en posnetek sicer zmeraj pravilno napovedan, vendar je bilo kar okoli 46% sličic posnetka zmeraj napačno napovedanih, ravno zaradi zakasnelega negativnega odziva.

6. ZAKLJUČEK

V članku smo prikazali uporabo metod za prepoznavo obraznega izraza za napovedovanje ali je osebi okus zaužite hrane ali pijače všeč. Na lastni bazi posnetkov smo dobili zelo dobre rezultate na množici pitja pijače in dobre rezultate na množici okušanja hrane. Pri pijači je bila povprečna natančnost 92%, pri hrani pa 81%. Podrobno smo tudi analizirali napake. Ugotovili smo, da se večina napak pojavi pri analizi posnetkov okušanja hrane, kadar je na posnetku neprijeten odziv na okus hrane, toda izraz na obrazu okuševalca ni dovolj izrazit.

Možno izboljšavo metode vidimo v avtomatski izbiri sličic po zaužitju hrane s pomočjo metod za prepoznavanje prekrivanja delov obraza, predvsem na območju okoli ust.

Uporabili bi lahko metode za istočasno zaznavanje lokacije in prekrivanja obraznih značilk [11, 12] ali za zaznavanje prekrivanja glede na spremembe v vektorskem polju slike [13, 14]. Za prepoznavanje prekrivanja bi lahko uporabili tudi naprednejše senzorje kot so Kinect. Bazo posnetkov, bi bilo potrebno dopolniti z večjim številom posnetkov, lahko tudi v zahtevnejših pogojih. Tudi samo metodo za prepoznavo izrazov bi bilo smiselno preveriti z naprednejšim postopkom normalizacije in značilkami kot so LPQ, HOG, PHOG, LBP-TOP in LPQ-TOP.

LITERATURA

1. K. Wendin, BH. Allesen-Holm, WLP. Bredie (2011), Do facial reactions add new dimensions to measuring sensory responses to basic tastes?, *Food Quality and Preference*, št. 4, zv. 22, str. 346-354.

2. R.A. de Wijk, V. Kooijman, R.H. Verhoevenb, N.T. Holthuysen, C. de Graafc (2012), Autonomic nervous system responses on and facial expressions to the sight, smell, and taste of liked and disliked foods, *Food Quality and Preference*, št. 2, zv. 26, str. 196-203.

3. R.A. de Wijk, W. He, M.G. Mensink, R.H. Verhoeven, C. de Graaf (2014), ANS responses and facial expressions differentiate between the taste of commercial breakfast drinks, *PloS one*, št. 4. zv. 9.

4. L. Danner, L. Sidorkina, M. Joechl, K. Duerrschmid (2014), Make a face! Implicit and explicit measurement of facial expressions elicited by orange juices using face reading technology, *Food Quality and Preference*, zv. 32, str. 167-172.

5. G. Juodeikiene, L. Basinskiene, D. Vidmantiene, D. Klupsaite, E. Bartkiene (2014), The use of face reading technology to predict consumer acceptance of confectionery products, *V 9th Baltic Conference on Food Science and Technology "Food for Consumer Well-Being" FOODBALT 2014*, str. 276-279.

6. *http://www.noldus.com/human-behavior-research/products/facereader*
Program FaceReader za prepoznavo 6 osnovnih čustev.

7. P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews (2010), The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, *V Computer Vision and Pattern Recognition Workshops (CVPRW)*, str. 94-101.

8. M.F. Valstar, M. Mehu, B. Jiang, M. Pantic, K. Scherer (2012), Meta-analysis of the first facial expression recognition challenge, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, št. 4, zv. 42, str. 966-979.

9. J.M. Saragih, S. Lucey, J.F. Cohn (2009), Face alignment through subspace constrained mean-shifts, *IEEE 12th International Conference on Computer Vision*, str. 1034-1041.

10. V. Kazemi, J. Sullivan (2014), One millisecond face alignment with an ensemble of regression trees, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, str. 1867-1874.

11. G. Ghiasi, C.C. Fowlkes (2014), Occlusion coherence: Localizing occluded faces with a hierarchical deformable part model, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, str. 1899-1906.

12. X.P. Burgos-Artizzu, P. Perona, P. Dollár (2013), Robust face landmark estimation under occlusion, *IEEE International Conference on Computer Vision (ICCV)*, str. 1513-1520.

13. M. Mahmoud, R. El-Kaliouby, A. Goneid (2009), Towards communicative face occlusions: machine detection of hand-over-face gestures, *Proceedings of the 6th International Conference on Image Analysis and Recognition (ICIAR)*, str. 481-490.

14. J. Xu, X. Zhang (2015), A Real-Time Hand Detection System during Hand over Face Occlusion, *International Journal of Multimedia and Ubiquitous Engineering*, št. 8, zv. 10, str. 287-302.