

UNIVERZA V LJUBLJANI  
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Klemen Strušnik

**Računalniška prepoznavna obraznih  
izrazov**

DIPLOMSKO DELO

VISOKOŠOLSKE STROKOVNE ŠTUDIJSKE PROGRAM PRVE  
STOPNJE RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: doc. dr. Iztok Lebar Bajec

Ljubljana 2013



Rezultati diplomskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja. <sup>1</sup>

*Besedilo je oblikovano z urejevalnikom besedil  $\LaTeX$ .*

---

<sup>1</sup>V dogovorju z mentorjem lahko kandidat diplomsko delo s pripadajočo izvorno kodo izda tudi pod katero izmed alternativnih licenc, ki ponuja določen del pravic vsem: npr. Creative Commons, GNU GPL. V tem primeru na to mesto vstavite opis licence, na primer tekst [?]



Št. naloge: 00480/2013

Datum: 15.04.2013



Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **KLEMEN STRUŠNIK**

Naslov: **RAČUNALNIŠKA PREPOZNAVA OBRAZNIH IZRAZOV**  
**COMPUTER RECOGNITION OF FACIAL EXPRESSIONS**

Vrsta naloge: Diplomsko delo visokošolskega strokovnega študija prve stopnje

Tematika naloge:

Računalniki in njim sorodne sodobne elektronske naprave že nekaj časa niso zgolj orodje, ampak vse bolj postajajo naši neločljivi spremljevalci. Spremljevalci, ki so (žal) prisotni tudi na mestih, do katerih naj bi šel še cesar sam. Slednje pomeni, da so ob nas v »dobrem in slabem,« da so morda z nami tudi v trenutkih, ko si ne želimo bližine »biološke inteligence.« Če bi bile elektronske naprave sposobne razbrati naše trenutno emocionalno stanje, bi nam lahko bile v veliko pomoč. S primernimi aktivnostmi bi lahko ali prilagodile delovanje ali morda celo vplivale na naše počutje.

V diplomski nalogi preučite trenutne razmere raziskav na področju računalniške prepoznavne obraznih izrazov. Osredotočite se na zgolj nekaj algoritmov in enega izmed njih poskusno implementirajte v okolju Matlab. Pridobljene izkušnje ustrezno komentirajte.

Mentor:

doc. dr. Iztok Lebar Bajec



Dekan:

prof. dr. Nikolaj Zimic



## IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Spodaj podpisani Klemen Strušnik, z vpisno številko **63080370**, sem avtor diplomskega dela z naslovom:

*Računalniška prepoznavna obraznih izrazov*

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom doc. dr. Iztoka Lebarja Bajca,
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki "Dela FRI".

V Ljubljani, 20. avgusta 2013

Podpis avtorja:



*Zahvaljujem se mentorju doc. dr. Iztoku Lebarju Bajcu za strpnost, moralno podporo in strokovno pomoč. Zahvalil bi se še bivšim sošolcem, ki so mi dali motivacijo, da sem lahko diplomsko nalogo pripeljal do konca.*



# Kazalo

Povzetek

Abstract

Seznam kratic in pojmov

<b>1</b>	<b>Uvod</b>	<b>1</b>
1.1	Motivacija in cilji . . . . .	1
1.2	Emocije . . . . .	2
1.3	Trenutno stanje . . . . .	4
<b>2</b>	<b>Algoritem prepoznavanja emocije</b>	<b>8</b>
2.1	Koraki . . . . .	8
2.2	Priprava podatkov . . . . .	10
2.3	Klasifikacija . . . . .	13
2.4	Metoda glavnih komponent - PCA . . . . .	14
2.5	Lokalni Gaborjevi binarni vzorci - LGBP . . . . .	14
2.6	Aktivni videzni model - AAM . . . . .	18
2.7	Združitev LGBP in AAM . . . . .	24
<b>3</b>	<b>Rezultati</b>	<b>25</b>
<b>4</b>	<b>Sklepne ugotovitve</b>	<b>27</b>



# Povzetek

V diplomski nalogi preizkusim eno izmed bolj znanih metod za modeliranje obraznih slik iz teh nato izluščim relevantne podatke, katerim kasneje zreduciram dimenzijo. S temi podatki lahko manipuliram obliko in teksturo obrazne slike. Iz teh slik s pomočjo LGBP tvorim histograme, ki predstavljajo vhod ene izmed bolj znanih metod strojnega učenja.

V prvem delu definiram emocijo in proces prepoznavanja obraznih izrazov, nato pa izpostavim faktorje, ki lahko vplivajo na rezultat same prepoznavne. Opisujem orodja in vmesnike za prepoznavo obraznih izrazov in lastnosti obraza, ki so trenutno na trgu.

V drugem delu so predstavljeni koraki za predprocesiranje podatkov, opisani so algoritmi relevantni za razumevanje samega procesa prepoznavne z LGBP in AAM in potek implementacije teh algoritmov v Matlab okolju. V zadnjem delu so predstavljeni rezultati klasifikacije obdelanih podatkov in dojetje teh rezultatov predvsem s kognitivnega vidika.

Za konec so podane še sklepne ugotovitve in morebitneboljšave uporabljenih metodologij.



# Abstract

In this diploma I try one of the more popular methods for modelling facial images and the extracted data's dimension is later on reduced. I can manipulate the shape and texture with the images reduced data. From these changed images the LGBP algorithm creates histograms, that represents the input of one of the most popular methods for machine learning.

In the first part I define emotions and the process of recognizing facial expressions. I mention the factors that can effect on the result of the recognition and some tools for recognizing facial features and facial expressions. In the second part the steps for preprocessing data are mentioned and relevant algorithms for understanding the process of recognition with LGBP and AAM are described - with the steps for implementation in Matlab. The last part consists of classification results of the processed data and the meaning behind them.

In the end part I explain my findings and propose some optimization techniques for this application.



# Seznam kratic in pojmov

AAM - Aktivni videzni model (angl. Active Appearance Model)

ASM - Aktivni oblikovni model (angl. Active Shape Model)

LGBP - Lokalni Gaborjevi binarni vzorci (angl. Local Gabor Binary Patterns)

PCA - Metoda glavnih komponent (angl. Principal Component Analysis)



# Poglavje 1

## Uvod

### 1.1 Motivacija in cilji

Sama tematika diplomske naloge, se loti vprašanja, ali so računalniki zmožni sočustvovanja oz. empatije. Slednje je možno le ob prepoznavi stanja osebk; tedaj ga je možno sprogramirati na način, ki bo pripomogel k izboljšanju tega stanja oz. počutja te osebe. Prvi korak pri tem je prepoznavna emocionalnega stanja iz obraznih izrazov. Zaenkrat poznamo sedem emocij: jeza, strah, gnus, zavist, sreča, žalost in presenečenje [3]. Pri prepoznavi pa je v igri preveliko faktorjev, da bi lahko zgradili učinkovito in robustno aplikacijo. Med te faktorje bi lahko všteli slabo prostorsko osvetlitev ali zaraščenost obraza. Včasih pa tudi sami ne moremo razbrati med strahom in presenečenjem, ker lahko sam osebek prikriva svojo emocijo ali pa ima kak obrambni mehanizem, kjer se v primeru živčnosti ali strahu začne smejeti. Diploma se dotakne futuristične ideje o robotih, ki reagirajo glede na naše stanje. Tako smo lahko jezni in nam robot citira kak izrek, ki nas spravi v nevtralno stanje, temu lahko sledi šala in se dejansko počutimo malce boljše. Z diplomsko nalogo želim raziskati ali je tak proces moč ustvariti z računalnikom.

## 1.2 Emocije

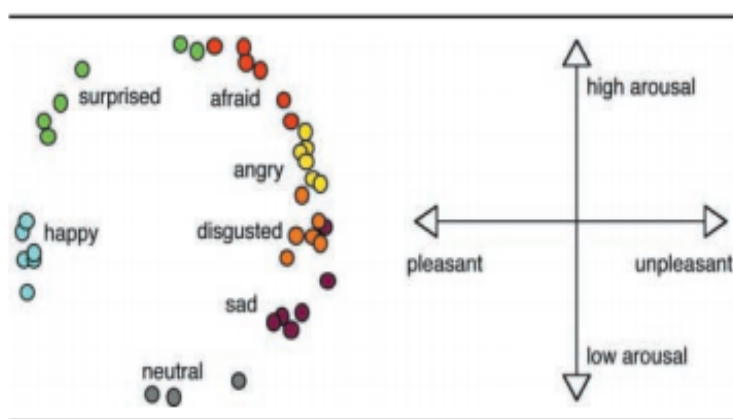
Slika 1.1 prikazuje sedem glavnih emocij. Za prepoznavo obraznih izrazov teh emocij pa ne zadostujejo le fizične lastnosti vidnega, ampak je za kategorizacijo emocije potrebno neko predznanje o svetu, nek spomin. Najpreprostejši način prepoznave je primerjava neke spominske slike iz preteklosti in slike, ki je trenutno opazovana. Vendar spomin ni vse, upoštevati je potrebno še asociativni mehanizem, ki vidno primerja z drugimi stimulusi v svetu na kate-  
 re lahko asociira trenutno opazovan obrazni izraz. Primer takih stimulusov vključuje znanje o tem kje je bil ta obraz že viden, kaj je bilo o tem rečenega, od koga je bil ta obraz in kakšen je bil odziv nanj.



Slika 1.1: Primeri sedmih glavnih emocij.

Nekatera odkritja iz računalniškega modeliranja bi trdila, da so za prepoznavo emocije iz obraznih izrazov dovolj le geometrične lastnosti izrazov, nakar nekatere kulturne študije trdijo, da je bistvo emocije v očeh. Prepoznavna identitete obraza je odvisna tudi od informacije o prostorskih povezavah med individualnimi obraznimi lastnostmi. Nekateri računalniški modeli pa nudijo dokaz, da je tovrstna prepoznavna potratna. V kontrast nekatere študije na človeških subjektih kažejo, da percepcija emocije potrebuje tudi nekaj procesiranj povezav med večimi obraznimi lastnostmi. Kar pomeni, da je sicer možno interpretirati obrazni izraz na podlagi individualnih obra-

znih lastnosti, vendar se računalniški modeli bolje izkažejo pri procesiranju večih obraznih lastnosti hkrati [1]. Te lastnosti se lahko tudi razlikujejo med različnimi emocijami: notranjo srečo lahko prepoznamo z eno lastnostjo tj. smehom, nakar diskriminacija med negativnimi emocijami potrebuje dodatno informacijo o konfiguraciji obraza. Če povprečimo podatke o obraznih izrazih, lahko s pomočjo metode glavnih komponent - *PCA* (angl. Principal Component Analysis) vidimo diskriminacijo med samimi emocijami, kot je prikazano v sliki 1.2.



Slika 1.2: Prikaz emocij v nižjih dimenzijah s PCA metodo.

### 1.3 Trenutno stanje

V svetu kognitivne znanosti je bilo zasnovanih veliko metod za prepoznavo izrazov vendar zaenkrat še ne obstaja model s katerim bi lahko definitivno rešili problem prepoznave obraznih izrazov. V proces prepoznave je potrebno vključiti le slike, ki vsebujejo ključne podatke o izrazu, zato je potrebno izluščiti ozadje in poenotiti velikost obrazov, detektirati robove in ohraniti prostorsko informacijo, ter poiskati t.i. lastnosti, ki bi lahko enolično definirale emocijo. Na razpolago imamo bolj skopo količino podatkov, zato je potrebno iz slike izluščiti čim več informacij. To lahko postorimo z uporabo večih metod hkrati, npr. poleg aktivnega videznega modela - *AAM* (angl. Active Appearance Model) in lokalnih Gaborjevih binarnih vzorcev - *LGBP* (angl. Local Gabor Binary Patterns), pristopov opisanih v diplomski bi lahko dodali še prepoznavo na podlagi razdalj med akcijskimi enotami (angl. Action Units) [14]. Akcijska enota je definirana kot razdalja med posameznimi elementi na obrazu npr. dvig obrvi ali razteg ustnic glede na razdalje pri brezizraznem obrazu.

Ko so elementi združeni v nek sklop, se je potrebno ozreti še na samo praktičnost aplikacije, saj je dobro proces avtomatizirati. Trenutne raziskave na področju avtomatske prepoznave sledijo naslednjim smernicam:

- gradnja bolj robustnih sistemov za pridobitev obraza, izluščitev obraznih podatkov in reprezentacija, prepoznavo obraznih izrazov relativno na gibanje obraza, prekrivanje, spremembe v svetlobi in nizko intenzivni izrazi;
- pridobitev več lastnosti obraza za prepoznavo večih izrazov in doseči visoko natančnost pri napovedovanju;
- prepoznavo obraznih akcijskih enot in kombiniranje le-teh.

### 1.3.1 Zaznavanje obraza in obraznih znamenitosti

Avtomatska zaznava obraza in obraznih znamenitosti je lahko uporabljena na več področjih: emocijska in paralingvistična komunikacija, klinična psihologija, psihiatrija, neurologija, določanje bolečine, prepoznavna laži, inteligentna okolja ali multimodalni vmesnik človek-računalnik (HCI). Nekatera izmed bolj znanih orodij, ki podpirajo vsaj avtomatsko zaznavo obraza, so:

- Windows Live fotogalerija vključuje prepoznavo obrazov;
- Google Picasa ima vgrajen obrazni prepoznavni sistem, ki asociira obraze z ljudmi;
- Sony's Picture Motion Browser analizira fotografije in razločuje med večini osebami v sliki;
- Applov iPhoto organizator slik ima lastnost imenovano 'Faces' v katerih dovoljuje uporabnikom, da povezujejo imena z obrazi v seriji fotografij;
- Betaface je spletna storitev za biometrično obdelavo slik;
- OpenBR odprtokodni obrazni prepoznavni sistem in raziskovalna platforma za biometrični razvoj algoritmov.

### 1.3.2 Prepoznavanje emocije

Za prepoznavo emocije je potrebno po zaznavi obraza dodatno analizirati in prepoznati obrazne znamenitosti. V večini primerov se sestavijo obrazni modeli, ki se kasneje klasificirajo z uporabo ustreznih metod strojnega učenja. Seveda se na trgu za ohranjanje konkurenčnosti uporabijo tudi druge metode, ki običajno niso javno dostopne. Kot zanimivost bi omenil, da vsa prej omenjena orodja v svojih rešitvah nudijo tudi aplikacijski vmesnik za programiranje. Glavna akterja, ki sta danes prisotna na trgu sta:

- Nviso <sup>1</sup> s analizira emocije iz obraznih izrazov na slikah in v videih na podlagi 43 ključnih točk (glej sliko 1.3);

---

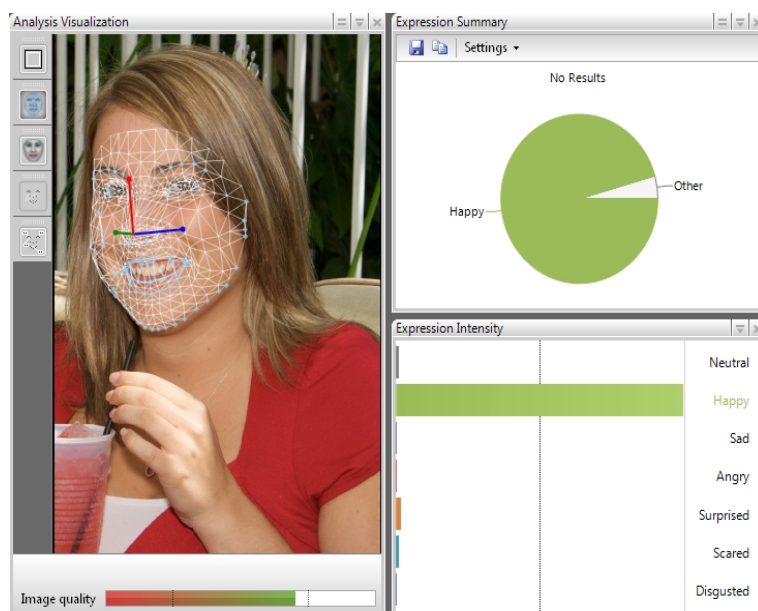
<sup>1</sup><http://nviso.ch/>

- FaceReader <sup>2</sup> analizira emocije iz obraznih izrazov samo na slikah na podlagi 500 ključnih točk (glej sliko 1.4).



Slika 1.3: Prepoznavna emocij z Nvivo.

<sup>2</sup><http://www.noldus.com/human-behavior-research/products/facereader>



Slika 1.4: Prepoznavna emocij z FaceReader.

## Poglavje 2

# Algoritem prepoznavanja emocije

### 2.1 Koraki

Modelsko osnovana interpretacija slik spremenljivih objektov je v zadnjem obdobju precej popularna tema [4]. Dodatna sposobnost razlage slike v smislu množice modelskih parametrov ponuja naravno osnovo za interpretacijo le-te. Model je zelo občutljiv, zato mora biti tem bolj popoln, kar pomeni da mora biti sposoben sintetizacije bližnje aproksimacije katerekoli slike ciljnega objekta. Modelsko osnovane metode z razmeroma majhnim številom parametrov so kar uspešne [3], vendar nobena ne uporablja polnih fotorealističnih modelov. Razlika pa je v tem da fotorealistični modeli vsebujejo bolj realistične reprezentacije originalne slike. Ti fotorealistični modeli tipično vključujejo veliko število parametrov, ponavadi od 50 do 500. Velikost števila parametrov je primerna, saj se tako lahko soočajo z variabilnostjo ciljnih objektov.

Pristop, ki ga uporablja fotorealističen model, z razlikovanjem opazovane in modelske slike se uporablja v fazi predprocesiranja podatkov in sicer tik preden se množico parametrov modela zreducira z metodo glavnih komponent. Optimizacija algoritmov v neobdelanem večdimenzionalnem prostoru

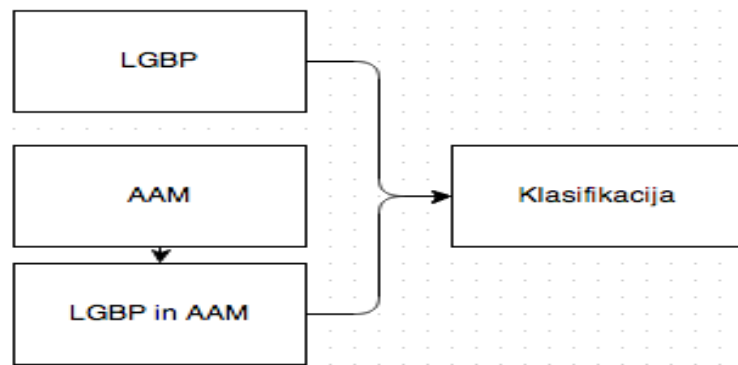
je sicer dobra, vendar počasna.

V literaturi je bilo opisanih veliko modelsko osnovanih načinov interpretacije slik [3]. Pri vseh je cilj doseči dobre rezultate z uporabo modela, ki vsebuje najboljše približke testnih slik na objektu modela. Pridobljeni model daje tudi osnovo za širok spekter aplikacij, ki opisujejo sliko preko modelskih parametrov. Ti parametri so uporabni za interpretacijo na visokem nivoju. Na primer, ko analiziramo obraz slike lahko te parametre uporabimo za prepoznavanje identitete, poze ali izraza na obrazu. Zato da interpretiramo novo sliko, mora biti uporabljena primerna metoda za iskanje najboljšega ujemanja med sliko in modelom.

Pred revolucijo modelsko osnovanih načinov interpretacije slik, pa je analiza potekala na podlagi iskanja vzorcev. Ojala in sodelavci [10] predlagajo klasifikacijo na podlagi lokalnih binarnih vzorcev - LBP (angl. Local Binary Pattern), ki se kasneje nadgradi z Gaborjevim filtrom [12] za iskanje robov in nastanejo lokalni Gaborjevi binarni vzorci - LGBP (angl. Local Gabor Binary Patterns). Pri LGBP pristopu je obraz modeliran kot sekvenca histogramov s konkatencijo histogramov lokalnih regij vseh Gaborjevih magnitudnih binarnih vzorčnih načrtov [12]. Kar v splošnem pomeni, da Gaborjev filter vrne robove, lokalne regije pa predstavljajo kvadratki, ki so binarno vzorčeni na podlagi intenzitete robov. Histograme binarnih vzorcev lokalnih regij se nato združi in ta združen histogram predstavlja vir razločanja.

Cootes in sodelavci modelirajo obliko in lokalni sivinski izgled z uporabo aktivnega oblikovnega modela - *ASM* (angl. Active Shape Model) za iskanje fleksibilnih objektov v novi sliki [2]. Lanitis in sodelavci [8] so uporabili ta način za interpretacijo obraznih slik in ga nadgradili z iskanjem oblike z uporabo *ASM*, kjer je obraz lahko zvit na normaliziran model in uporabljen za interpretacijo slike. Edwards in sodelavci [5], razširijo to delo za stvaritev skupnega modela oblik in sivinskih izgledov, ampak se preveč zanašajo na *ASM* pri iskanju obrazov v novih slikah. Cootes in sodelavci [3] so modelirali tudi obliko in sivinsko nivojski izgled z uporabo aktivnega videznega modela - *AAM* (angl. Active Appearance Model) za iskanje fleksibilnih objektov v

novih slikah. Pristop, uporabljen v tej nalogi je združitev LGBP in AAM [11], kjer lahko s spreminjanjem parametrov AAM tvorimo veliko količino LGBP histogramov in tako zmanjšamo dvom v fazi napovedovanja. Pri tem se je LGBP zaradi enostavnosti razvil pred AAM, saj lahko deluje tudi kot samostojen algoritem za prepoznavanje emocije in se kasneje združil z AAM, ki predstavlja jedro naše aplikacije. Razvojni koraki so opisani v sliki 2.1



Slika 2.1: Razvojni koraki aplikacije.

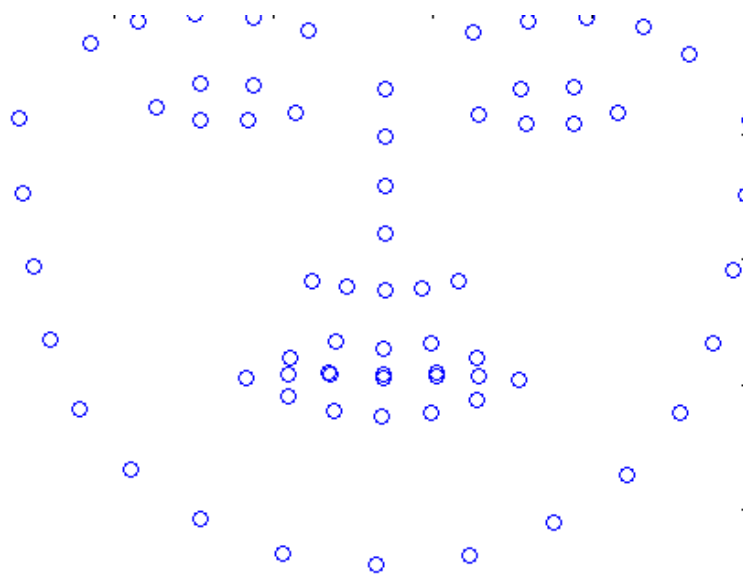
## 2.2 Priprava podatkov

Preden se lotimo obravnave posameznih metod je potrebno omeniti nabor podatkov, ki jih bomo uporabili pri sami tvorbi modelov. Baza se imenuje CK+ [9] in jo sestavlja 327 obraznih slik (glej tabelo 2.1). Obrazni izrazi 210 odraslih ljudi so bili posneti z dvema strojno sinhroniziranimi Panasonic AG-7500 kamerama. Prostovoljci so bili stari od 18 do 50 let, 69% jih je bilo ženskega spola, 81% Euro-Ameriške, 13% Afro-Ameriške, ter 6% ostalih narodnosti.

Tabela 2.1: Število slik na emocijo.

Emocija	N
Jeza	45
Zavist	18
Gnus	59
Strah	25
Sreča	69
Žalost	28
Presenečenje	83

Naši podatki so sestavljeni iz teksture obrazne slike in iz niza znamenitosti (angl. landmarks), in predstavljajo parametre modela za opazovano obrazno sliko iz podatkovne baze CK+.



Slika 2.2: Slika znamenitosti.

Znamenitost je definirana kot točka v nekem koordinatnem sistemu in predstavlja obliko obraza. Znamenitosti so standardizirane, na vseh slikah je enako število znamenitosti in za posamezno emocijo se nahajajo v podobnih

regijah. Na posnetku obraza na levi strani slike 2.3 sta razvidna še ozadje in trup osebe in te podatke je potrebno za večjo natančnost pri napovedovanju odstraniti. To lahko postorimo tako, da posnetke pretvorimo v sivinske, in vse kar je izven koordinat znamenitosti očrnimo ( $RGB = \{0,0,0\}$ ). Nato izluščimo le kvadrant v katerem se nahajajo vse znamenitosti in velikost slike poenotimo na  $45 \times 45$ . Izpis Matlab izvorne kode, ki po tem postopku pripravi podatke, je prikazan v algoritmu 1. Podatki so tako pripravljeni za LGBP in AAM.

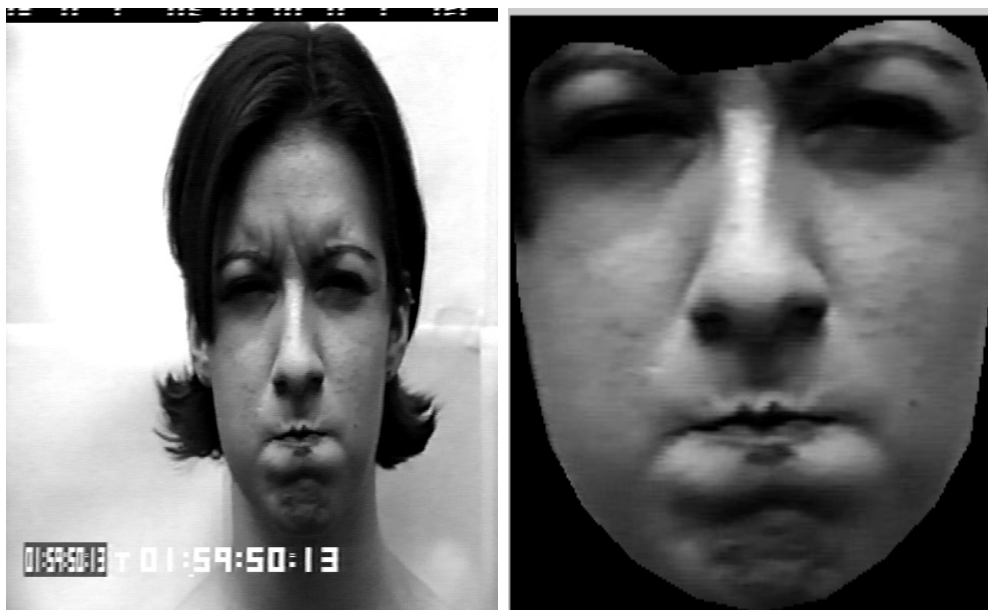
---

**Algorithm 1** Matlab koda za pripravo podatkov.

---

```
xmin = min(shape1(:,1)) ;
ymin = min(shape1(:,2));
width = max(shape1(:,1))-min(shape1(:,1));
height = max(shape1(:,2))-min(shape1(:,2));
im1 = imcrop(im1, [xmin, ymin, width, height]); %izluscimo le
    kvadrant v katerem je obraz
landmarks_xdiff = shape1(:,1) - min(shape1(:,1)); %postavimo v
    novo izhodišce
landmarks_ydiff = shape1(:,2) - min(shape1(:,2));
shape1(:,1) = landmarks_xdiff;
shape1(:,2) = landmarks_ydiff;
```

---



Slika 2.3: Razlika med originalno in pripravljeno sliko.

## 2.3 Klasifikacija

Rezultat LGBP in AAM sta zgolj združen histogram ali model obraza, ki lahko služita nadaljnji razpoznavi emocije. V ta namen se uporabljajo različni klasifikacijski algoritmi. Metoda podpornih vektorjev - SVM (angl. Support Vector Machines) spada med te algoritme strojnega učenja in se uporablja za analizo podatkov in prepoznavi vzorcev v klasifikacijski ali regresijski analizi. Pri večini algoritmov strojnega učenja želimo zmanjšati število atributov, pri metodi SVM pa uporabimo vse razpoložljive attribute, tudi manj pomembne in jih z linearno kombinacijo uporabimo za napovedovanje odvisne spremenljivke [7]. Primerne so za učenje na velikih množicah primerov, ki vsebujejo veliko število pomembnih ali nepomembnih atributov. Metode dosežejo visoko stopnjo natančnosti, a je njihovo odločitev dostikrat težko razumeti. Klasifikacijske metode SVM so generalno namenjene razločevanju dveh razredov. V našem problemu imamo množico primerov, ki pripadajo večim razredom. Za napovedovanje tovrstnih podatkov en razred ločimo od ostalih.

## 2.4 Metoda glavnih komponent - PCA

Analiza multivariantnih podatkov igra ključno vlogo pri analizi podatkov. Multivariantni podatki so sestavljeni iz več različnih atributov ali spremenljivk izvzetih iz nekega opazovanja. Če je  $p$  število spremenljivk v podatkovni bazi, lahko vsako spremenljivko interpretiramo kot dimenzijo v  $p$ -dimenzionalnem hiperprostoru. Večdimenzionalni hiperprostor si je dostikrat zelo težko predstavljati in obvladovati.

Glavna naloga metode glavnih komponent je redukcija dimenzionalnosti in združevanje skupnih opazovanj. PCA je matematično definirana kot ortogonalna linearna transformacija podatkov na novi koordinatni sistem, tako da največja varianca, katerekoli projekcije podatkov, leži na prvi koordinati imenovani prva glavna komponenta, druga največja varianca na drugi glavni komponenti, itd. V bolj posplošenem smislu, PCA analizira varianco multivariantnih podatkov.

Pred PCA je potrebno narediti normalizacijo s povprečjem. Temu sledi izračun kovariančne matrike in lastnih vektorjev te matrike. Za izračun teh vektorjev se uporabi singularna dekompozicija - SVD (angl. Singular Value Decomposition), ta omogoča, da za realno matriko  $A$  velikosti  $n \times m$ , dobimo  $n \times n$  ortogonalno matriko  $U$  in  $n \times m$  diagonalno matriko  $S$  ter  $m \times m$  ortogonalno matriko  $V$  pri čemer velja  $A = USV$ . Pri tem velja, da so koeficienti, ki jih bomo spreminjali v algoritmu 2 predstavljeni kot  $U_{red}$ , uteži PCA, so pa podatki predstavljeni v zmanjšani dimenziji in označeni s spremenljivko  $z$ . Za pravilno uporabo PCA je potrebno podatke prej normalizirati s povprečjem, kot je predstavljeno v algoritmu 2.

## 2.5 Lokalni Gaborjevi binarni vzorci - LGBP

Lokalni binarni vzorci, LBP, spadajo med neparametrične metode, ki učinkovito izluščijo lokalne strukture slike. Uporabljajo se predvsem pri analizi obraznih slik, kar vključuje prepoznavo obraza, analizo obraznih izrazov in njihovo klasifikacijo.

**Algorithm 2** PCA algoritem

---

```

X = x-mean(x); %normalizacija s povprecjem
sigma = (X'*X)./m; %izracun kovariancne matrike
[U,S,V] = svd(sigma); %izracun eigenvektorjev
Ured = U(1,1:k); %k-stevilo izbranih komponent
z = Ured'*X; %z predstavlja x v zreducirani dimenziji

```

---

LBP izluščijo lokalne strukture slike tako, da primerjajo piksel z njegovimi okoliškimi. Najbolj pomembna lastnost LBP je toleranca do sprememb v osvetlitvi. Prvotno so bili LBP namenjeni prepoznavi teksta [10] in so se izkazali kot preprost, a močen pristop za opis lokalnih struktur. Primeri uporabe LBP: analiza obraznih slik, rekonstruiranje slik in videa, modeliranje okolice, vizualna inšpekcija, analiza gibanja in biomedicinska obdelava slik [6]. Orientacija glave v večpoglednih obraznih slikah je zelo pomembna za oceno poze neke obrazne slike in LBP prevzamejo to dobro lastnost z uporabo Gabor filtra. V LGBP lahko Gabor filter z osmimi različnimi orientacijami odraža orientacijo glave v večpoglednih obraznih slikah in nato LBP operator, na osnovi Gabor lastnosti pridobi lokalno informacijo na različnih orientacijah in na različnih skalah. Gabor filtri so bili uporabljeni v aplikacijah za segmentacijo tekstur, detekcijo tarče, analizo dokumentov, prepoznavo robov in za kodiranje slik [15]. Gabor filter si lahko predstavljamo kot sinusoidno ravnino določene frekvence in orientacije, ki je modulirana z Gaussovo ovojnico, kar prikazuje enačba (2.1).

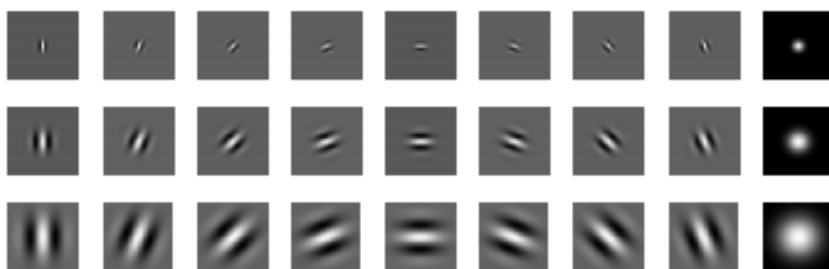
$$g(x, y) = a(x, y)c(x, y) \quad (2.1)$$

$$a(x, y) = \frac{1}{(2\pi)\sigma_x\sigma_y} * e^{-0.5*\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} \quad (2.2)$$

$$c(x, y) = \cos(2\pi(F_x x + F_y y)) \quad (2.3)$$

Enačba (2.2) predstavlja Gaussovo komponento in enačba (2.3) sinusoidno komponento, kjer sta  $x$  in  $y$  prostorski spremenljivki. Standardna devia-

cija  $(\sigma_x, \sigma_y)$  v enačbi (2.2) opisujeta velikost Gaussove ovojnice in definirata skalo filtra skupaj s prostorskimi in spektralnimi osmi.  $(F_x, F_y)$  predstavlja frekvenco sinusoidne komponente in tako center frekvence filtra v dvodimenzionalni frekvenčni domeni. Orientacija filtra je definirana kot enotski vektor od izvora  $(x, y)$  do centra frekvence  $(F_x, F_y)$  filtra. Slika 2.4 prikazuje skupek Gabor filtrov pri 3 skalah in 9 orientacijah (skupaj 27 filtrov). Osvetljeni in zasenčeni deli pa označujejo pozitivne in negativne vrednosti filtra.

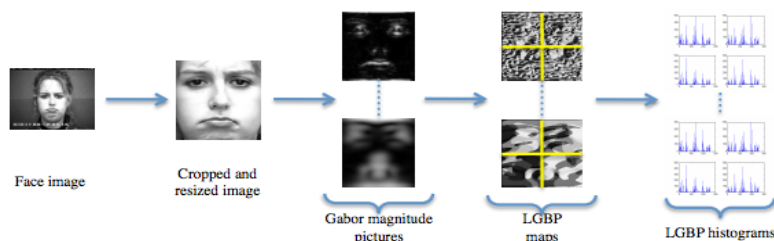


Slika 2.4: 2D Gabor filter s 3 skalami in 9 orientacijami.

Kombinacija Gabor filtra in LBP izboljšuje reprezentacijo moči v večpoglednih obraznih slikah. Histogram lokalnih regij naredi diskriminatorne lastnosti, ki so robustne na napake pri poravnavi. LGBP lahko deluje tudi kot samostojna metoda za izluščitev podatkov iz obraznih slik. Implementacijo vključuje pridobitev slik, obdelavo slik z Gabor filtrom, razdelitev na kvadrante in izračun histogramov na posameznih kvadrantih slike (glej sliko 2.5). Ti histogrami pa se lahko kasneje uporabijo kot učni primeri za SVM klasifikator.

Za vsak piksel na sliki filtrirani z Gabor filtrom tvorimo 8-bitno binarno labelo na podlagi 8 okoliških pikslov. Posameznim pikslom, ki vsebujejo okoliške bomo rekli centralni. Velikost centralnega piksla primerjamo z velikostjo okoliških. Tem okoliškim  $f_p$ , nato spremenimo vrednosti v binarne relativno na centralni piksel  $f_c$ , kot prikazuje enačba (2.4).

$$S(f_p - f_c) = \begin{cases} 1 & \text{if } f_p < f_c, \\ 0 & \text{drugače} \end{cases} \quad (2.4)$$



Slika 2.5: LGBP proces tvorbe histogramov

$$LBP = \sum_{p=0}^7 S(f_p - f_c) * 2^p \quad (2.5)$$

Te z enačbo (2.5) pretvorimo v desetiški sistem in iz njih dobimo histograme za števila od 0 do 255. Pri prepoznavanju vzorcev za določeno emocijo, se je potrebno osredotočiti le na najbolj pomembne vzorce, hkrati pa je z ohranjanjem informacije dobro tem bolj zmanjšati količino informacij. Za ta namen se uporabijo t.i. uniformni vzorci. LBP je uniformen, če njegov binarni vzorec vsebuje do največ dve bitni tranziciji iz 0 v 1 in obratno. Za primer so števila 00000000 (0 tranzicij), 01110000 (2 tranziciji) in 11001111 (2 tranziciji) uniformna, nakar 11001001 (4 tranzicije) in 01010011 (5 tranzicij) niso. Pri tem ima histogram ločen stolpec za vsak uniformen vzorec in vsi neuniformni so združeni v svojem. Ojala [10] je opazil, da pri njihovih eksperimentih s teksturnimi slikami uniformni vzorci za 8 okoliških pikslov vsebujejo 90% vseh vzorcev. Namen uporabe Gabor filtra je, da ko bomo kasneje slike rotirali se bo ohranjala prostorska informacija o detektiranih robovih našega obraza. Frekvenca in orientacija pa sta v Gabor filtrih podobna kot pri človeškem vizualnem sistemu in sta posebej primerni za reprezentacijo tekstur in za razne diskriminacije. LGBP metoda opisana v [11] izkoristi vse kvadrante na sliki, naš pristop pa uporablja le kvadrante v okolici oči in ust, saj tako dobimo boljšo natančnost pri napovedovanju in hitrejšo pripravo podatkov. Ko pridobimo histograme po kvadrantih, le-te združimo v en niz, ki predstavlja en učni primer za klasifikacijo s SVM. LGBP algoritem

je prikazan v algoritmu 3.

---

**Algorithm 3** Matlab koda za LGBP.

---

```
mask = gabormask([1 8],sigma);  
imgb = conv2(im,mask); % izvedemo konvolucijo med filtrom in sliko  
u = uniform(imgb); % izracunamo uniformne vzroce  
q = quadrants(u); % izluscimo kvadrante  
h = hist(q); % izracunamo histograme
```

---

## 2.6 Aktivni videzni model - AAM

AAM je nadgradnja ASM, in poleg znamenitosti uporablja tudi podatke o teksturi. ASM predstavlja samo podatke o obliki, AAM pa še o teksturi. Zaradi te delitve bomo lahko kasneje s spreminjanjem koeficientov spreminjali le obliko, le teksturo, ali oboje našega razvitega modela. Vsaki emociji bo dodeljen svoj model, saj bodo tako podatki bolj enolično definirani in lažje napovedljivi. Vsak model pa ima svojo specifično obliko, ki bo predstavljala to emocijo. Iz slike 1.2 so razvidne razlike od povprečja, in iz grafa lahko sklepamo, da je razlika med obraznimi izrazi sicer težka, a vendarle tudi možna. V naslednjih korakih, je opisan potek dela potreben za implementacijo ASM modela:

1. povprečimo vse istoležne znamenitosti za izbrano emocijo (dobimo povprečeno obliko),
2. vsako posamezno znamenitost odštejemo od povprečene oblike (dobimo odstopanja od povp. oblike),
3. izvedemo kompresijo s PCA metodo (zmanjšamo količino parametrov za obliko).

Prej smo v poglavju o PCA omenili tudi normalizacijo s povprečjem. Ta normalizacija je izvedena v točkah 1 in 2. Če želimo ustvariti ASM iz slik za npr. eno emocijo, moramo izračunati povprečno obliko, pozicije znamenitosti teh slik odštejemo od povprečne oblike in jih preoblikujemo, tako da bodo vsi podatki v eni vrstici. Z uporabo PCA zmanjšamo dimenzijo podatkov in če za malo spremenimo koeficient PCA dobimo spremenjeno obliko obraza. Rekonstrukcija nazaj v množico znamenitosti poteka z uporabo enačbe

$$x = \bar{x} + P_s b_s, \quad (2.6)$$

kjer spremenljivka  $\bar{x}$  predstavlja povprečno obliko,  $P_s$  koeficiente in  $b_s$  uteži PCA. Ta pristop je koristen predvsem zaradi kompresije velike količine podatkov in lahke manipulacije s temi podatki. Sliki 2.6 in 2.7 prikazujeta, kako z manipulacijo koeficientov oblike preidemo iz originalne v spremenjeno obliko. Če bi hoteli enako storiti brez PCA, bi morali za vse obrazne slike, torej vsako teksturo in pozicijo znamenitosti posebej spremeniti za željeno vrednost. Ta proces je precej potraten in je lažje spreminjati le tabelo vrednosti koeficientov.

---

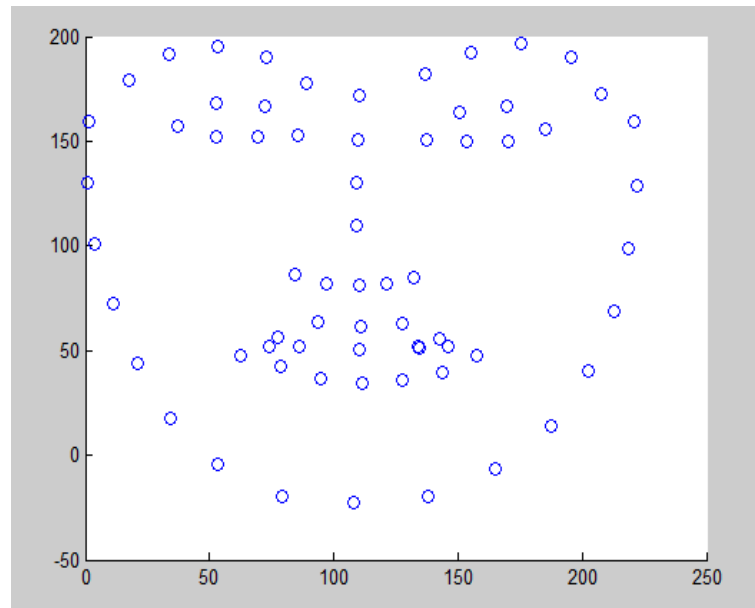
**Algorithm 4** Matlab koda za ASM
 

---

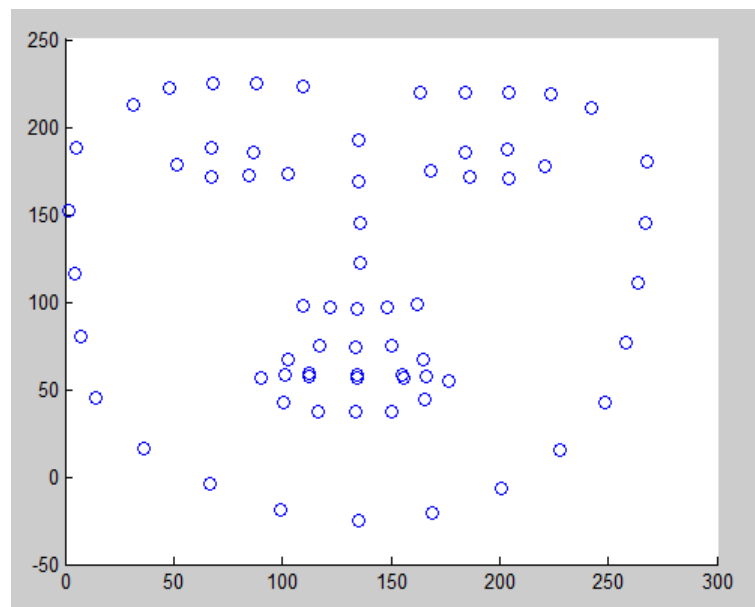
```
shapem_1 = asm_mean(shapes_1, images_1); % povprečenje oblik;
shapes1_pca = shapes_4_pca(shapes_1, shapem_1); %preoblikujemo
    tabelo v velikosti [st_oblik, X*Y ], kjer sta X in Y velikost
    dvodimenzionalne tabele
[coeff_shapes_1, score_shapes_1, latent_shapes_1] =
    pca(shapes1_pca); %ASM: z oblikami manipuliramo tako, da
    spreminjamo koeficient coeff_shapes_1
```

---

Zelo podoben postopek se uporablja pri AAM, kjer najprej izvedemo normalizacijo s povprečjem na naših teksturah. Za kreiranje povprečne teksture je potrebno vse slike za posamezno emocijo prenesti na prej omenjeno povprečno obliko. Ta postopek se imenuje zvijanje (angl. warping). Pri tem se uporabi Delaunayeva triangulacija za povezovanje znamenitosti v sliki neke



Slika 2.6: Oblika originalnih znamenitosti.



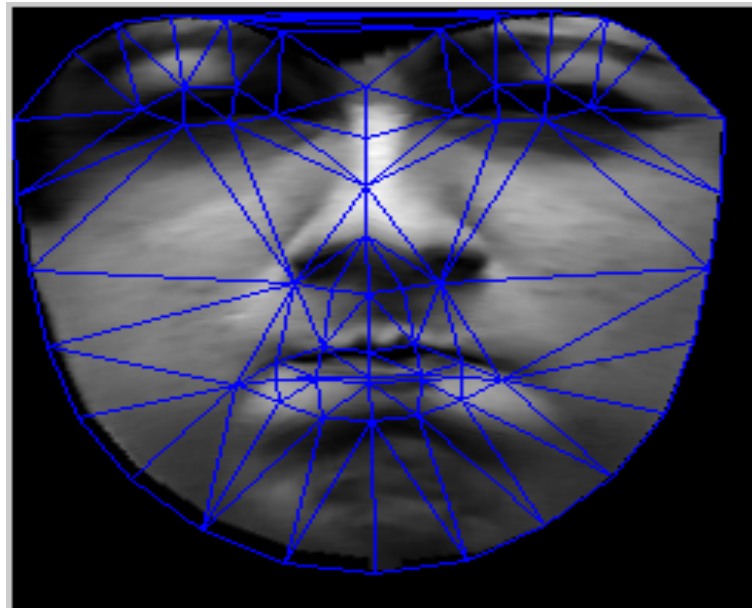
Slika 2.7: Oblika znamenitosti s spremenjenimi koeficienti.

emocije in nato posamezne piskle, ki so znotraj trikotnika prenese na trikotnike povprečne oblike. Rezultat Delanuayeve triangulacije je razviden iz slike 2.8. Ta prenos pikselov iz enega trikotnika na drugega opravimo z uporabo Baricentričnih koordinat, ki zagotovijo, da se preneseni piksel vedno nahaja v domeni trikotnika povprečne oblike. Proces zvijanja in celoten AAM, delujeta nekako takole:

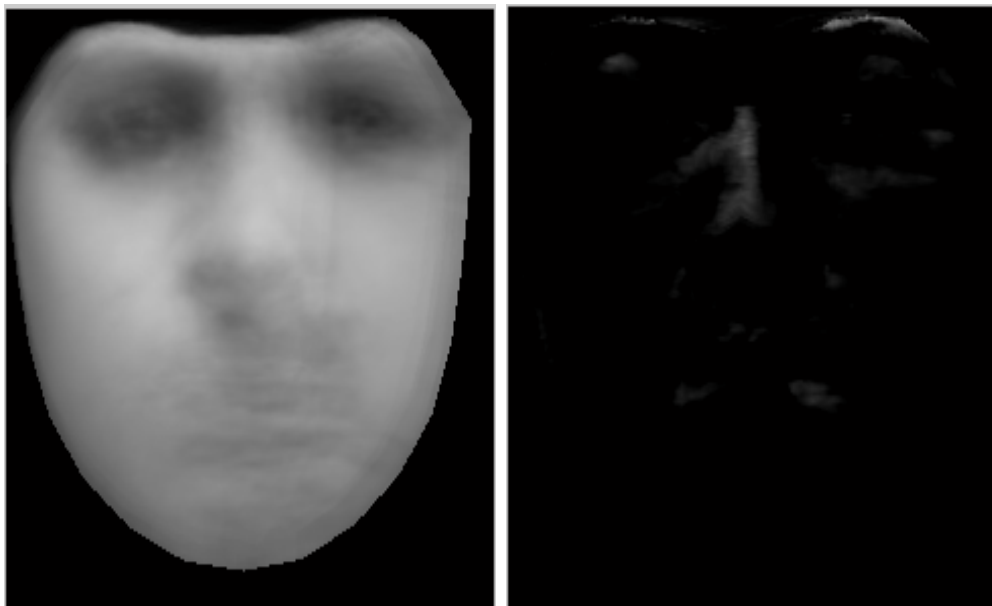
1. izvedemo Delaunayev triangulacijo nad povprečjem (bližnje povprečne znamenitosti povežemo v trikotnike),
2. izvedemo Delaunayev triangulacijo na znamenitostih naših testnih slik,
3. z baricentričnimi koordinatami piksele znotraj trikotnikov iz druge točke preslikamo v trikotnike iz prve točke (proces zvijanja),
4. gornji postopek zvijanja na povprečje izvedemo z vsemi slikami izbrane emocije in izračunamo povprečno teksturo,
5. teksture zbranih slik odštejemo od povprečja (dobimo odstopanja od povprečne teksture),
6. izvedemo kompresijo s PCA (zmanjšamo količino parametrov),
7. koeficiente ASM in AAM združimo in spravimo čez tretji PCA (s koeficienti lahko spreminjamo obliko in teksturo).

Pri tem je pomembno, da se najprej izračuna povprečje vseh slik in se to povprečje najprej zvije na povprečno obliko, ki je rezultat ASM modela. Razlog za to je predvsem v tem, da bomo kasneje slike odšteli od tega povprečja kot je prikazano na sliki 2.9, in nad temi razlikami izvedli drugi PCA. Rekonstrukcija iz koeficientov PCA nazaj v originalno teksturo poteka, tako kot pri oblikah z enačbo

$$x = \bar{x} + P_g b_g, \quad (2.7)$$



Slika 2.8: Delaunay triangulacija.



Slika 2.9: Na levi je povprečje na desni pa razlika.

le da tokrat  $\bar{x}$  predstavlja povprečene teksture zvite na povprečno obliko za posamezno emocijo  $i$ ,  $P_g$  koeficiente in  $b_g$  uteži PCA.

Ko imamo odstopanja tekstur od povprečne v zmanjšani dimenziji, koeficiente teksture združimo s koeficienti oblike. Ker je sestavljen nabor koeficientov ponovno velik spet uporabimo PCA za redukcijo njihove dimenzije. S pridobljenimi koeficienti spreminjamo tako obliko kot teksturo. Celoten proces od zvijanja do prehoda v originalno dimenzijo je opisan v algoritmu 5. S spreminjanjem koeficientov teksture in oblike bomo z modelom dobili veliko število variacij opazovane slike, na katerih bomo izvedli LGBP. Ta nam bo vrnil histograme, ki so potrebni za končno napovedovanje na SVM klasifikatorju.

---

**Algorithm 5** Matlab koda za AAM.

---

```
w1 = warp(images_1,shapes_1,shapem_1); %zvijemo na povprečno
    obliko
greyscalesm_1 = aam_mean(w1); %izračunamo povprečno teksturo
greyscales1_pca = greyscales_4_pca(w1,greyscalesm_1,im_size);
    %pripravimo podatke za PCA
[coeff_greyscales_1,score_greyscales_1,latent_greyscales_1] =
    pca(greyscales1_pca);
coeff_param= coeff_param + 0.005; %spreminjamo koeficiente
coeffs = score_param*coeff_param'; %rekonstrukcija podatkov AAM+ASM
temp_grey = score_greyscale * coeffs(:,137:end); %rekonstrukcija
    teksture (1:137 predstavlja obliko)
reversed_image_diff = reshape(greyscalesm_1,im_size(1),im_size(2));
```

---

## 2.7 Združitev LGBP in AAM

Torej pridobili smo LGBP histograme in možnost spreminjanja oblike in teksture z AAM za vsako sliko. Vrednosti nekih realnih slik ne moremo natančno predvideti, vendar lahko v ta namen skušamo pridobiti najboljši približek tej sliki. Nad sliko je bil izveden AAM algoritem in pridobili smo veliko količino opazovanih slik, ki so bile z različnimi oblikami in teksturami le variacije originalnih slik. Koeficiente za spreminjanje teh lastnosti sem določil eksperimentalno, za dodatno povečanje testnih primerov pa sem sliko rotiral za pet stopinj v levo in desno. Te slike sem nato popeljal skozi LGBP in s SVM klasificiral kreirane histograme.

V naših testnih primerih ne vemo za katero emocijo gre, zato je vseeno ali je sama slika zvita na obliko strahu ali na obliko sreče. Tu se je zgradil en model za vse oblike emocij, in se nato ta model poljubno spreminjal. Tako smo dobili tudi veliko večjo testno množico in tudi izboljšali kvaliteto napovedovanja.

## Poglavje 3

### Rezultati

Pri LGBP sem izmed 327 obraznih slik uporabil 21 za testno množico in 306 za učno množico. Kot že povedano, potrebujemo za napovedovanje s SVM veliko učno množico. Vsako izmed slik sem rotiral za deset stopinj v vsako stran v koraku po 0.5 stopinj. Tako sem pridobil 12240 slik, te so bile velikosti 42x42. Iz slik izluščil le kvadrante, ki vsebujejo največ informacij o emociji, torej oči in usta. Te slike sem spravil skozi LGBP in histograme napovedal s prej omenjenim klasifikatorjem.

Rezultati klasifikacije so prikazani v tabeli 3.1. Anomalija pri teh rezultatih je vsekakor obrazni izraz presenečenja, saj vizualno ni ravno podoben preziru ali gnusu. Pri ostalih pa bi lahko opravičili odločitev SVM, saj so si negativni obrazni izrazi jeze, gnusa, strahu in žalosti precej podobni. V uvodnem delu smo omenili, da lahko nasmejan obraz uporabimo kot obrambni mehanizem v neprijetni situaciji in ravno to prikazujejo rezultati klasifikacije sreče.

Za namen napovedovanja podatkov, ki so posledica združitve LGBP in AAM, sem uporabil SVM model kreiran v predhodnem napovedovanju. Za testno množico sem izbral eno sliko za vsako emocijo, te slike pa so bile izvzete iz bolj spontanij npr. predsedniških volitev ali televizijskih novic. Sam model je pri vseh preferiral emocijo sreče kjer je sliko, ki vsebuje srečno emocijo napovedal z nad 90% natančnostjo, nakar vse ostale emocije uvrstil

v pravi razred z 10% natančnostjo ali manj. Razlog za tovrstno maržo napake pa se vsekakor skriva v eksperimentalnih vrednostih sprememb koeficientov PCA.

Emocija	Jeza	Prezir	Gnus	Presenečenje	Sreča	Žalost	Strah
Jeza	<b>21.43</b>	0.00	64.29	0.00	0.00	7.14	7.14
Prezir	0.00	<b>66.67</b>	6.67	20.0	0.00	6.67	0.00
Gnus	12.50	0.00	<b>50.00</b>	0.00	0.00	12.50	25.00
Presenečenje	0.00	40.00	40.00	<b>20.00</b>	0.00	0.00	0.00
Sreča	10.42	6.25	12.50	20.83	<b>29.17</b>	10.42	10.42
Žalost	0.00	0.00	40.00	0.00	0.00	<b>60.00</b>	0.00
Strah	24.00	0.00	4.00	4.00	4.00	16.00	<b>48.00</b>

Slika 3.1: Razvrstitev testnih primerov za LGBP.

# Poglavje 4

## Sklepne ugotovitve

Začetna ideja je bila razviti spletno storitev, ki bo uporabnikom nudila avtomatsko prepoznavo emocije iz obraznih slik. Če bi želeli sam proces prepoznave avtomatizirati, bi bila potrebna avtomatska prepoznavna znamenitosti. Za začetek bi uporabil spletno storitev, ki prepozna par ključnih znamenitosti imenovana Betaface. Seveda bi bila ta soodvisnost za potencialno dolgotrajno aplikacijo neprimerna in bi tu bila bolj primerna lastna raziskava na tem področju. Pri samem testiranju so bili uporabljeni tudi realni podatki, vendar rezultat klasifikacije ni vreden omembe. Sumim, da je za bolj natančnejšo analizo realnih podatkov potrebna predvsem večja podatkovna zbirka. CK+ baza [9] je kar specifična in iz slik so emocije zelo lepo razvidne, kar je seveda slabo, če imaš v testnih oz. realnih slikah popačeno emocijo oz. sliko oslABLJENO z nekim šumom, ki ga lahko povzroči le slaba osvelitev. Poleg tega število slik na emocijo ni poenoteno in imamo npr. za gnus 59 slik slik, za strah pa le 25, kar se odraža pri preferiranju ciljnega razreda v procesu napovedovanja. Za nadaljnjo raziskavo predlagam naslednji delovni tok:

Ker so videi le skupek slik, bi verjetno lahko AAM in LGBP model uporabili tudi v tem mediju. Seveda sama ideja o prepoznavi emocije je dokaj suhoparna, zato bi lahko samo aplikacijo naredili bolj živo in bi glede na stanje slikane osebe vrnil kak spodbujevalni izraz. Za piko na i bi tovrstno

- priprava večje količine podatkov, vključno z ročno dodeljenimi znamenitostmi (avtomatiziran proces),
- dodati AAM iskanje z linearno regresijo,
- detekcija anomalij in izbira atributov na podlagi neke hevristične ocene,
- združitev AAM z LGBP,
- preizkus različnih metod prepoznave obraznih lastnosti.

spletno storitev uporabil v npr. Lego Mindstrom robotu, ki bi te lahko spravil v dobro voljo ali celo motiviral.

Vsem raziskovalcem, ki bi se radi pogolobili v metodo modeliranja slik, priporočam v branje [13], saj zelo izčrpno opiše cel proces kreiranja modela. Poleg same teorije, opiše še par primerov uporabe v medicini npr. prepoznavo objektov iz magnetno-resonačnih slik. Za kontemplacijo LGBP metode pa je prava izbira [12], saj jasno in kratko opiše predlagano metodo. Obe metodi lahko ločeno uporabljamo za prepoznavo obraznih izrazov, Senechal pa je dobre lastnosti obeh, združil v aplikacijo opisano v [11].

# Literatura

- [1] R. Adolphs. Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, 1(1):21–62, 2002.
- [2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape model - their training and applications. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.
- [3] T.F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 484–498. Springer, 1998.
- [4] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(10):979–984, 1999.
- [5] G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face recognition using the active appearance model. *5th European Conference on Computer Vision*, pages 1–20, 1998.
- [6] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen. Local binary patterns and its application to facial image analysis. *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 41(4):4–17, 2011.
- [7] Igor Kononeko and Marko Robnik Šikonja. *Inteligentni sistemi*. Založba FE in FRI, 2010.

- 
- [8] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.
- [9] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended Cohn-Kanade dataset (CK+): A complete dataset of action unit and emotion-specified expression. pages 94–101, San Francisco, USA, 2010.
- [10] T. Ojala, M. Pietikainen, and T. Aenpaa. Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [11] T. Senechal, V. Rapp, H. Salam, R. Segquier, K. Bailly, and L. Prevost. Combining aam coefficients with LGBP histograms in the multi-kernel SVM framework to detect facial action units. *Proc. IEEE Intl Conf. Automatic Face and Gesture Recognition*, pages 860–865, 2011.
- [12] C. Shan, S. Gong, and P. W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, 2009.
- [13] M. B. Stegmann. Active appearance models: Theory, extensions and cases. Masters thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2000.
- [14] Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(2):762–767, 2001.
- [15] T. P. Weldon, W. Higgins, and D. Dunn. Gabor filter design for multiple texture segmentation. *Optical Engineering*, 35(10):2852–2863, 1996.