

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Gregor Žerovnik

**Slikovna podatkovna baza tekstov v
naravnih scenah**

DIPLOMSKO DELO
NA VISOKOŠOLSKEM STROKOVNEM ŠTUDIJU

Mentor: doc. dr. Peter Peer
Asistent: mag. Andrej Ikica

Ljubljana, 2010



Št. naloge: 00522/2010

Datum: 05.04.2010

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

Kandidat: **GREGOR ŽEROVNIK**


Naslov: **SLIKOVNA PODATKOVNA BAZA TEKSTOV V NARAVNIH SCENAH**
IMAGE DATABASE OF TEXTS IN NATURAL SCENES

Vrsta naloge: Diplomsko delo visokošolskega strokovnega študija

Tematika naloge:


V nalogi najprej preučite dosegljive slikovne podatkovne baze za potrebe računalniškega vida, s poudarkom na bazah za razpoznavo teksta. Nato preučite metodologije zajema slik, gradnje baze, v reprezentativnih bazah. Potem zasnujte lastno metodologijo za bazo FRI OCR ter ustvarite bazo. Na primeru razpoznave pokažite tudi uporabnost nove baze.

Mentor:


doc. dr. Peter Peer



Dekan:


prof. dr. Nikolaj Zimic

Rezultati diplomskega dela so intelektualna lastnina Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavlanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje Fakultete za računalništvo in informatiko ter mentorja.

Besedilo je oblikovano z urejevalnikom besedil \LaTeX .

Namesto te strani **vstavite** original izdane teme diplomskega dela s podpisom mentorja in dekana ter žigom fakultete, ki ga diplomant dvigne v študentskem referatu, preden odda izdelek v vezavo!

IZJAVA O AVTORSTVU

diplomskega dela

Spodaj podpisani Gregor Žerovnik,

z vpisno številko 63020324,

sem avtor diplomskega dela z naslovom:

Slikovna podatkovna baza tekstov v naravnih scenah

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom doc. dr. Petra Peera
- 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, dne 14. 10. 2010

Podpis avtorja:

Zahvala

Zahvalil bi se svojemu mentorju doc. dr. Petru Peeru za samo idejo diplomske naloge ter za konstantno pomoč pri njeni izdelavi. Zahvala gre tudi mag. Andreju Ikici za izčrpno pomoč in nasvete predvsem pri praktičnem delu diplomske naloge.

Kazalo

| | |
|--|----------|
| Povzetek | 1 |
| Abstract | 2 |
| 1 Uvod | 3 |
| 1.1 Slikovne podatkovne baze | 3 |
| 1.2 Metodologija zajema | 4 |
| 2 Slikovne baze v računalniškem vidu | 5 |
| 2.1 Slikovne baze za optično razpoznavo znakov | 5 |
| 2.1.1 NIST Special Database 2 | 6 |
| 2.1.2 Arabase | 9 |
| 2.1.3 InftyCDB-1 | 11 |
| 2.1.4 Primerjava baz | 13 |
| 2.2 Podatkovne baze slik obrazov | 14 |
| 2.2.1 Standardne slike obrazov | 14 |
| 2.2.2 Termične slike obrazov | 15 |
| 2.2.3 3D slike obrazov | 15 |
| 2.2.4 Primerjava baz | 17 |
| 2.3 Podatkovne baze slik šarenic | 18 |
| 2.3.1 Primerjava baz | 19 |
| 2.4 Podatkovne baze slik prstnih odtisov | 20 |
| 2.4.1 3D slike prstnih odtisov | 21 |
| 2.4.2 Primerjava baz | 21 |
| 2.5 Podatkovne baze slik podpisov | 22 |
| 2.5.1 Primerjava baz | 23 |
| 2.6 Podatkovne baze slik odtisa dlani | 23 |
| 2.6.1 Primerjava baz | 24 |
| 2.7 Podatkovne baze slik gibanja | 25 |

| | | |
|----------|---|-----------|
| 2.7.1 | Primerjava baz | 26 |
| 2.8 | Podatkovne baze slik ušes | 26 |
| 2.8.1 | Primerjava baz | 27 |
| 2.9 | Ostale podatkovne baze | 27 |
| 3 | Metodologija zajema | 28 |
| 3.1 | Preučevanje metodologije drugih baz | 28 |
| 3.1.1 | CASIA-IrisV3 | 28 |
| 3.1.2 | FERET | 29 |
| 3.1.3 | NIST Special Database 6 | 32 |
| 3.2 | Metodologija zajema podatkovne baze CVL OCR DB | 36 |
| 3.2.1 | Primerjava metodologije zajema baze CVL OCR DB z ostalimi OCR bazami | 39 |
| 3.3 | Rezultati in ugotovitve | 39 |
| 4 | Primer detekcije in razpoznavne teksta v naravnih scenah | 42 |
| 4.1 | Detekcija teksta v naravnih scenah | 42 |
| 4.2 | Razpoznavna teksta na slikah naravnih scen | 43 |
| 4.2.1 | Klasični OCR | 43 |
| 4.2.2 | OCR na slikah naravnih scen | 43 |
| 4.3 | Primer detekcije in razpoznavne teksta | 44 |
| 5 | Zaključek in nadaljne delo | 47 |
| | Seznam slik | 49 |
| | Seznam tabel | 51 |
| | Literatura | 52 |

Seznam uporabljenih kratic in simbolov

OCR - (angl. Optical Character Recognition) optična razpoznavna znakov

CVL OCR DB - (angl. Computer Vision Laboratory Optical Character Recognition DataBase) podatkovna baza za optično razpoznavo znakov laboratorija za računalniški vid na Fakulteti za računalništvo in informatiko, Univerze v Ljubljani

NIST - (angl. National Institute of Standards and Technology) Nacionalni inštitut standardov in tehnologij

ARABASE - (angl. A Relational Arabic dataBASE) relacijska podatkovna baza arabskih znakov

ASCII - (angl. American Standard Code for Information Interchange) Ameriški standard za izmenjavo informacij

FRVT - (angl. Face Recognition Vendor Test) razpoznavne obrazov komercialnih sistemov

CASIA - (angl. Chinese Academy of Sciences, institute of Automation) Kitajska akademija znanosti, Inštitut za avtomatiko

FERET - (angl. Face REcognition Technology) tehnologija razpoznavanja obrazov

FBI - (angl. Federal Bureau of Investigation) Državni urad za preiskave

FVC - (angl. Fingerprint Verification Competition) tekmovanje v razpoznavi prstnih odtisov

MIT - (angl. Massachusetts Institute of Technology) tehnološki inštitut Massachusetts

ANN - (angl. Artificial Neural Network) umetna nevronska mreža

SVM - (angl. Support Vector Machine) metoda podpornih vektorjev

K-NN - (angl. K-Nearest Neighbour) klasifikacija na principu najbližjih sosedov

Povzetek

Obstaja veliko podatkovnih baz slik. Ključno jih je imeti dobro urejene ter kategorizirane, za kar poskrbi ustrezna metodologija zajema gradnje baze. V diplomskem delu sem se posvetil predvsem bazam s področja računalniškega vida, metodologijam zajema ter zgradil našo podatkovno bazo slik CVL OCR DB. V uvodnem delu sem na kratko opisal osnovna pojma podatkovna baza ter metodologija. V osrednjem delu sem predstavil bolj znane slikovne baze, ki se uporabljajo na področju računalniškega vida. Največji poudarek je na bazah za optično razpoznavo znakov (angl. OCR). Po opisu sem jih med sabo tudi primerjal. Zatem sem se posvetil sami metodologiji zajema baz, izbrane tudi opisal ter med sabo zopet primerjal. Sledi predstavitev naše slikovne baze CVL OCR DB, katero sem tudi primerjal z ostalimi bazami na področju optične razpoznavne znakov. Zaključni del pa je namenjen ilustraciji primera uporabe, analizi rezultatov ter nadaljnem delu.

Ključne besede:

metodologija zajema podatkovnih baz slik, struktura, kategorije, atributi, računalniški vid, optična razpoznavna znakov, baza CVL OCR DB

Abstract

There are many image databases, but the key part is having them well structured and organized. A well formed database assembly methodology can take care of exactly that. In this diploma thesis I enlightened the databases from the computer vision area and their methodologies and created our own image database called CVL OCR DB. In the introductory part I briefly described the basics of databases and methodology. In the middle part I presented the most important image databases from the majority of computer vision subareas with the emphasis on the optical character recognition (OCR) databases. After their description I also compared them between each other. After that I focused on the methodology of the chosen databases, which were also described and compared between each other. Following that I presented our own database CVL OCR DB, which was compared to other OCR databases. The final part focuses mostly on the usage illustration, analysis of results and on the future work.

Key words:

image database methodology, structure, categories, attributes, computer vision, optical character recognition, CVL OCR DB

Poglavje 1

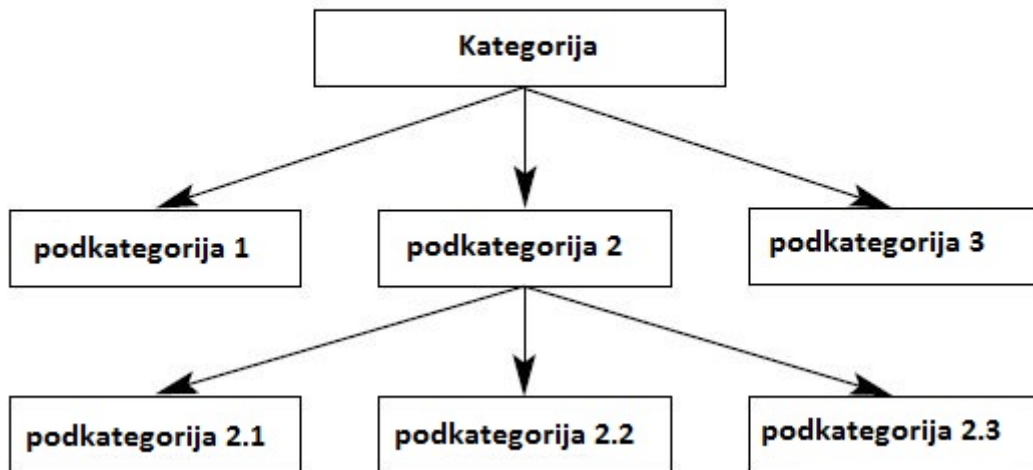
Uvod

V današnjem svetu čutimo porast potrebe po varnosti, tako fizični kot varnosti informacij in podatkov. Za varnost v digitalnem svetu med drugimi skrbijo mnogi biometrični sistemi, ki zagotavljajo avtomatično razpoznavanje posameznika na podlagi njegovih unikatnih značilnosti. Bodisi preko prstnih odtisov, šarenice, obraza, odtisov dlani itn. Prav tako pa lahko zagotavljajo tudi razpoznavanje besedila oz. znakov iz slik najrazličnejših besedil. Na tem področju se nahaja na primer tudi razpoznavanje elektronskih podpisov. Dejstvo pa je, da vsi takšni sistemi za svoje delovanje zahtevajo ogromne baze slik, pridobljenih na različne načine. Te baze morajo biti ustrezno urejene in kategorizirane, za kar pa skrbi metodologija zajema baz. V diplomskem delu nas vidik varnosti ne bo toliko zanimal. Zanimale pa nas bodo slikovne podatkovne baze in njihova metodologija.

1.1 Slikovne podatkovne baze

So organizirane zbirke slik, ki se lahko uporabljajo za več namenov. V grobem sestojijo iz treh nivojev. In sicer: zunanji, konceptualni in notranji nivo. Zunanji nivo je tisti, ki definira, kako uporabniki razumejo organizacijo podatkov. Posamezna baza ima lahko mnogo pogledov zunanjega nivoja. Notranji nivo definira, na kakšen način so podatki shranjeni in procesirani. Skrbi za stroške, optimalno delovanje, skalabilnost in ostale operacijske stvari. Konceptualni nivo pa predstavlja povezavo med zunanjim in notranjim nivojem. Priskrbi skupni pogled baze na najbolj osnoven način ter tako združi mnoge zunanje poglede v neko koherentno celoto. Prav tako obstaja tudi veliko modelov strukturiranja baz oz. samega prikazovanja baz. Kot so na primer: tabelarični model, hierarhični model (glej sliko 1.1), mrežni model,

relacijski model itn. V diplomskem delu je struktura baz največkrat prikazana prav s hierarhičnim modelom oz. večkrat imenovanim kar drevesnim modelom.



Slika 1.1: Prikaz hierarhičnega modela baze.

1.2 Metodologija zajema

Glede slikovnih baz v sklopu računalniškega vida (s poudarkom na področju optične razpoznavne znakov) nas najbolj zanima njihova metodologija zajema, oz. zasnova samih atributov baz ter posledično tudi njihovih kategorij in podkategorij na katere se razdeli. Pri samem zajemanju slik pride namreč namenoma (včasih tudi nenamenoma) do različnih situacij kot so na primer: slikanje pod kotom, razlike v osvetlitvi, različne kamere, zamaknjene slike itn. Metodologija pa načeloma zasnuje opis potrebnih kategorij za reševanje teh specifičnih problemov (v našem primeru pri računalniškem vidu). Torej definira vse pogoje zajemanja, da je rekonstrukcija pogojev čim lažja. Za vsako bazo je metodologija ključnega pomena, saj je ustrezna določitev kategorij pomembna za kakršnokoli interakcijo z njo, bodisi s strani samega algoritma, ki jo uporablja, bodisi s strani samega vzdrževanja in dopolnjevanja ter same gradnje baze.

Poglavje 2

Slikovne baze v računalniškem vidu

2.1 Slikovne baze za optično razpoznavo znakov

Optična razpoznavna znakov (angl. Optical Character Recognition - OCR), se ukvarja s prenosom skeniranih slik pisanega, tipkanega ali tiskanega teksta v strojno-kodirani tekst. Je široko uporabljen za pretvorbo dokumentov, knjig in drugih oblik teksta v elektronske datoteke z namenom digitalizacije arhiva v pisarni, objave teksta na spletu ipd. OCR tako omogoči urejanje teksta, iskanje zelenih besed ali fraz, bolj urejen arhiv, tiskanje brez pomoči raznih optičnih čitalcev in apliciranje tehnik kot so strojno prevajanje, strojno branje, ipd. OCR je področje raziskav v domeni razpoznave vzorcev, umetni inteligenci in, kar je za nas najbolj pomembno, v računalniškem vidu.

Tri glavna področja OCR-ja, ki so največkrat v bazah kar specializirane podatkovne množice so:

- Indirektna ročna pisava je ustvarjena tipično s pisanjem s pisalom na list, nato skenirana in digitalizirana. Običajno vsebuje izolirane alfanumerične znake ali besede. Pri tem področju je pomembno upoštevati, da je oblika podatkovne baze odvisna od tega, kako dobro subjekti, ki producirajo podatke in pogoje, pod katerimi so zbrani, dejansko predstavljajo samo aplikacijo. Idealno bi subjekti morali biti izbrani iz iste populacije in podatki zbrani pod enakimi pogoji.
- Direktna ročna pisava je pisana direktno na digitalizirano tablico z elektronskim pisalom (izhod je zaporednje x in y koordinat, ki izražajo pozi-

cijo pisala, kot tudi druge informacije, na primer moč pritiska s pisalom. Vsebuje tudi izolirane znake in besede, prav tako pa tudi tekstovne fraze.

- Strojno tiskani tekst pa se pojavlja pogosto v dnevni uporabi in je ustvarjen največkrat z zunanjimi procesi, laserskimi, toner ali igličnimi tiskalniki. Strojno tiskane tekstovne podatkovne baze največkrat vsebujejo slike besed ali kar celotnih strani. Te strani so lahko zbrane iz različnih virov, varirajo pa tudi v formatih (pisava, razporeditev vrstic, velikost pisave ipd.)

Vsako področje ima unikatne karakteristike, ki vplivajo na obliko podatkovne baze. Te karakteristike določajo kako dobro se bodo lahko eksperimentalni rezultati pri dotični podatkovni bazi prenesli tudi na drugo aplikacijo.

Za primerjavo sem si izbral tri baze, in sicer NIST Special Database 2 [1], ki se ukvarja s slikami davčnih obrazcev, bazo ARABASE [2], ki sem jo vključil kot zanimivost, saj se ne ukvarja z latinskimi črkami, temveč s pisano in tiskano arabsko pisavo in pa bazo matematičnih znakov in simbolov InftyCDB-1 [3]. Vse baze so opisane v spodnjih poglavjih, na koncu pa so med sabo tudi primerjane.

2.1.1 NIST Special Database 2

Osnovni podatki

Baza vsebuje binarne slike davčnih obrazcev (slika 2.1) in sicer 5.590 slik čez celo stran ter zavzema približno 610 MB prostora. Vsaka slika je shranjena v črno-belem formatu. Vnosna polja so bila avtomatsko generirana z računalnikom z namenom prikritja dejanskih osebnih podatkov. Vrednosti, vnešene v ta polja, se ločeno shranijo v ASCII tekst datoteko. Ta datoteka (ena na en obrazec) potem služi kot odgovorjajoča datoteka, ki se jo lahko uporabi za prirejanje vrednosti, ki so bile predlagane s strani sistema prepoznavanja. Glavne informacije baze so velikost, datum, tip, znaki ipd.

Hierarhija

Predstavlja hierarhični podatkovni model, v katerem so podatki predstavljeni v podobi drevesne strukture, kar je lepo vidno na sliki 2.2. Takšna struktura zelo lepo ponazori zgradbo baze in omogoča ponavljajoče informacije z uporabo razmerja oče/sin (oz. starš/otrok). Vsak oče ima lahko mnogo sinov, vsak sin pa ima lahko samo enega očeta. Sin brez naslednikov se smatra kot list drevesa (to zgradbo lahko razberemo iz slike 2.3).

Your first name and initial (if joint return, also give spouse's name and initial)
Suffolk U. & Taylor M. Ramsey

Present home address (number, street, and apt. no. or rural route) (if a P.O. box, see page 6 or 7)
25300 Early Road

City, town or post office, state, and ZIP code
Rockdale, HI 31807

Do you want \$1 to go to this fund? X Y
 If joint return, does your spouse want \$1 to go to this fund? X Y

1 X Single
 2 Married filing joint return (even if only one had income)
 3 Married filing separate return. Enter spouse's social security no. above and full
 4 Head of household (with qualifying person). (See page 7 of instruction
 your dependent, enter child's name here.
 5 Qualifying widow(er) with dependent child (year spouse died \geq 19

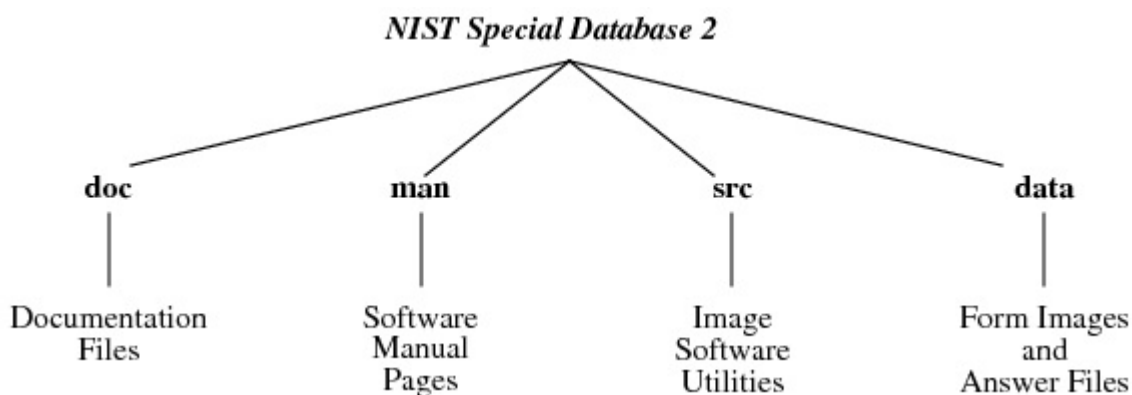
a Yourself If someone (such as your parent) can claim you as a dependent, \oplus
 But be sure to check the box on line 33b on page 2.

b Spouse

c Dependents:

| (1) Name (first, initial, and last name) | (2) Check if under age 5 | (3) If age 5 or older, dependent's social security number | (4) Religion |
|--|--------------------------|---|--------------|
| Alvin Evans | | A59:06:7960 | StPis |
| Piedmont Kingman | | A93:28:3784 | StPro |

Slika 2.1: Delna slika davčnega obrazca baze NIST Special Database 2.

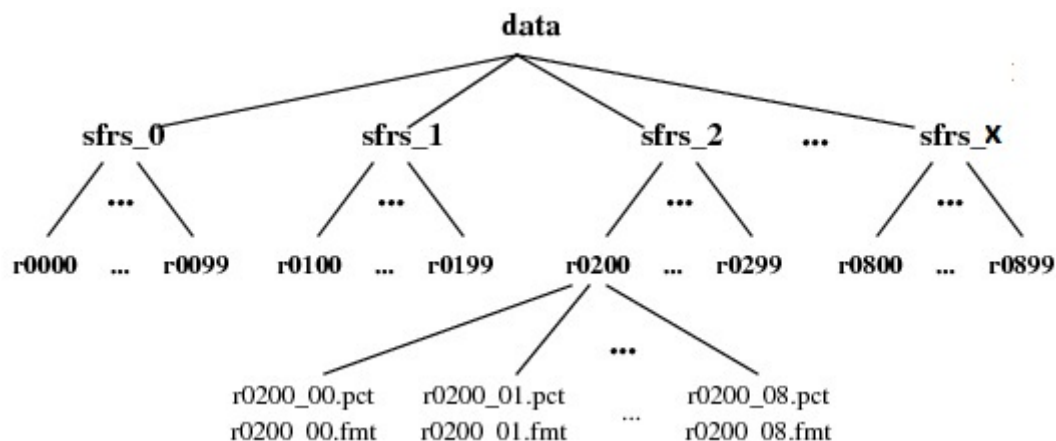


Slika 2.2: Prva stopnja drevesa v bazi. Direktorji doc, man in src vsebujejo dokumentacijo in pripomočke za manipulacijo slik (na primer izvorno kodo za dostopanje do same baze). Direktorij data, pa vsebuje datoteke slik in njihove informacije.

Dokumentacijske tabele vnosnih polj

Še zadnja informacija, ki jo nosi baza je zbirka dokumentacijskih tabel (slika 2.4, slika 2.5). Te tabele vsebujejo splošne informacije o vsakem vnosnem polju

na danem obrazcu. Informacije priskrbijo sistemski razvijalci in s tem vodijo proces prepoznavanja njihovega dokumentnega sistema. Tabele vsebujejo tip podatkov in kontekst povezan z vsakim vnosnim poljem najdenim na obrazcu.



Slika 2.3: Prikaz členitve drugega reda. Direktoriji z imeni sfrs-0...sfrs-X vsak vsebujejo po 100 slik (r0000-r0099) sestavljenih davčnih obrazcev. Direktoriji r0200 predstavljajo oddane vloge, ki se naprej delijo še na odrezke obrazcev ipd.

| TAG | DEFINITION |
|---------------------|--|
| A, CA | Alphanumeric Fields |
| F, FF, FP, FPER, FU | Floating Point Fields |
| I | Integer Fields |
| ICON | Non-Character Fields (box markings, signatures) |

Slika 2.4: Skupek možnih tipov vnosnih polj. (A, CA - alfanumerična polja; F, FF, FP, FPER, FU - polja decimalnih števil s plavajočo vejico; I - celoštevilčna polja; ICON - polja, ki ne predstavljajo znakov.

| TAG | DEFINITION |
|------|-------------------------|
| DATA | Generic Data |
| NAME | Names of People |
| SSN | Social Security Numbers |

Slika 2.5: Skupek možnih kontekstnih vnosnih polj. (DATA - generični podatki; NAME - imena ljudi; SSN - številke zavarovanja).

2.1.2 Arabase

Je baza za raziskovanje indirektno in direktno optične razpoznavo arabske ročne pisave kot tudi optične razpoznave strojno tiskanega teksta. Vsebuje digitalne slike dokumentov, tekstovne fraze, besede, podbesede (unikatne za arabski jezik), izolirane znake, številke, podpise ipd.

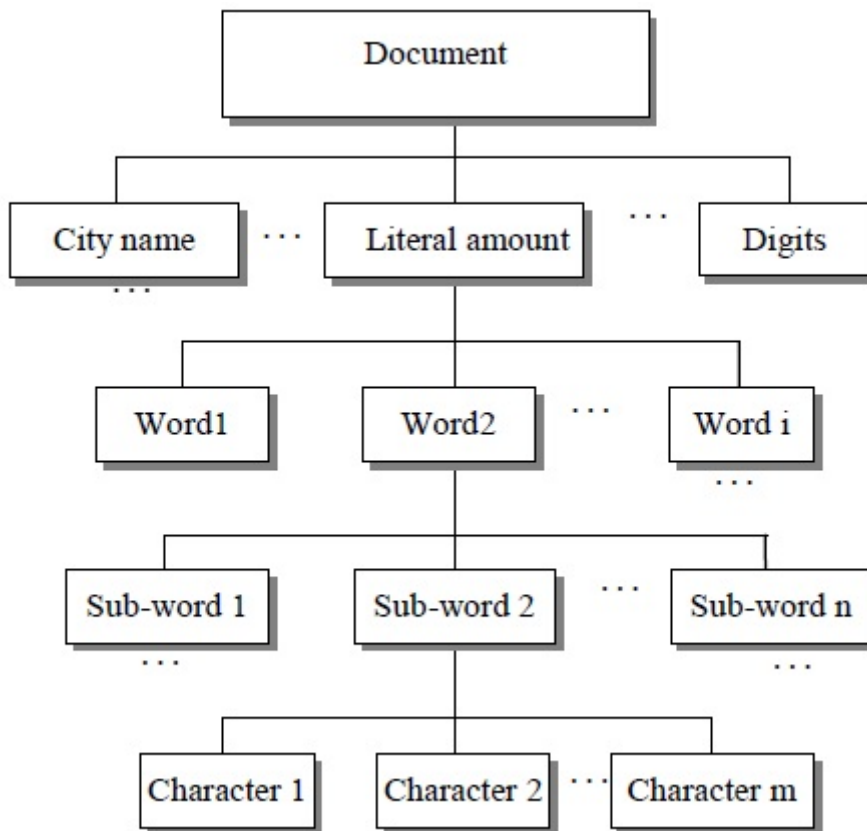
Zbiranje podatkov

Največji problem, ki nastane na tej stopnji je zbiranje podatkov iz resničnih dokumentov, da bi jih lahko uporabili za dejanske aplikacije kot so urejanje na pošti ali prepoznavanje pisanih zneskov v bančništvu. Arabase vsebuje širok spekter strojno tiskanih in ročno napisanih podatkov. Do sedaj je pri njihovem zbiranju pomagalo 400 ljudi, večinoma Tunizijcev. Uporabljeni so različni A4 obrazci z že vnaprej določenimi okenci, v katere mora subjekt napisati zgoraj (nad okencem) navaden tekst (številke, imena mest ipd.). Za ta način izpolnjevanja so dobili navdih pri podatkovni bazi UNIPEN[4]. Vsak izpolnjen obrazec je nato še ročno pregledan. Vključenih je tudi veliko slik teksta iz raznih časopisov in knjig, objavljenih s strani tunizijske nacionalne knjižnice. Prav tako je vključen tekst, ki je večkrat pretipkan z različnimi fonti, različnimi velikostmi in stiskan z različnimi stroji. Torej vse te natiskane slike in indirektno ročno napisane (razen tiste, ki so bile objavljene že elektronsko) so črno-belo skenirane z dvema optičnima čitalcema z ločljivostjo od 200 do 600 dpi.

Organizacija podatkov

Za vsako področje prepoznavanja (tiskano, indirektno in direktno pisano) je Arabase razdeljen na podbaze (slika 2.6), kjer vsaka predstavlja specifični del podatka (imena mest, številke ipd.). Vsaka podbaza pa se znajde v svojem direktoriju, ki vsebuje vzorce testnih slik (1/3) in vzorce učnih slik (2/3).

Veliko dela je bilo vložnega v razvoj konceptualnega modela baze. Naloga ni lahka zaradi tako raznolikih delov informacij in njihovih odnosov.



Slika 2.6: Struktura Arabasa prikazana okoli koncepta dokumenta, ki predstavlja najvišji nivo podatkov. Naslednje stopnje pa se ravnajo po padajočem segmentacijskem principu, sestavljajo pa ga deli teksta, besede, podbesede in znaki.

Dokument je opisan na grobo s štirimi atributi: referenčna koda (ime dokumenta), tip (tiskan, pisan indirektno in direktno), kvaliteta (stopnja degradacije 0-10) in kontekst (opis podatkov v dokumentu: imena mest ipd.). Vsebina in podvsebine so opisane z naslednjimi atributi: njihovo kodiranje, njihova slika ter njihova segmentirana slika. Sam namen baze ARABASE za raziskovanje prepoznavanja pisave je načeloma razdeljen na dva dela in sicer na učno fazo (učenje sistema) in na testno fazo. Baza bo z neprestano rastjo, eksperimentiranjem in dodajanjem novih kategorij, hitro prešla v eno največjih baz arabsko pišočega prebivalstva.

2.1.3 InftyCDB-1

Je baza okarakteriziranih (angl. ground truth) dejanskih matematičnih znakov in simbolov. Bazo sestavlja 467 strani tridesetih matematičnih člankov.[3] Je javna in pride najbolj do izraza, če se uporablja za naslednje namene:

- razvoj in ocenjevanje prepoznavne znakov in znanstvenih simbolov
- razvoj in ocenjevanje prepoznavne matematičnih formul
- analiza besed in matematičnih dokumentov

Lastnosti vsakega znaka/simbola so sestavljene iz tipa, pisave (italic/pokončna, odebeljena/neodebeljena), kvalitete (načeta, prelomljena), in povezave (relativna pozicija) itn. Slikovne informacije so shranjene ločeno v enotah za besede in enotah za matematične formule tako, da je lahko uporabljena kot slikovna baza besed ali slikovna baza matematičnih formul. Slikovne informacije so urejene po padajočem abecednem vrstnem redu neodvisno od vsebine. V bazi zaradi avtorskih pravic niso vključene slike celotnih strani. „Znak“ ne pomeni samo navadnih črk (na primer „A“), ampak tudi matematične simbole (na primer „+“), če ni prej drugače navedeno. „Kategorija“ pomeni najvišjo stopnjo klasifikacije znaka, „tip“ pa pomeni skupek kategorij s podobnimi lastnostmi. Na primer „A“, „B“ in „C“ so tri kategorije, ki pripadajo istemu tipu (Roman), primeri ostalih tipov so recimo Italic, Calligraph, German, Script ipd. Vsak znak spada bodisi na tekstovno področje ali na matematično področje.

Zbiranje podatkov

Dokumenti, ki jih vsebuje baza, so v angleškem jeziku in vsebujejo 467 strani, 688.570 znakov, 108.914 besed in 21.056 matematičnih izrazov. Iz baze pa so izključene matrike, tabele in slike. Vse strani so skenirane v ločljivosti 600 dpi in avtomatično spravljene v binarno obliko. Kvaliteta slike strani se spreminja sorazmerno s kvaliteto originalne strani. Zato mnogo strani vsebuje razne šume, kot so na primer načeti ali zlomljeni znaki.

Okarakterizacija podatkov

Vsak znak je opisan(angl. ground truth) z naslednjimi atributi:

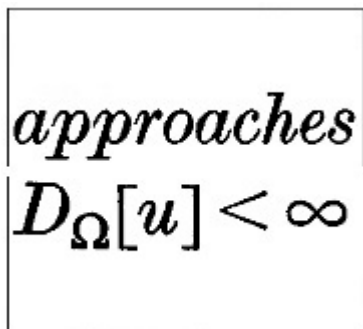
- tip, kategorija in pisava
- tekstovno ali matematično področje

- normalni ali redki znak
- velikost (višina in širina)
- povezava (predstavlja pozicijsko relacijo s predhodnim oz. starševskim znakom)
- lokacija v sliki besede ali formule
- pot do slikovnega dokumenta (ime direktorija + ime datoteke).

Segmentacija besed je načeloma dosežena s presledki. Po drugi strani pa je zaporedje matematičnih izrazov v vrsti združeno v eno formulo ne glede na presledke.

Struktura baze

Baza je sestavljena iz dveh delov: (i) tekstovni podatki in (ii) slikovni podatki, ki sta med seboj povezana. Tekstovni podatki so v Microsoft Access ali CSV formatu, medtem ko so slikovni podatki sistematično imenovane PNG datoteke. Pri tekstovnih podatkih je za vsak znak pripetih 25 atributov. Primer slikovne datoteke je viden na sliki 2.7.



Slika 2.7: Primer slikovne datoteke baze InftyCDB-1.

Pri slikovnih podatkih pa so, da bi zmanjšali število slikovnih datotek, slike iste besede v istem članku shranjene v eno slikovno datoteko. Besede tipa *Italic* in pokončnega tipa so še vedno shranjene v različne slikovne datoteke, prav tako se besede z velikimi začetnicami razlikujejo od tistih z malimi tako, da sta na primer besedi „Test“ in „test“ shranjeni v različnih datotekah. Slike formul v članku so grupirane na enak način, če so izrazi isti.

2.1.4 Primerjava baz

V vseh treh zgoraj opisanih bazah je bila poiskana njihova sestava, lastnosti in kategorije ter izolirani atributi, s katerimi je baza kategorizirana. S tem, ko je vse to naštetu že zgoraj, je s tabelo 2.1 predstavljena pregledna primerjava atributov med bazami. In sicer na naslednji način: Stolpci predstavljajo posamezne baze, vrstice pa lastnosti po katerih je možno iskati v bazi. Če so lastnosti baze zavedene in so hkrati spremenljive je okvirček označen z besedo "Da". V primeru, da atribut ne obstaja oziroma obstaja, vendar ostaja nespremenljiv skozi celotno bazo, je označen z besedo "Ne". V primeru, da lastnosti v samem opisu baze ni podane pa je oznaka "/". V primeru nejasnih situacij pa je poleg besede v oklepaju napisana kratka razlaga. Enako velja za vse nadaljne tabele.

| | NIST | ARABASE | InftyCDB-1 |
|--|------|---------|------------|
| format slike | / | / | Ne |
| barva slike | Ne | Ne | Ne |
| velikost slike | Da | Da | Da |
| ločljivost slike | Ne | Da | / |
| datum nastanka | Da | Da | / |
| gostota znakov | Da | / | / |
| globina znakov | Da | / | / |
| smer skeniranja | Da | / | Ne |
| direktno/indirektno/strojno besede/podbesede | Ne | Da | Ne |
| znak | Da | Da | Da |
| število | Da | Da | Da |
| podpis | Da | Da | / |

Tabela 2.1: Prikaz atributov v bazah NIST, ARABASE ter InftyCDB-1. Znak "/" označuje, da lastnost v bazi ni podana.

Za bolj podrobno razlago si izberimo lastnost ločljivost, za katero je pri bazi NIST oznaka „Ne“ kar pomeni, da je ločljivost slike sicer podana, vendar je enaka za vse slike, torej bi bila njena vključitev pod argumete brezpredmetna. Pri bazi ARABASE pa se nahaja oznaka „Da“ kar pomeni, da je ločljivost podana in hkrati tudi spremenljiva. Pri bazi InftyCDB-1 pa je oznaka „/“ kar pomeni, da ločljivosti v opisu baze sploh ni podane. Če pa pogledamo na primer lastnost datum nastanka pa vidimo, da je oznaka pri bazah NIST in InftyCDB-1 „Ne“ kar pomeni, da datuma poleg slike ni podanega, medtem ko

je pri bazi ARABASE „Da“ kar pomeni, da je datum zajema slike podan ter prav tako spremenljiv.

2.2 Podatkovne baze slik obrazov

V zadnjih 10 letih je prepoznavanje obraza postalo izredno popularno področje in je konstantno v porastu. Zaradi njegove narave se za njega ne zanima samo računalniška znanost, temveč tudi psihologija in nevrološke znanstvene veje. Obstaja splošno mnenje, da bodo napredki v računalniškem vidu prinesli mnoge uporabne podatke in vpoglede ravno v ta področja in pripomogli k vedno boljšemu razumevanju delovanja človeških možgan in obratno.

Nekaj smeri raziskav (sodeč po FRVT 2002 - Face Recognition Vendor Test[5]):

- prepoznavanje zunaj zajetih slik obrazov
- prepoznavanje nefrontalno zajetih slik obrazov
- prepoznavanje pri nizkih stopnjah napačne odobritve
- razumevanje, zakaj so moški obrazi bolj prepoznavni kot ženski
- večje razumevanje efekta demografskega faktorja na rezultate
- razvoj boljših statističnih metod za razumevanje rezultatov
- razvoj boljših modelov za napovedovanje obnašanja identifikacije pri velikih galerijah

Podatkovne zbirke slik obrazov poleg standardnih slik razdelimo še v dve veji in sicer termalne in pa 3D slike obrazov.

2.2.1 Standardne slike obrazov

Standardne slike obrazov so zajete z različnimi digitalnimi kamerami in v večini primerov shranjene kot 8-bitne črno bele slike. Kvaliteta slike je odvisna od svetlobnih pogojev, premikanja glave, kamere in ostalih dejavnikov. Ponavadi se slika obraz istega subjekta pod različnimi koti oz. v različnih pozicijah (frontalno in profil), prav tako pa se zajamejo tudi slike različnih izrazov na

obrazu, kot je razvidno iz slike 2.8. Bazi, ki sem ju primerjal in vsebujeta standardno zajete slike sta opisani v tabeli 2.2.



Slika 2.8: Vzorčna zbirka slik obrazov podatkovne baze YALE.

2.2.2 Termične slike obrazov

Takšne slike so zajete s posebnimi infra-rdečimi (IR) kamerami, ki zaznavajo toplotna oddajanja obraza ter jih pretvorijo v sliko. Tako se znebimo mnogih nezaželenih zunanjih dejavnikov, ki bi nas pri standardnem zajemanju slike lahko motili, kot so na primer razne spremembe svetlobe ipd. Potencial zajemanja termičnih IR slik pa je v literaturi prejel bolj malo pozornosti. Območje spreminjanja temperature človeškega telesa je na žalost dokaj malo. Najbolje je slike zajemati pri sobni temperaturi, saj takrat biometrična kalibracija kar najbolje opravi začetno segmentacijo točk kože in ponudi bolj čisto sliko. Najbolj obširna baza termalnih slik obrazov je baza Equinox [6], dva primerka slik iz te baze lahko vidimo na sliki 2.9.

2.2.3 3D slike obrazov

V preteklih nekaj desetletjih se je večina dela fokusirala na izvor 2D intenzitete barvnih slik. Ker pa je natančnost 2D prepoznavne obrazov pogojena s spreminjanjem poz, izrazov in osvetlitve, je še vedno težko razviti robustni

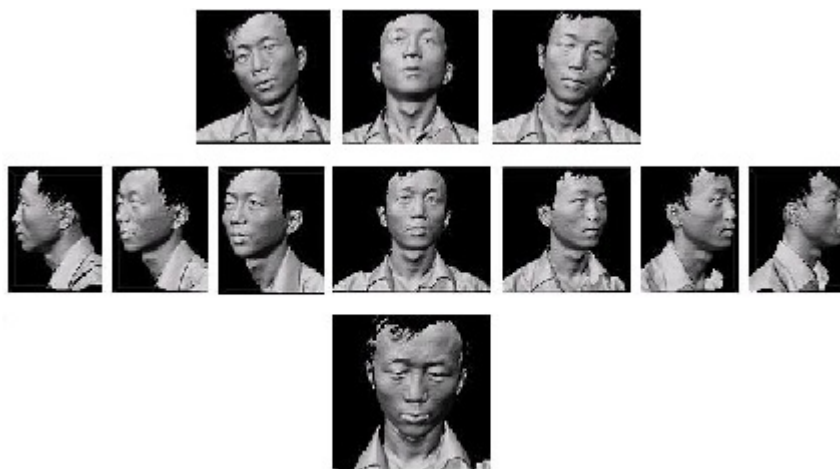
avtomatični 2D sistem prepoznavanja. 3D podatki obraza obljublajo svetlo pot k razumevanju karakteristik človeškega obraza v 3D območju, in predstavljajo potencial, da rezultate takega sistema tudi izboljšajo. Vse več raziskav se zadnja leta posveča prav 3D sistemu prepoznavanja obrazov. Dejstvo pa je da 3D kamere (slika 2.10) niso tako pogoste kot 2D kamere in da je 3D bazo slik obrazov kar drago zgraditi, kar pripelje do problema pri potrjevanju predlaganih metod v univerzalni platformi. Stvari pa se iz leta v letu izboljšujejo. CASIA je na primer izdala zelo obsežno javno bazo 3D slik obrazov z imenom CASIA 3D Face Database [7], primerki 3D slik naključnega subjekta so vidni na sliki 2.11.



Slika 2.9: Standardno zajeta slike (levo) in Termično zajeta slika (desno).



Slika 2.10: Potek zajemanja 3D slike baze CASIA 3D Face Database.



Slika 2.11: Primeri 3D slik iste osebe iz baze CASIA 3D Face Database.

2.2.4 Primerjava baz

Za primerjavo sem si, zaradi lažje predstave, vzel tri baze s standardnim zajemom slik. In sicer bazo FERET [8], The Yale Face Database [9] in pa CVL Face Database [10]. Njihova kategorizacija, razlike in podobnosti so razvidne iz tabele 2.2.

| | FERET | YALE | CVL |
|------------------------------|-------|------|-----|
| format slike | Ne | Ne | Ne |
| barva slike | Ne | Ne | Ne |
| velikost slike | Da | Ne | Ne |
| ločljivost slike | Ne | Ne | Ne |
| različne kamere | Da | / | / |
| perioda zajemanja podatkov | Da | / | / |
| različna svetloba | Da | Da | Ne |
| pozicije glave | Da | Da | Da |
| obrazni izrazi | Da | Da | Da |
| smer osvetitve | / | Da | Ne |
| vklučevanje ozadja | Ne | Da | Ne |
| upoštevanje brade, očal ipd. | Da | Ne | / |
| upoštevanje dolžine las | Ne | / | / |
| retuširanje oblačil | Da | Ne | Ne |

Tabela 2.2: Primerjava baz FERET, YALE in CVL.

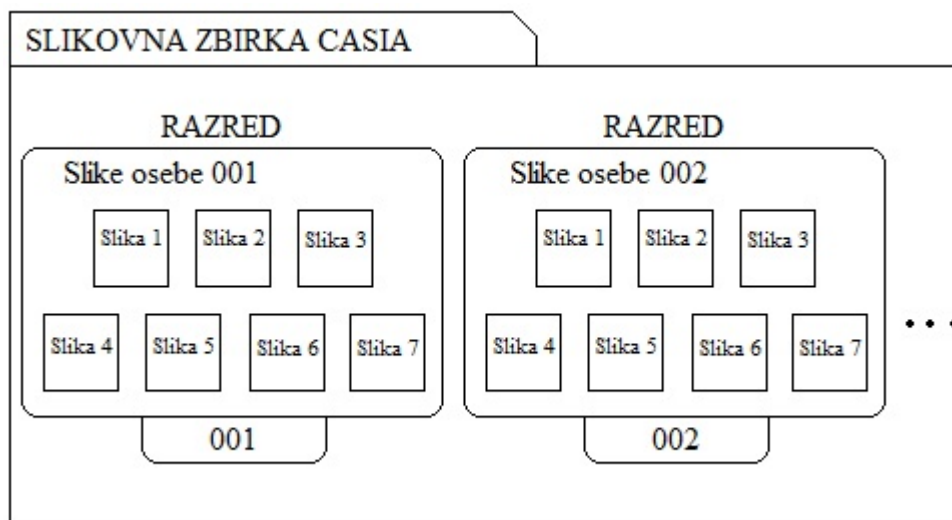
2.3 Podatkovne baze slik šarenic

Detekcija na podlagi šarenice s hitrim razvojem pridobivanja slik postaja ena izmed pomembnih komponent moderne družbe, s širokim spektrom uporabe, na primer na nacionalnih identifikacijskih karticah, potnih listih, v bančništvu, e-poslovanju, forenziki ipd. Od leta 1990 so raziskave na tem področju dosegle ogromen razvoj. Moramo pa priznati, da je delovanje sistemov detekcije šarenice v nekontroliranih okoljih še vedno nezanesljivo. (Na sliki 2.12 lahko vidimo lastnoročno slikanje šarenice v kontroliranem okolju ter naknadno lociranje šarenice). Lociranje šarenice, nelinearna normalizacija, odstranjevanje šumov, detekcija življenjskih znakov, široko zastavljena identifikacija in mnoga druga področja raziskovanja še potrebujejo nadaljne raziskave. Uspeh teh raziskav pa je pogosto odvisen od razpoložljivosti skrbno zasnovanih in dovolj velikih baz slik šarenic.



Slika 2.12: Slikanje očesa za bazo CASIA-IrisV3 ter lociranje šarenice.

V nadaljevanju bom predstavil bazi CASIA-IrisV3 [11] in pa UBIRIS.v2 [12]. Osnovno sestavo baze CASIA lahko vidimo na sliki 2.13. Obe sta v svetu poznani bazi, sta javni in brezplačni, tako da jih za svoje raziskave uporabljajo mnogi raziskovalci in znanstveniki. Casia sestoji iz treh podbaz CASIA-IrisV3-Interval, CASIA-IrisV3-Lamp, CASIA-IrisV3-Twins ter vsebuje več kot 22.035 slik šarenice od več kot 700 subjektov. Ubiris baza je sicer malce manjša, saj sestoji iz slik zgolj 261 subjektov, vendar se razlikuje od CASIA baze tudi v drugih stvareh, saj vključuje slikanje v premikanju, rotiranje glave, spreminjanje razdalje slikanja ipd. Zanimivo je tudi, da ima CASIA skoraj vse subjekte azijskega porekla, za razliko od Ubirisa, ki ima sestavo raznoliko. V bazi CASIE zasledimo tudi posvečanje posebne pozornosti slikam šarenic dvojčkov, ki imajo izredno podobne šarenice in jih je še toliko težje ločiti.



Slika 2.13: Tipična osnovna zgradba baze CASIA.

2.3.1 Primerjava baz

Bazi CASIA-IrisV3 ter UBIRIS.v2. sta primerjani v tabeli 2.3.

| | CASIA-IrisV3 | UBIRIS.v2. |
|----------------------------|--------------|------------|
| format slike | .Ne | Ne |
| barva slike | Ne | Ne |
| velikost slike | Da | Ne |
| ločljivost slike | Ne | Ne |
| različne kamere | Da | / |
| perioda zajemanja podatkov | Da | / |
| različna svetloba | Da | Da |
| levo ali desno oko | Da | Da |
| starost subjekta | Da | Da |
| razdalja slikanja | / | Da |
| rotiranje glave | / | Da |
| premikanje med slikanjem | / | Da |
| kvaliteta slike | Da | Da |
| znotraj lab. ali zunaj | Da | Da |
| enojajčni dvojčki | Da | Ne |

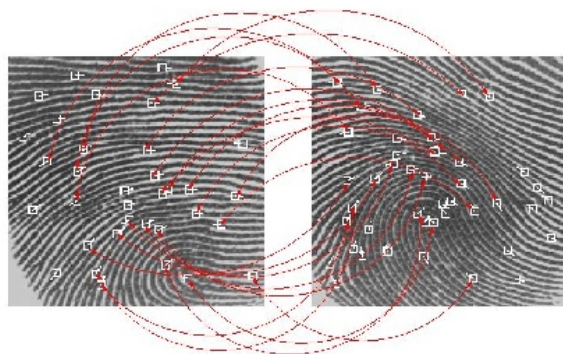
Tabela 2.3: Primerjava baz CASIA-IrisV3 in UBIRIS.v2.

2.4 Podatkovne baze slik prstnih odtisov

Med vsemi biometričnimi tehnikami je metoda identifikacije, ki bazira na prstnem odtisu ena izmed najstarejših. Vsak posameznik ima unikaten in nespremenljiv prstni odtis. Sestavlja ga vrsta robov in globeli na sami površini blazinice prsta, njegova unikatnost pa je določena z vzorcem le-teh in pa tudi z značilnimi točkami. Značilne točke so lokalne značilnosti roba, ki se pojavijo bodisi na razpotju roba ali pa na njegovem koncu (slika 2.14 in slika 2.15). Ogromna količina prstnih odtisov je vsak dan zbrana in shranjena v raznih aplikacijah, za potrebo forenzike, kontrole dostopa, registracijo prometnih dovoljenj ipd. Avtomatična razpoznavna ljudi na podlagi prstnih odtisov zahteva primerjanje vhodnega prstnega odtisa z ogromnimi bazami, ki lahko vsebujejo tudi do 70 milijonov prstnih odtisov (podatkovna baza FBI). Da bi zmanjšali čas iskanja, je zaželeno vse te slike prstnih odtisov ustrezno klasificirati tako, da se vhodni prstni odtis primerja samo z delčkom baze oz. podbazo.



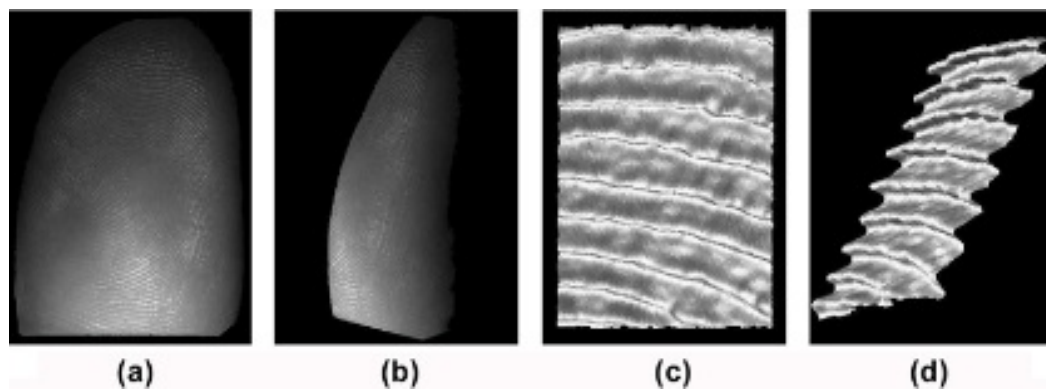
Slika 2.14: Primer prstnega odtisa z že označenimi značilnimi točkami.



Slika 2.15: Primerjanje značilnih točk med različnimi prstnimi odtisi.

2.4.1 3D slike prstnih odtisov

Tradicionalno pridobivanje 2D slik prstnih odtisov uporablja kontaktno metodo, ki predstavlja nenadzorovana popačenja pri pritiskanju in rotiranju prsta po površini. Da bi te probleme zaobšli obstajajo posebni 3D skenerji prstnih odtisov. Prikaz 3D prstnega odtisa vidimo na sliki 2.16. Sistem se zanaša na 3D pridobivanje slik s strukturiranimi svetlobnimi iluminacijami, katere pridobijo potrebne informacije s triangulacijo med projektorjem in parom kamer z uporabo zaporedja časovno kompleksnih okrnjenih vzorcev svetlobe. Prstna blazinica ne pride v stik s površino.



Slika 2.16: 3D prstni odtis iz univerzitetne baze v Kentuckyju prikazan iz različnih zornih kotov. Frontalno (a), stransko (b), frontalno z globinske perspektive (c), stransko z globinske perspektive (d).

Bazo 3D slik prstnih odtisov so naredili na univerzi v Kentucky-ju ter vsebuje 441 3D slik enajstih subjektov. Vsi prstni odtisi so sploščeni in v ločljivosti 500 dpi [13].

2.4.2 Primerjava baz

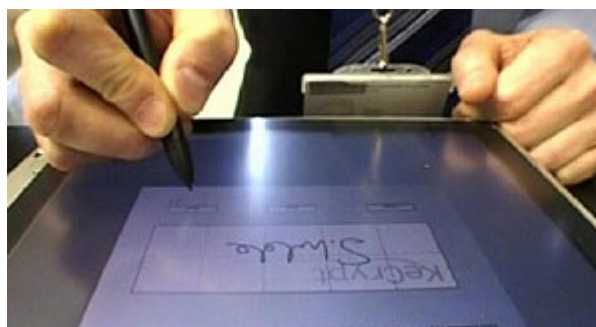
V tabeli 2.4 je razvidna primerjava baz Fingerprint Verification Competition 2004 (FVC 2004) [14] ter NIST Special Database 4 Fingerprint Database [15].

| | FVC 2004 | NIST4 |
|----------------------------|----------|-------|
| format slike | Ne | Ne |
| barva slike | Ne | Ne |
| velikost slike | Da | Ne |
| gostota točk | Ne | Da |
| ločljivost slike | Da | Ne |
| različni skenerji prstov | Da | Ne |
| časovna oznaka | Da | Da |
| moč pritiska prsta | Da | Ne |
| vlažnost prstov | Da | Ne |
| perioda zajemanja podatkov | Da | Ne |
| velikost slike v memoriji | Ne | Da |
| spolu subjekta | Ne | Da |

Tabela 2.4: Primerjava med bazama FVC 2004 in NIST4.

2.5 Podatkovne baze slik podpisov

Računalniška verifikacija ročnega podpisa je proces potrjevanja identitete uporabnika. Primarna prednost, ki jo ima takšna verifikacija pred ostalimi biometričnimi tehnologijami je ta, da je podpis že sam po sebi najbolj razširjen, pogosto uporabljen in sprejet način za verifikacijo oseb. Kar pomeni, da nam že sama zgodovina govori, da je podpis zaupanja vreden. Primer podpisa na elektronsko tablico prikazuje slika 2.17.



Slika 2.17: Podpis na e-tablico.

2.5.1 Primerjava baz

Izbral sem si bazi slik podpisov SVC2004 [16], in pa GPDS300signature [17], saj sta izmed vseh ponujali največ javno dostopnih podatkov, ju primerjal in rezultate predstavil v tabeli 2.5.

| | SVC 2004 | GPDS300 |
|----------------------------|----------|---------|
| uporaba elektronske table | Da | Da |
| orientacija pisala | Da | Da |
| hitrost pisala | Da | Da |
| moč pritiska pisala | Da | Da |
| vrsta podpisa | Da | Da |
| časovna oznaka | Da | / |
| barva slike | / | Ne |
| ločljivost slike | / | Ne |
| perioda zajemanja podatkov | / | Ne |
| rodna država subjekta | Da | / |

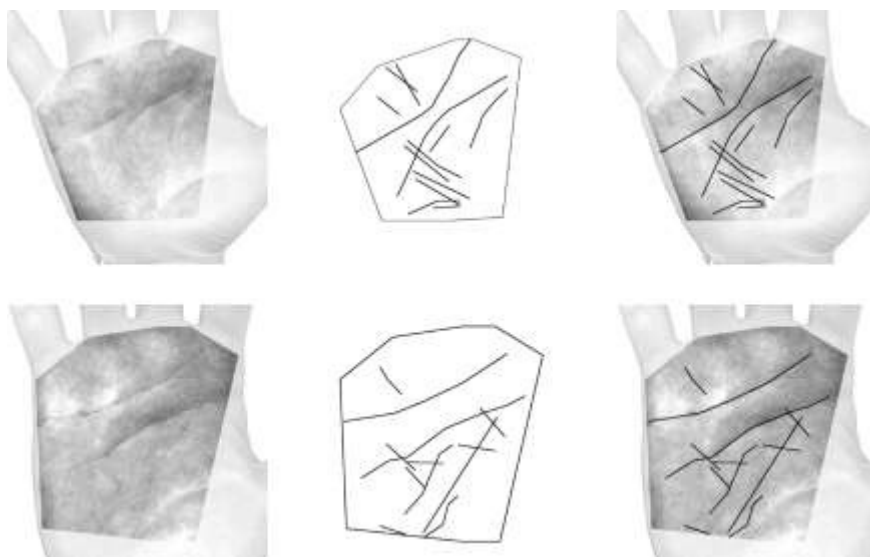
Tabela 2.5: Primerjava baz SVC 2004 in GPDS300.

2.6 Podatkovne baze slik odtisa dlani

Na področju biometrije je odtis dlani še novinec, vendar obljublja veliko. Do sedaj še ni bilo na voljo standardne kamere za slikanje odtisov dlani, mnoge raziskovalne skupine so poskušale pridobiti slike z uporabo optičnih čitalcev, digitalnih kamer ipd. Biometrični raziskovalni inštitut (UGC/CRC) na politehnični univerziteti v Hong Kongu [18] pa je razvil eno izmed prvih realno časovnih naprav za zajemanje slik odtisov dlani in ustvaril tudi bazo [19]. Primer slike dlani zajete brez dotikanja samega skenerja vidimo na sliki 2.18. Takšne slike brez dotika dlani (angl. Touchless Hand Print), so lahko uporabljene tudi za razvoj 3D slik dlani. Biometrični sistem na podlagi slik dlani podeduje mnogo karakteristik uveljavljenega prstnega odtisa, saj oba črpata informacije iz strukture robov prisotnih na sami dlani oz. blazinicah prstov, ki so unikatni za vsakega posameznika, saj robovi potekajo pri vsakem človeku drugače, kar je lepo razvidno iz slike 2.19.



Slika 2.18: Primer slike dlani iz baze IIT Delhi Touchless Palmprint.



Slika 2.19: Potek značilnih robov na dlani.

2.6.1 Primerjava baz

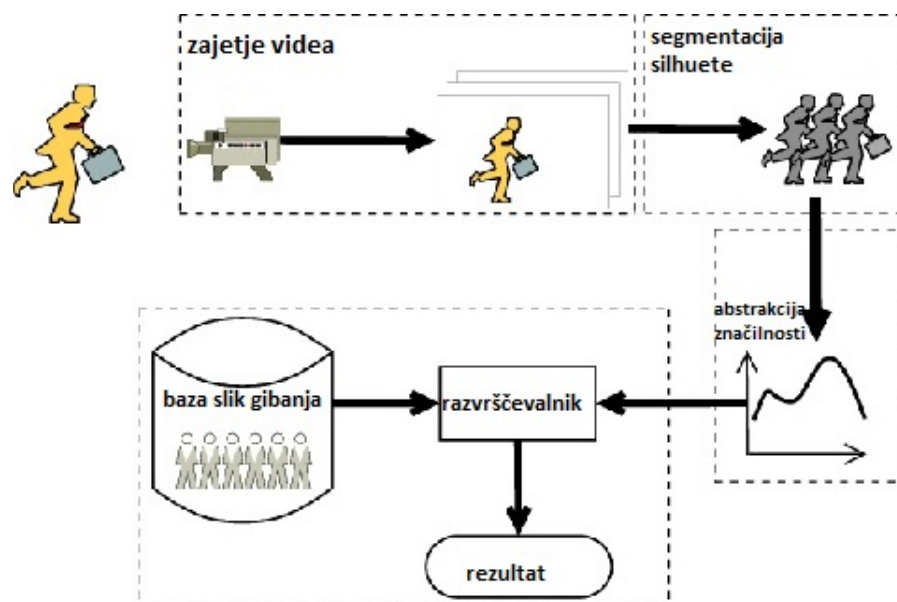
Izbrani bazi slik dlani sta CASIA Palmprint [20] in pa IIT Delhi Touchless Palmprint [21], ki ima slike dlani zajete brez skenerja. Značilnosti obeh so predstavljene v tabeli 2.6.

| | CASIA | Delhi |
|-------------------------------|-------|-------|
| različne kamere | Da | Ne |
| iluminacija okoli leče kamere | Ne | Da |
| leva ali desna dlan | Da | Da |
| spol subjekta | Da | Ne |
| restrikcija dlani | Ne | Ne |
| normalizacija slike | / | Da |
| ločljivost slike | / | Ne |

Tabela 2.6: Primerjava baz CASIA in Delhi.

2.7 Podatkovne baze slik gibanja

Z vedno večjo potrebo po video nadzornih sistemih postaja identifikacija na daljavo vedno pomembnejša. Takšno identifikacijo lahko priskrbi biometrični sistem na podlagi gibanja (slika 2.20) (angl. gait recognition), ki zadnje čase dobiva vse več pozornosti. Gibanje pomeni v grobem način hoje oz. obnašanje pri hoji. V primerjavi z drugimi biometričnimi sistemi ima veliko prednosti, na primer identifikacija na distanci, neinvazivnost, izločenje potrebe po kontaktu ipd. Sicer za to področje obstajajo tudi mnoge baze video posnetkov, vendar si bomo za naše potrebe izbrali zgolj baze slik gibanja.



Slika 2.20: Tipičen biometričen sistem na osnovi gibanja.

2.7.1 Primerjava baz

V tabeli 2.7 sta primerjani bazi The Southampton Human ID at a Distance Gait Database [22], ter CASIA Gait [23] (katero sestavljajo 3 podbaze A, B in C).

| | Southampton | CASIA |
|-------------------------------------|-------------|-------|
| različne kamere | Da | Da |
| koti snemanja | Da | Da |
| hitrost gibanja | Da | Da |
| perioda snemanja | Ne | Ne |
| dolžina hoje | Ne | Da |
| pripomočki in obleke (torbice ipd.) | Da | Da |
| lokacija snemanja | Da | Da |
| ozadje posnetka | Da | Da |
| tekoči trak za hojo | Da | Ne |

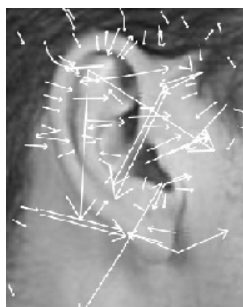
Tabela 2.7: Primerjava baz Southampton in CASIA.

2.8 Podatkovne baze slik ušes

V biometričnih sistemih zaradi neinvazivnosti pogosto srečamo obraz in ušesa (slika 2.21). Ušesa so lahko zelo uporabna, saj ohranjajo identično strukturo oz. pozicijo značilnih točk (slika 2.22) skozi vsa leta. Lahko so uporabljena samostojno ali v kombinaciji z obrazom, saj so večkrat zelo majhna ali pa prekrita z lasmi, uhani ipd.



Slika 2.21: Sistem prepoznave obraza v kombinaciji z ušesi.



Slika 2.22: Točke interesa (oz. značilne točke) prikazane nad sliko ušesa v bazi XM2VTS.

2.8.1 Primerjava baz

Ravno zaradi razloga navezave obraz-uho obstaja bolj malo samostojnih baz slik ušes. Tabela 2.8 primerja bazo slik ušes XM2VTS [24] in pa bazo slik ušes in obrazov MIT [25].

| | MIT | XM2VTS |
|---------------------|-----|--------|
| uho in obraz | Da | Ne |
| slik na subjekt | Da | Da |
| perioda slikanja | Ne | Da |
| namen slik | Da | Da |
| velikost slike | Ne | Da |
| različna svetloba | Da | Ne |
| normalizacija slike | Da | Ne |

Tabela 2.8: Primerjava med bazama MIT in XM2VTS.

2.9 Ostale podatkovne baze

Poleg navedenih slikovnih baz obstaja še mnogo drugih slikovnih baz področij računalniškega vida, katere je potrebno omeniti. Nekaj izbranih področij je navedenih spodaj:

- prepoznavanje objektov [26]
- lokalizacija robotov [27]
- prepoznavanje urbanega okolja [28]

Poglavje 3

Metodologija zajema

Čeprav sem že v prejšnjem poglavju odkril nekaj lastnosti metodologij zajema, gradnje baz, se bom v tem poglavju v metodologijo poglobil, predvsem z namenom, da razčlenim metodologijo zajema baze CVL OCR DB, ki bo vsebovala slike teksta v naravnih scenah.

3.1 Preučevanje metodologije drugih baz

3.1.1 CASIA-IrisV3

Kot vse vodilne slikovne baze na področju biometrije ima tudi CASIA-IrisV3 [11] zelo dobro premišljeno metodologijo zajema. Vsebuje 22.035 slik od več kot 700 oseb. Vse slike so sivinske s končnico JPEG. Zajete so bile pod skoraj infrardečo osvetlitvijo, vsebuje pa večinoma osebe kitajskega porekla (razen nekaj izjem v bazi CASIA-IrisV3-Interval), več statistik in značilnosti je zbranih v tabeli 3.1. CASIA-IrisV3 se razdeli na tri podbaze in sicer CASIA-IrisV3-Interval, CASIA-IrisV3-Lamp in pa CASIA-IrisV3-Twins, vse tri so bile zbrane v različnih časovnih obdobjih. Vsaka izmed njih pa je drugače zgrajena oz. vsebuje drugačno kategorizacijo slik.

CASIA-IrisV3-Interval

Slike za to podbazo so bile zajete s (za to posebej razvito) kamero. So zelo čiste in natančne. Vsebuje podbazo imenovano CASIA V1.0, ki je bila narejena zaradi povpraševanja mnogih raziskovalcev oziroma raziskovalnih ekip. V tej bazi imajo zaradi patentne prijave vse slike že avtomatično zaznane in nadomeščene območje zenice s krogom konstantne intenzitete. Tako na tem

področju tudi izničijo svetlobne odboje ter omogočijo lažjo detekcijo šarenice. Lastnosti te baze lahko razberemo iz tabele 3.1.

CASIA-IrisV3-Lamp in CASIA-IrisV3-Twins

Slike obeh baz so bile pridobljene z ročnim OKI senzorjem šarenice. Senzor ima tudi svetilko, ki so jo z namenom, da bi pridobili čim več slik pod različnimi pogoji, tudi ugašali in prižigali. Lastnosti obeh so predstavljene v tabeli 3.1. Takšne slike spadajo v bazo CASIA-IrisV3-Lamp. Medtem ko baza CASIA-IrisV3-Twins kot prva javno ponuja slike šarenic stotih parov dvojčkov, njihove šarenice so si med sabo lahko zelo podobne.

Lastnosti so predstavljene v tabeli 3.1, ki povzema značilnosti vseh treh podbaz. Zaradi podobnosti teh treh baz sem poleg obstoja atributa dodal še dodatne informacije.

| | IrisV3-Interval | IrisV3-Lamp | IrisV3-Twins |
|---------------------|-----------------|--------------|--------------|
| različne kamere | Ne (lastna) | Ne(OKI) | Ne (OKI) |
| okolje | Ne (znotraj) | Ne (znotraj) | Ne (zunaj) |
| različna svetloba | Ne | Da | Ne |
| perioda slikanja | Da (dve) | Ne (ena) | Ne (ena) |
| datum nastanka | Da | Da | / |
| število subjektov | Ne (249) | Ne (411) | Ne (200) |
| število podrazredov | Ne (395) | Ne (819) | Ne (400) |
| število slik | Ne (2655) | Ne (16213) | Ne (3183) |
| velikost slike | Ne (320*280) | Ne (640*480) | Ne (640*480) |

Tabela 3.1: Statistika in značilnosti slikovnih podbaz CASIA-IrisV3-Interval, CASIA-IrisV3-Lamp, CASIA-IrisV3-Twins.

3.1.2 FERET

Slikovna baza FERET [8] je bila ustvarjena kot podpora vladnem testiranju in ocenitvi programov za prepoznavo obrazov. Zadnja verzija vsebuje kar 14051 sivinskih slik človeških obrazov. Slike so zajete iz različnih vidikov in sicer frontalno ter s profila, pod različnimi koti, narejene pa so tudi naključne slike. Prav tako se spreminja izraz na obrazih subjektov (osnovni in alternativni). Slikanje istih subjektov poteka v različnih periodah, od enega dneva do treh let. Spremenljiva je tudi osvetlitev obraza pri slikanju, uporablja pa se dve

kameri. Loči se tudi slike subjektov, ki nosijo očala, ali imajo različno dolge lase čez neko določeno časovno obdobje ter kombinacije obojega. Obleke so avtomatično retuširane. Spreminja pa se tudi svetlost same slike ter njena velikost. Končnica datoteke slik je tif.

Feretovo poimenovanje (t.i. numenklatura) posamezne slike izgleda takole: nnnnxxfffq-yymmdd.ext, kjer določeni deli imena pomenijo naslednje:

- **nnnn**: Je petmestna številka, ki unikatno predstavlja subjekt.
- **xx**: Predstavljata dve črki, ki označujeta vrsto slike. Kombinacije teh dveh črk so predstavljene v tabeli 3.2.

| Dvočrkovna koda | Kot poziranja (stopinje) | Opis | Številka v bazi | Število subjektov |
|-----------------|--------------------------|--|-----------------|-------------------|
| Fa | 0 | osnovni obrazni izraz | 1762 | 1010 |
| Fb | 0 | alternativni obrazni izraz | 1518 | 1009 |
| ba | 0 | frontalno serije "b" | 200 | 200 |
| bj | 0 | alternativni obrazni izraz za ba | 200 | 200 |
| bk | 0 | različna osvetlitev za ba | 200 | 200 |
| bb | +60 | Subjekt je obrnjen na svojo levo (fotografova desna stran) | 200 | 200 |
| bc | +40 | | 200 | 200 |
| bd | +25 | | 200 | 200 |
| be | +15 | | 200 | 200 |
| bf | -15 | Subjekt je obrnjen na svojo desno (fotografova leva stran) | 200 | 200 |
| bg | -25 | | 200 | 200 |
| Bh | -40 | | 200 | 200 |
| bi | -60 | | 200 | 200 |
| ql | -22.5 | Četrtno levo in desno | 763 | 508 |
| qr | +22.5 | | 763 | 508 |
| hl | -67.5 | Polovico levo in desno | 1246 | 904 |
| hr | +67.5 | | 1298 | 939 |
| pl | -90 | Profil levo in desno | 1318 | 974 |
| pr | +90 | | 1342 | 980 |
| Ra | +45 | Naključne slike (Pozitivni koti predstavljajo obraze subjektov obrnjene na fotografovo levo stran) | 322 | 264 |
| Rb | +10 | | 322 | 264 |
| Rc | -10 | | 613 | 429 |
| Rd | -45 | | 292 | 238 |
| Re | -80 | | 292 | 238 |

Tabela 3.2: Dvočrkovne kode, ki označujejo vrsto slike v bazi FERET.

1. **fa** - standardna frontalna slika
 2. **fb** - alternativna frontalna slika, zajeta tik za **fa** sliko
 3. **ba** - frontalne slike, ki so popolnoma proporcionalne slikam **fa** serije
 4. **bj** - alternativna frontalna slika sliki **ba** in proporcionalna **fb** sliki
 5. **bk** - prav tako frontalna slika, vendar zajeta pod drugačnimi svetlobnimi pogoji
 6. **bb...bi** - serija slik zajetih z namenom raziskovanja vpliva kotov poze subjekta. Slike **bf...bi** so simetrično proporcionalne slikam **bb...be**
 7. **ra...re** - naključne orientacije. Točna vrednost njihovega kota je neznana. Pokaže se, da so koti poz naključni, vendar konsistentni. Koti poz v tabeli so bili izračunani z ročnim merjenjem medočesnih razdalj v sami sliki in njeni frontalni različici
- **fff**: Je zaporedje treh binarnih zastavic, ki po vrstnem redu pomenijo naslednje:
 1. Označuje, če je sliki dovoljena uporaba v publikaciji. Ta zastavica se pri dotični bazi sicer ne uporablja, saj so v njihovi distribuciji na voljo vse slike
 2. Če je ta zastavica nastavljena na 1 to indicira, da je slika grafično prirejena.
 3. Označuje, če je bila slika zajeta z ASA 200 ali ASA 400 filmom.
 - **q**: Je modifikator, ki ni vedno prisoten. Ko pa je, lahko pomeni naslednje:

1. **a** - prisotnost očal
2. **b** - isti subjekt z različno dolžino las
3. **c** - prisotnost očal in različna dolžina las
4. **d** - elektronsko proporcionalna ter dodan graf
5. **e** - digitalno spremenjena oblačila
6. **f** - svetlost slike zmanjšana za 40 odstotkov
7. **g** - svetlost slike zmanjšana za 80 odstotkov
8. **h** - velikost slike zmanjšana za 10 odstotkov
9. **i** - velikost slike zmanjšana za 20 odstotkov
10. **j** - velikost slike zmanjšana za 30 odstotkov

Opaziti je, da so spremembe **d...j** rezultat uporabe mnogih indirektnih (angl. offline) operacij nad dejanskimi slikami v bazi. Starševska slika je tista, ki ne vsebuje modifikatorja **q**.

- **yymmdd**: Ta tri polja predstavljajo datum zajema slike (leto, mesec, dan).
- **.ext**: Končnica datoteke je tif. Slike v bazi prav tako vsebujejo končnico bz2, saj so bile skrčene z bzip2 skrčevalcem, ki je prav tako vključen v sami distribuciji baze.

3.1.3 NIST Special Database 6

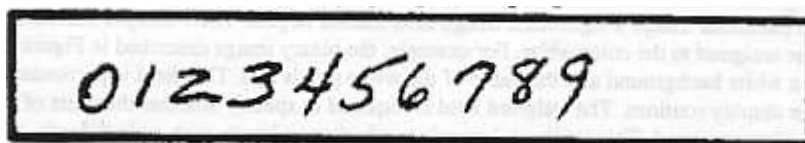
Osnovni podatki

Izbrana NIST Special Database 6 [29] vsebuje slike sestavljenih dokumentov oz. strukturiranih davčnih obrazcev. Zajema 5.595 celostranskih slik, že izpolnjenih davčnih obrazcev. Vsaka slika je shranjena v črno belem formatu ter

mora po digitalizaciji vsebovati določene osnovne podatke oz. lastosti, ki so shranjeni v strojno berljivi glavi slikovne datoteke. Program, namenjen manipuliranju s podatki dotične slike, namreč najprej prebere glavo ter nato določi primerno interpretacijo kasnejših podatkov. Vrednosti glave izbrane izolirane podslike prikazane na sliki 3.1 so prikazane v tabeli 3.3.

| | |
|---------------------------|------------------------------|
| Identiteta | TestneSlika.pct "0123456789" |
| Velikost glave | 288 (bitov) |
| Ustvarjena | čet Jan 4 17:34:21 1990 |
| Širina | 656 (točk) |
| Višina | 135 (točk) |
| Ločljivost | 300 (točk na enoto) |
| Kompresija | 2 (koda) |
| Dolžina kompresije | 874 (bitov) |
| Skeniranje | 16 (bitno) |
| Slikovna podatkovna enota | 16 (bitov) |
| Vrstni red bitov | visoko-nizko |
| Nastavitev stolpcev | 0 (slikovnih elementov) |
| Podatkovne enote | nedoločena |
| Vrstni red skeniranja | vrstično generalno |
| Podatkovne enote | nedoločena |
| X os | 192 (slikovnih elementov) |
| Y os | 732 (slikovnih elementov) |

Tabela 3.3: Glava slikovne datoteke, ki prikazuje vrednosti posameznih atributov v NIST Special Database 6.



Slika 3.1: Naključna slika izoliranega okvirja baze NIST Special Database 6.

Vsebina in organizacija baze

Da bi zagotovili realne vzorce ročne pisave na strukturiranih obrazcih, jih je priporočljivo vzeti od čim več ljudi. Ta baza vsebuje segmentirane vzorce

od 2.100 subjektov. Cilj je bil prikazati unikatna avtorstva za vsak obrazec. En pristop k temu bi bila uporaba znakov izključno enega avtorja pri sestavljanju posameznega obrazca in sicer tako, da vsak napiše 13 unikatnih pojavitev številke "0" ter eno unikatno ponovitev vsake velike in male črke od „A“ do „Z“. S tem omogočimo izraz večine naravnih variacij, ki se pojavljajo v digitalizirani ročni pisavi. Drugi pristop pa analizira in loči subjekte glede na podobnost njihove ročne pisave. Za ugotavljanje podobnosti so bile nad slikami uporabljene štiri različne meritve. To so bile višina slikovnih točk, širina slikovnih točk, gostota slikovnih točk ter nagib. Višina slikovnih točk je število točkovnih vrstic med najvišjo in najnižjo črno točko znaka v sliki. Širina slikovnih točk pa je število točkovnih stolpcev med najbolj levo in najbolj desno črno točko znaka v sliki. Medtem ko je gostota slikovnih točk izračunana z deljenjem števila slikovnih točk v sliki znaka z rezultatom množenja višine in širine. Nagib pa je izračunan kvantifikator, ki predstavlja delež horizontalnega nagiba, zahtevanega za transformacijo znaka v njegovo navpično verzijo. Te meritve so bile uporabljene za razdelitev slik znakov v 16 neodvisnih skupin. Štiri podobnostne vrednosti so bile izračunane za slike vseh znakov različnih razredov, na primer vse „0 „ ustvarijo 4 distribucije za vsak razred. Srednje vrednosti so izračunane za vsako od teh distribucij. Zatem so 4 podobnostne vrednosti za vsako sliko znaka primerjane s srednjimi vrednostmi teh štirih distribucij meritev, navezujoč se na razred znaka. Če je vrednost podobnosti večja kot srednja vrednost, potem sliki znaka pripada binarno število 1. Če pa je vrednost podobnosti manjša kot srednja vrednost, potem sliki znaka pripada binarno število 0. Na koncu dobimo štiri binarne identifikatorje za vsako sliko znaka, en identifikator za eno meritev podobnosti. Če združimo te štiri identifikatorje v sekvenčnem vrstnem redu, ustvarimo 16 unikatnih binarnih sekvenc. Za dotično podatkovno bazo so vse slike znakov, ki so imele isto unikatno binarno sekvenco, postale ena izmed šestnajstih možnih ročnih pisav. Te ročne pisave so bile uporabljene za sintetiziranje obrazcev v sami bazi. Na sliki 3.2 lahko vidimo sintetizirani naslov z ročno pisavo, ki vsebuje znake za katere so bile vse štiri podobnostne vrednosti manjše od srednje vrednosti njihovih meritvenih distribucij.

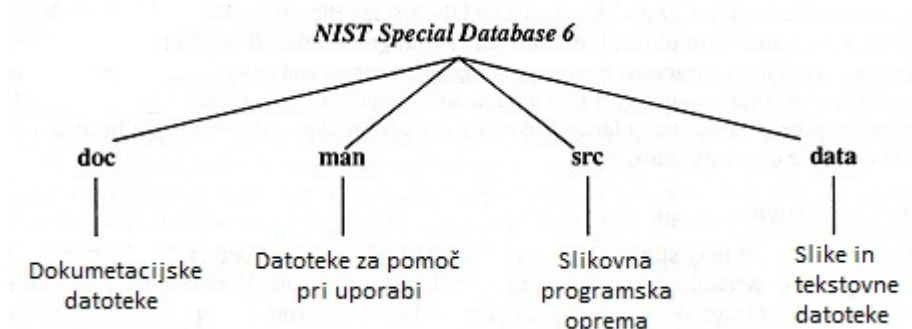
| | | |
|-----------------------|--|-----------|
| L A B E L | Your first name and initial (if joint return, also give spouse's name and initial) | Last name |
| | Suffolk u. & Taylor M. Ramsey | |
| | Present home address (number, street, and apt. no. or rural route). (If a P.O. Box, see page 6 of instructions.) | |
| H O M E | 25300 Early Road | |
| | City, town or post office, state, and ZIP code | |
| | Rockdale, HI 31907 | |

Slika 3.2: Primer slike naslovnega okvirja v bazi NIST Special Database 6.

Teh 5.595 celostranskih slik obrazcev vključno s 5.595 tekstovnimi datotekami zasede 630 MB prostora. Binarne slike so skrčene. Neskrčena baza bi tako zasedla okoli 5.95 GB prostora.

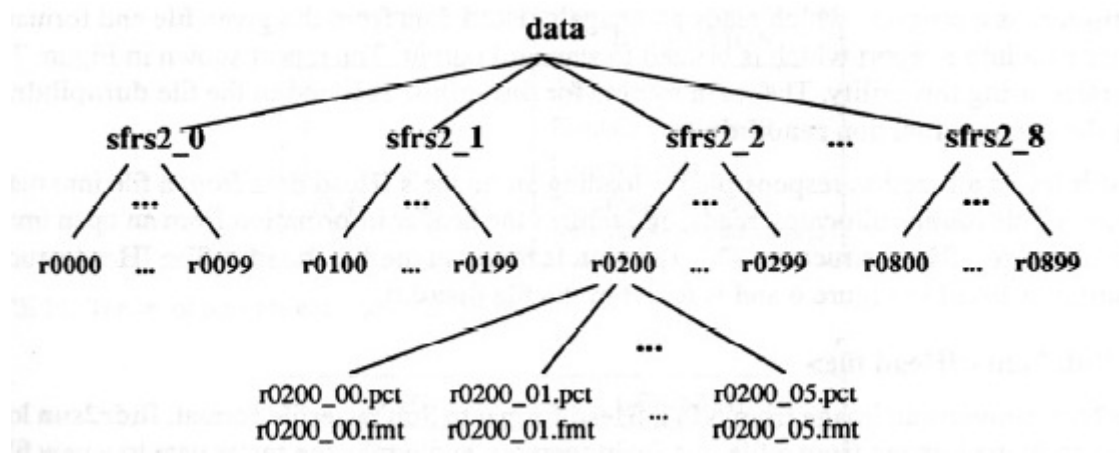
Hierarhija

Če najprej pogledamo osnovno zgradbo baze, se ta deli na mape data, doc, man in src. Doc, man in src vsebujejo dokumentacijo in pripomočke potrebne za manipulacijo vseh slikovnih podatkov. Medtem ko mapa data vsebuje datoteke slik vključno z njihovimi odgovarjajočimi tekstovnimi datotekami. Organizacija teh datotek ja predstavljena na sliki 3.3.



Slika 3.3: Osnovna drevesna zgradba baze.

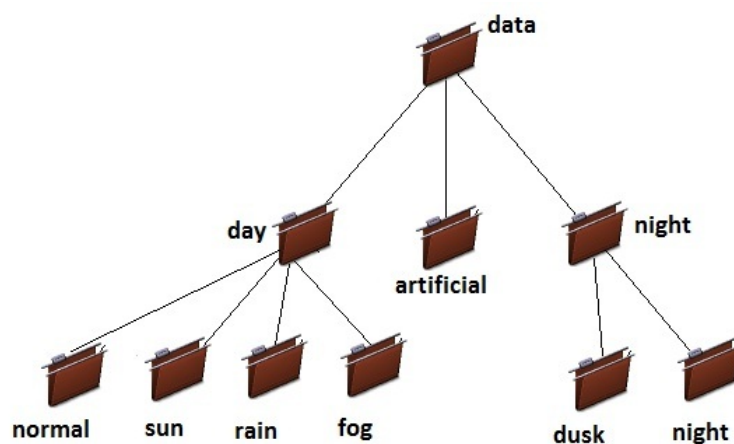
Vseh 5.595 slik je v mapi data razdeljenih še na 9 podmap (glej sliko 3.4). Podmapi sfrs2-0, sfrs2-1, do sfrs2-8 vsebujejo po 100 sintetiziranih davčnih vlog iz naključno izpolnjenih računalniških obrazcev. Torej 900 davčnih vlog vsega skupaj. Vsaka vloga je predstavljena kot mapa. Primeri map vlog segajo od r0200 do r0299. Vse slike povezane z vlogo 200 so shranjene v mapi r0200. V povprečju je v vsaki mapi shranjenih 6,22 slik, r0200 vsebuje 6 sintetiziranih obrazcev shranjenih pod datoteke r0200-00.pct, r0200-01.pct, do r0200-05.pct, kjer zadnji dve številki predstavljata indeks vsake slike. Za vsako sliko obrazca obstaja z njo povezana informacijska datoteka. Za sliko r0200-00.pct je r0200-00.fmt, za sliko r0200-01.pct je r0200-01.fmt ipd. Na tak način je vseh 5.595 slik shranjenih vključno z njihovimi formatnimi datotekami, kar posledično nanese vsega skupaj 11.190 datotek.



Slika 3.4: Organizacija slik in datotek v bazi Nist Special Database 6.

3.2 Metodologija zajema podatkovne baze CVL OCR DB

Slike v bazi so barvne in shranjene v formatu jpg. Zajete so bile z mobilnim fotoaparatom ločljivosti 5 Mp in z digitalnim fotoaparatom z ločljivostjo 7 Mp, pod pravim kotom in iz povprečne razdalje 2 metrov. Slike v bazi so predstavljene v obliki drevesa, pri katerem vozlišča predstavljajo mape, saj se lahko tako nazorno prikaže dejansko zgradbo baze po kategorijah (glej sliko 3.5).



Slika 3.5: Drevesna struktra osnovne zgradbe baze CVL OCR DB.

3.2. METODOLOGIJA ZAJEMA PODATKOVNE BAZE CVL OCR DB37

Z namenom, da bi pri slikanju dosegli čimbolj različne svetlobne pogoje, so bile slike zajete ponoči in podnevi ter ob različnih vremenskih razmerah. Nočne slike vključujejo mračne slike (slikane brez fleša) ter slike zajete s flešom v popolni temi. Dnevne fotografije pa vključujejo slike zajete pri močnem soncu (pojav bleščanja in odbleskov), v megli (slabša vidljivost), v dežju in pri normalni svetlobi. Na koncu so dodane še fotografije pri umetni svetlobi, ki so bile zajete znotraj stavb, v nakupovalnih centrih ter raznih zgradbah neodvisnih od naravne svetlobe. Na podlagi teh svetlobnih oz. vremenskih pogojev je struktura baze slik tudi predstavljena.

Ko so slike posortirane po kategorijah, zgoraj opisanih svetlobnih ter vremenskih pogojev, se naprej delijo v grobem na tri dele, in sicer:

- Celotne slike: So slike, na katerih je poleg samega tekstovnega območja vidno še celotno ozadje. Ime vsake celotne slike vsebuje oznako "full". Primer imena slike izgleda takole: d-r-00001-full.jpg (kjer d pomeni dan, r pomeni dež, 00001 oznako slike, .jpg pa predstavlja format slike). Iz teh slik se nato določi območje besedila, bodisi ročno kot smo to naredili mi, bodisi z uporabo programa za detekcijo področja besedila v sliki (glej sliko 3.6). Na koncu se to območje shrani v lastno sliko.



Slika 3.6: Označitev območja besedila za izrez v celotni sliki.

- Slike besedila: Po izločitvi ozadja in ostalih motenj iz celotne slike ostane zgolj besedilo, ki ga shranimo v sliko z oznako "text" v imenu. Primer imena izgleda takole: d-r-00001-text.jpg (kjer text torej pomeni, da gre za sliko, ki vsebuje zgolj besedilo). Iz teh slik se v nadaljevanju izolira vse znake (glej sliko 3.7), prav tako bodisi ročno kot smo naredili mi, bodisi s pomočjo programa za detekcijo znakov ter se vsakega naknadno tudi shrani v lastno sliko.



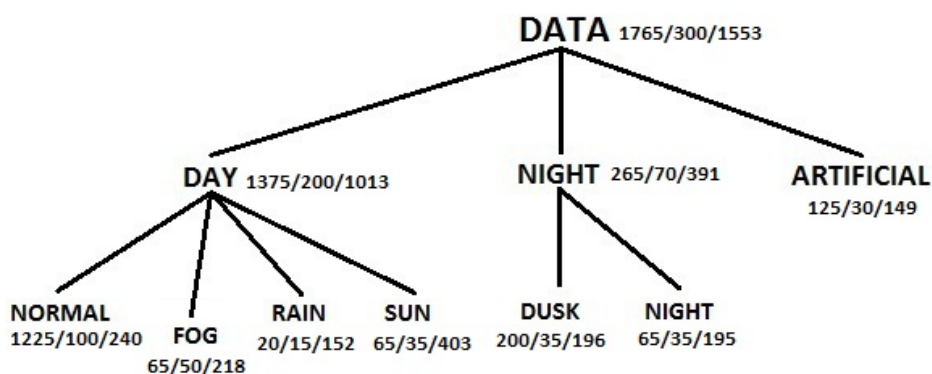
Slika 3.7: Izolacija znakov v sliki besedila.

- Slike znakov: Kot rečeno se po detekciji znakov iz slike besedila vsakega posebej shrani v lastno sliko ter ustrezno označi (slika samostojnega znaka je prikazana na sliki 3.8). Primer imena slike znaka „x“ izgleda takole: d-r-00001-char-x-lc-001.jpg (kjer char pomeni, da gre za znak, x pa predstavlja izbrani znak, lc pomeni da gre za mali znak, 001 pa pove, da gre za prvo ponovitev tega znaka). Iz slike smo vsak znak posebej izrezali ročno ter ga shranili v lastno sliko.



Slika 3.8: Samostojna slika izrezanega znaka iz besedila.

Imamo torej tri vrste slik, ki se lahko uporabljajo za različne namene. Celotne slike so namenjene predvsem testiranju metod za detekcijo teksta v slikah naravnih scen, slike besedila so namenjene sami razpoznavi črk, ki pa se naučijo prepoznavati črke s pomočjo slik znakov.

Slika 3.9: Število slik po kategorijah; baza skupaj vsebuje $x+y+z= 3618$ slik.

Baza skupaj vsebuje 3618 različnih slik tekstov v naravnih scenah (reklame, smerokazi, jedilniki ipd.). Na sliki 3.9 je število slik po mapah predstavljeno s spremenljivkami $x/y/z$, kjer „x“ pomeni število vseh zajetih slik v mapi, „y“ pomeni število slik kjer ohranimo le področje teksta, „z“ pa predstavlja končno število posameznih znakov iz slednjih slik.

3.2.1 Primerjava metodologije zajema baze CVL OCR DB z ostalimi OCR bazami

CVL OCR DB je bila primerjana z že omenjenimi bazami NIST Special Database 2, ARABASE ter InftyCDB-1. Za primerjavo so bile izbrane zgolj baze iz OCR področja z razlogom, da pridejo njihove skupne lastnosti in razlike najbolj do izraza. Primerjava vseh treh baz je prikazana v tabeli 3.4.

| | NIST | ARABASE | InftyCDB-1 | CVL OCR DB |
|-----------------------------|------|---------|------------|------------|
| končnica imena slik | / | / | Ne | Ne |
| črno-belo/barvno | Ne | Ne | Ne | Ne |
| različna svetloba | / | / | / | Da |
| velikost slike | Da | Da | Da | Da |
| ločljivost slike | Ne | Da | / | Ne |
| kvaliteta slike | / | / | / | Da |
| datum nastanka | Da | Da | / | Ne |
| gostota znakov | Da | / | / | Ne |
| globina znakov | Da | / | / | Ne |
| smer skeniranja | Da | / | Ne | Ne |
| direktna/indirektna/strojna | Ne | Da | Ne | Ne |
| ločevanje besed/podbesed | Da | Da | Da | Ne |
| znak | Da | Da | Da | Da |
| število | Da | Da | Da | Da |
| mali/veliki znak | / | / | / | Da |
| ločevanje podpisov | Da | Da | / | Ne |

Tabela 3.4: Lastnosti baz NIST, ARABASE, InftyCDB-1 ter CVL OCR DB.

3.3 Rezultati in ugotovitve

Izkazalo se je, da ima CVL OCR DB baza kar nekaj lastnosti, ki jih druge baze nimajo:

- Kvaliteta slike: Pri kvaliteti se lahko v CVL OCR DB bazi izbira med naslednjimi možnostmi (ki so pri vsaki sliki podane že v samem imenu):
 1. Deževne slike, pri katerih se poleg slabše svetlobe pojavi tudi okluzija slik zaradi dežnih kapljic, ki lahko v močnem nalivu opazno zmanjšajo kvaliteto slike.
 2. Meglene slike, pri katerih je izstopajoča zmanjšana vidljivost, ki prav tako vpliva na končno kvaliteto slike.
 3. Sončne slike, pri katerih se zaradi močnega sonca pojavijo nezaželeni odbleski. Bodisi direktni odbleski, če je v sami sliki zajeto sonce, bodisi odbleski, ki se odbijejo od površin ter tako drastično zmanjšajo kvaliteto slike.
 4. Normalne slike so slike v normalnih, pretežno oblačnih vremenskih razmerah. Take slike so posledično tudi najbolj kvalitetne.

Torej ugotovimo lahko, da vremenske razmere močno vplivajo na kvaliteto slik, saj posledično zaradi njih dobimo kar štiri nove kategorije kvalitete.

- Različna svetloba: Pri tej lastnosti, ki jo prav tako lahko razberemo že iz samega imena slike, se izbira med štirimi različnimi jakosti svetlobe. Slike so bile namreč (z izjemo slikanja pri umetni svetlobi) zajete ob različni urah v dnevu, s tem da je bil upoštevan tudi letni čas slikanja (september, oktober). Lastnost časa slikanja je definirana takole:
 1. podnevi: slikano od osme ure zjutraj do šeste ure zvečer
 2. ob mraku: slikano od šeste do osme ure zvečer
 3. ponoči: slikano od osme ure zvečer do sončnega vzhoda
 4. ob umetni svetlobi: slikano kadarkoli
- Ločevanje malih in velikih črk: Ta lastnost se lahko tako kot prejšnji dve, prav tako razbere že iz imena slike, saj z oznako „lc“ označimo male črke, z oznako „uc“ pa velike.

Vendar pa ji primanjkuje tudi še nekaj dodatnih lastnosti kot so na primer:

- ločljivost slike
- gostota znakov

- globina znakov
- ločevanje barvnih ter črno-belih slik
- datum zajetja slike ipd.

Poglavje 4

Primer detekcije in razpoznavе teksta v naravnih scenah

Slike naravnih scen, ki jih vsebuje podatkovna zbirka CVL OCR DB, med drugim služijo testiranju metod detekcije in razpoznavе teksta. To poglavje ilustrira uporabo baze v ta namen.

4.1 Detekcija teksta v naravnih scenah

Detekcija teksta v naravnih scenah je dokaj novo področje, ki se ukvarja s problemom iskanja regij na sliki, ki ustrezajo dejanskemu tekstu [30,31,32]. Problem detekcije teksta na slikah naravnih scen je izredno težaven, saj te vsebujejo ogromno število različnih tekstov, ki se razlikujejo glede na slog oz. velikost pisave. Prav tako so slike, kot že omenjeno, zajete ob različnih vremenskih in svetlobnih pogojih (sonce, tema, dež, megla ipd.) ter pod različnimi koti (prisotnost afinih, perspektivnih in projektivnih transformacij), kar prinaša dodatne probleme. Nenazadnje slike naravnih scen vsebujejo kompleksna ozadja, ki jih klasifikator zlahka napačno interpretira kot tekst. Predlaganih je bilo veliko metod za detekcijo teksta, ki se bolj ali manj uspešno spopadajo z omenjeno problematiko. Med drugim omenjene metode temeljijo na detekciji robov (angl. edge detection), binarizaciji (angl. binarization), analizi prostorske frekvence (angl. spatial-frequency image analysis) ter matematični morfologiji (angl. mathematical morphology) [32].

4.2 Razpoznavna teksta na slikah naravnih scen

Preden se poglobimo v OCR na slikah naravnih scen se ustavimo najprej pri klasičnem OCR sistemu.

4.2.1 Klasični OCR

Razpoznavna teksta na slikah naravnih scen ima korenine v klasičnem OCR (angl. Optical Character Recognition), ki se ukvarja z avtomatično razpoznavo pisanih oz. tiskanih besedil (večinoma) zajetih z uporabo računalniških skenerjev. Klasični OCR temelji na metodah računalniškega vida ter strojnega učenja in se v grobem deli na tri osnovne korake [33]:

- predprocesiranje (angl. preprocessing)
- ekstrakcija značilk (angl. feature extraction)
- klasifikacija (angl. classification)

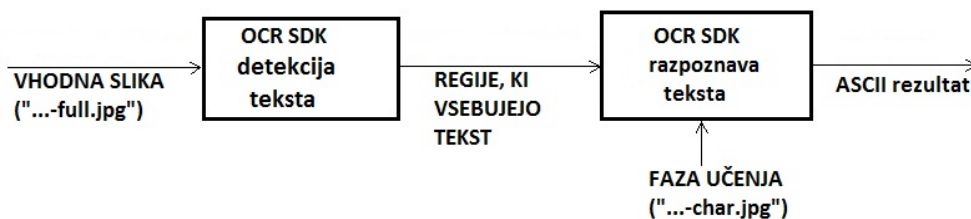
V fazi predprocesiranja se slike posameznih črk oz. znakov predhodno obdelajo. Obdelava vključuje uporabo različnih filtrov, uprakovanja (angl. thresholding), binarizacije (angl. binarization), normalizacije (angl. normalization) ipd. Predprocesirana slika vstopa v fazo ekstrakcije značilk, kjer se izračunajo določene značilnosti znaka. Med tipične značilke sodijo momenti, histogrami, Fourierjevi deskriptorji ipd. Faza klasifikacije na podlagi izračunanih značilk ustrezno ugotovi, za kateri znak gre. Tipični klasifikatorji, ki se uporabljajo, so nevronske mreže (angl. ANN), metoda podpornih vektorjev (angl. SVM), metoda najbližjih sosedov (angl. K-NN) ipd. Klasifikator je potrebno predhodno naučiti, katere vrednosti značilk ustrezajo določenim znakom. Ta postopek se izvede v fazi učenja.

4.2.2 OCR na slikah naravnih scen

Razpoznavna teksta na slikah naravnih scen je mnogo kompleksnejša od klasičnega OCR, saj so pogoji zajema izredno problematični. Dodatno razpoznavo teksta oteži sam koncept naravnih scen, ki za razliko od skeniranih dokumentov (ti večinoma vsebujejo črno besedilo na beli podlagi) vsebuje izjemno kompleksna ozadja ter široko paleto raznovrstnih slogov, velikosti in barv pisav.

4.3 Primer detekcije in razpoznavne teksta

Slike iz podatkovne zbirke CVL OCR DB smo uporabili za testiranje metod detekcije in razpoznavne teksta, ki so implementirane v posebni OCR SDK knjižnici, ki je nastala v Laboratoriju za računalniški vid na Fakulteti za računalništvo in informatiko v Ljubljani. Postopek testiranja je prikazan na sliki 4.1. Slike posameznih znakov („...-char.jpg“) smo uporabili v fazi učenja OCR klasifikatorja. Primeri slik posameznih znakov so prikazani na sliki 4.2. Nato smo slike celotnih scen („...-full.jpg“) kot vhod poslali modulu za detekcijo teksta. Ta je na sliki detektiral potencialne regije, torej regije, kjer se po vsej verjetnosti nahaja tekst. Primer dobljenih regij, ki vsebujejo tekst, je prikazan na sliki 4.3. Regije, ki vsebujejo tekst, predstavljajo vhod modula za razpoznavo teksta (OCR SDK). Ta dobljeno sliko analizira in vrne rezultat v ASCII obliki. Primeri razpoznavne teksta so prikazani na sliki 4.4.



Slika 4.1: Diagram testiranja.



Slika 4.2: Primeri slik posameznih črk oziroma znakov.



a.)



b.)



c.)



d.)



e.)



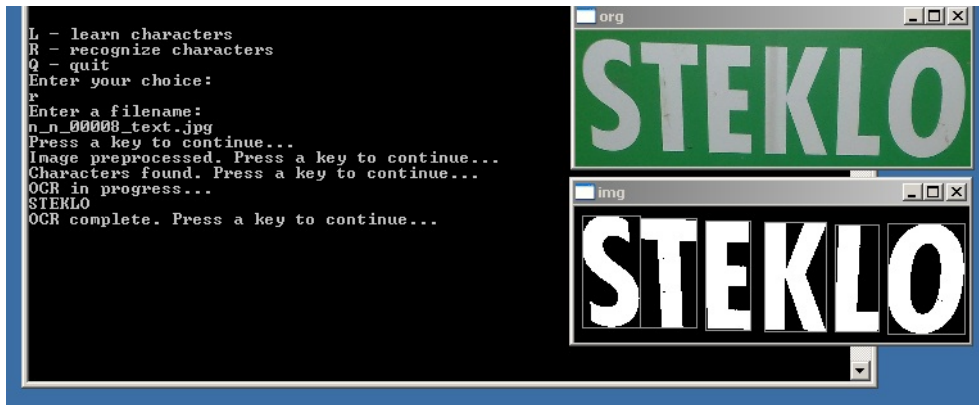
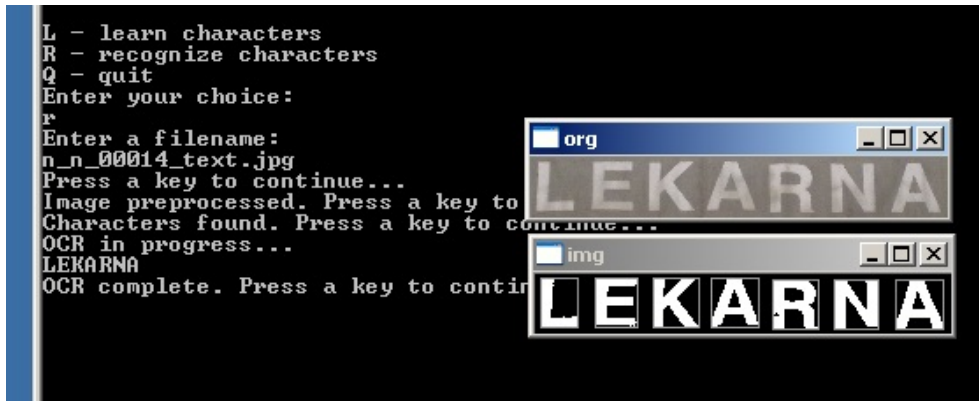
f.)



g.)

Slika 4.3: Primer detekcije teksta: a.-c.) pravilna detekcija, d.-g.) najdene tudi regije, ki ne ustrezajo tekstu.

46 POGLAVJE 4. PRIMER DETEKCIJE IN RAZPOZNAVE TEKSTA V NARAVNIH SCENAH



Slika 4.4: Primeri razpoznavanja teksta: razpoznani znaki so zapisani v predzadnji vrstici terminalnega okna, ki je v ozadju slik.

Poglavje 5

Zaključek in nadaljne delo

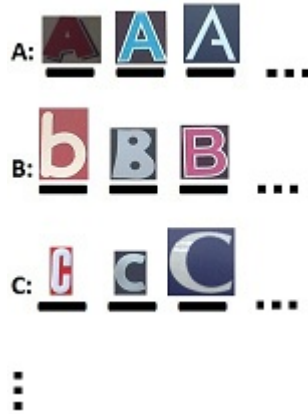
Iz zapisanega je razvidno, da je razvoj in uporaba ustrezne metodologije zajema za slikovne baze izredno pomembna. Tako lahko najbolj ustrezno določimo kategorije po katerih razdelimo slike. V okviru diplomskega dela je predstavljena kopica bolj znanih slikovnih baz, ki se uporabljajo v računalniškem vidu. Vsaka izmed njih ima svojo lastno metodologijo zajema, nekatere so enostavne, druge kompleksnejše, vendar že po sami zasnovi kategorij vidimo, da obstaja med bazami kar nekaj skupnih točk. Pri masovni primerjavi baz oz. primerjanju njihovih metodologij pa se lahko zelo hitro izpostavijo tudi njihove pomanjkljivosti. Na dosti področjih je ustvarjenih izredno malo slikovnih baz, bodisi recimo, ker so sama področja relativno nova (na primer prepoznavanje ušes), bodisi, ker je samo zajemaje slik predrago in omejeno samo na specializirane laboratorije. Veliko baz tudi ni enostavno dosegljivih. Prav tako bi omenil, da ima premalo baz zajete slike z večimi različnimi kamerami. Vsekakor pa je zasnova slikovnih baz oz. njihova metodologija zajema velikokrat odvisna od samih potreb algoritmov, ki jo bodo uporabljali.

Baza CVL OCR DB bo služila svojemu osnovnemu namenu testiranja razvitih algoritmov za razpoznavanje teksta v naravnih scenah, hkrati pa bo javno dostopna.

Kot nadaljne delo na slikovni bazi CVL OCR DB pa bi izpostavil naslednje možnosti:

- Več podkategorij baze z lastnostmi kot so: ločljivost slike, gostota znakov, globina znakov, časovna oznaka slike, razlikovanje kamer, različni koti slikanja, ločevanje barvnih in črno belih slik, več formatov slik, poševne črke, pisane črke.

- Podbaza abecede: Vse slike izoliranih znakov pridejo v tabelo abecede. Tako bi za vsak znak obstajala svoja tabela slik vseh možnih oblik tega znaka (glej sliko 5.1).



Slika 5.1: Tabela slik znakov urejenih po abecedi.

- Relacije: Dolgoročno je namen ustvariti oz. nadgraditi bazo tudi v smislu relacij posameznih znakov oz. besed ter njihovih ASCII zapisov.

Slike

| | | |
|------|---|----|
| 1.1 | Prikaz hierarhičnega modela baze | 4 |
| 2.1 | Delna slika davčnega obrazca baze NIST Special Database 2 . . . | 7 |
| 2.2 | Prva stopnja drevesa v bazi | 7 |
| 2.3 | Prikaz členitve drugega reda | 8 |
| 2.4 | Skupek možnih tipov vnosnih polj | 8 |
| 2.5 | Skupek možnih kontekstnih vnosnih polj | 9 |
| 2.6 | Struktura Arabasa | 10 |
| 2.7 | Primer slikovne datoteke baze InftyCDB-1 | 12 |
| 2.8 | Vzorčna zbirka slik obrazov podatkovne baze YALE | 15 |
| 2.9 | Standardno zajeta slika in Termično zajeta slika | 16 |
| 2.10 | Potek zajemanja 3D slike baze CASIA 3D Face Database | 16 |
| 2.11 | Primeri 3D slik iste osebe baze CASIA 3D Face Database | 17 |
| 2.12 | Slikanje očesa za bazo CASIA-IrisV3 ter lociranje šarenice | 18 |
| 2.13 | Tipična osnovna zgradba baze CASIA | 19 |
| 2.14 | Primer prstnega odtisa z že označenimi značilnimi točkami | 20 |
| 2.15 | Primerjanje značilnih točk med različnimi prstnimi odtisi | 20 |
| 2.16 | 3D prstni odtis iz univerzitetne baze v Kentuckyju prikazan iz različnih zornih kotov | 21 |
| 2.17 | Podpis na e-tablico | 22 |
| 2.18 | Primer slike dlani iz baze IIT Delhi Touchless Palmprint | 24 |
| 2.19 | Potek značilnih robov na dlani | 24 |
| 2.20 | Tipičen biometričen sistem na osnovi gibanja | 25 |
| 2.21 | Sistem prepoznave obraza v kombinaciji z ušesi | 26 |
| 2.22 | Točke interesa prikazane nad sliko ušesa v bazi XM2VTS | 27 |
| 3.1 | Naključna slika izoliranega okvirja baze NIST Special Database 6 | 33 |
| 3.2 | Primer slike naslovnega okvirja v bazi NIST Special Database 6 | 34 |
| 3.3 | Osnovna drevesna zgradba baze | 35 |
| 3.4 | Organizacija slik in datotek v bazi Nist Special Database 6 | 36 |

| | | |
|-----|---|----|
| 3.5 | Drevesna struktra osnovne zgradbe baze CVL OCR DB | 36 |
| 3.6 | Označitev območja besedila za izrez v celotni sliki | 37 |
| 3.7 | Izolacija znakov v sliki besedila | 38 |
| 3.8 | Samostojna slika izrezanega znaka iz besedila | 38 |
| 3.9 | Število slik po kategorijah | 38 |
| 4.1 | Diagram testiranja | 44 |
| 4.2 | Primeri slik posameznih črk oziroma znakov | 44 |
| 4.3 | Primer detekcije teksta | 45 |
| 4.4 | Primeri razpoznavne teksta | 46 |
| 5.1 | Tabela slik znakov urejenih po abecedi | 48 |

Tabele

| | | |
|-----|--|----|
| 2.1 | Prikaz atributov v bazah NIST, ARABASE ter InftyCDB-1 . . . | 13 |
| 2.2 | Primerjava baz FERET, YALE in CVL | 17 |
| 2.3 | Primerjava baz CASIA-IrisV3 in UBIRIS.v2. | 19 |
| 2.4 | Primerjava med bazama FVC 2004 in NIST4 | 22 |
| 2.5 | Primerjava baz SVC 2004 in GPDS300 | 23 |
| 2.6 | Primerjava baz CASIA in Delhi | 25 |
| 2.7 | Primerjava baz Southampton in CASIA | 26 |
| 2.8 | Primerjava med bazama MIT in XM2VTS | 27 |
| 3.1 | Statistika in značilnosti slikovnih podbaz CASIA-IrisV3-Interval, CASIA-IrisV3-Lamp, CASIA-IrisV3-Twins | 29 |
| 3.2 | Dvočrkovne kode, ki označujejo vrsto slike v bazi FERET | 30 |
| 3.3 | Glava slikovne datoteke, ki prikazuje vrednosti posameznih atrib- utov v NIST Special Database 6 | 33 |
| 3.4 | Lastnosti baz NIST, ARABASE, InftyCDB-1 ter CVL OCR DB | 39 |

Literatura

- [1] D. L. Dimmick, M. D. Garris, C. L. Wilson, *Nist Special Database 2, Structured Forms Database Users' Guide* National Institute of Standards and Technology, Advanced Systems Division, Image Recognition Group, December 1, 1991 (Dostopno na: <http://www.nist.gov/ts/msd/srd/nistsd2.cfm>).
- [2] N. B. Amara, O. Mazhoud, N. Bouzrara, N. Ellouze, *ARABASE: A Relational Database for Arabic OCR Systems*, National school of Engineer of Monastir, Tunisia, 17. 09. 2004 (Dostopno na: <http://www.ccis2k.org/iajit/PDF/vol.2,no.4/1-Najoua.pdf>).
- [3] M. Suzuki, *A Ground Truth Database of Characters, Symbols words and Formulas in Mathematical Documents; First Distribution*, Faculty of Mathematics and Graduate School of Mathematics, Kyushu University, 18. 05. 2005 (Dostopno na: <http://www.inftyproject.org/en/database.html>).
- [4] L. Schomaker, I. Guyon, *Unipen*, Ecole Polytechnique de Montreal, Montreal, 9. 1992 (Dostopno na: <http://www.unipen.org/>).
- [5] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, J. M. Bone, *Face Recognition Vendor Test*, 3. 2002 (Dostopno na: <http://www.frvt.org/FRVT2002/Default.htm>).
- [6] A. Socolinsky, B. Wolffz, D. Neuheisely, K. Evelandz, *Illumination Invariant Face Recognition Using Thermal Infrared Imagery*, Equinox Corporation (Dostopno na: <http://www.equinoxsensors.com/publications/face.pdf>).
- [7] Y. Wang, *Note on CASIA 3D Face Database*, Center for Biometrics and Security Research, Institute of Automation, Chinese Academy of Sciences (Dostopno na: <http://www.cbsr.ia.ac.cn/english/Databases.asp>).

- [8] P. J. Phillips, Pentland, *Face Recognition Technology (FERET) program* (Dostopno na: <http://www.frvt.org/FERET/default.htm>).
- [9] A. S. Georghiades, P. N. Belhumeur, D. J. Kriegman, *From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose*, IEEE Trans. Pattern Anal. Mach. Intelligence, 23 (6), str. 643-660, 2001 (Dostopno na: <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>).
- [10] P. Peer, *CVL Face Database*, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, (Dostopno na: <http://www.lrv.fri.uni-lj.si/facedb.html>).
- [11] T. Tan, *Note on CASIA-IrisV3*, Center for Biometrics and Security Research, National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences (Dostopno na: <http://www.cbsr.ia.ac.cn/english/IrisDatabases.asp>)
- [12] SOCIA (Soft Computing and Image Analysis) Group, *UBIRIS Noisy Visible Wavelength Iris Image Databases*, Department of Computer Science, University of Beira (Dostopno na: <http://iris.di.ubi.pt/ubiris1.html>).
- [13] Y. Wang, L. Hassebrook, D. Lau, *Noncontact, depth-detailed 3D fingerprinting*, University of Kentucky Lexington, 31. 12. 2009 (Dostopno na: <http://spie.org/x38623.xml?ArticleID=x38623>).
- [14] Biometric Systems Lab, Pattern Recognition and Image Processing Laboratory, Biometric Test Center, *FVC2004: the Third International Fingerprint Verification Competition*, University of Bologna, Michigan State University, San Jose State University, 2004 (Dostopno na: <http://bias.csr.unibo.it/fvc2004/>).
- [15] C. J. Watson, C. L. Wilson, *NIST Special Database 4 Fingerprint Database*, National Institute of Standards and Technology Advanced Systems Division Image Recognition Group, 17. 05. 1992, (<http://www.nist.gov/srd/nistsd4.cfm>).
- [16] D. Y. Yeung, *SVC 2004: First International Signature Verification Competition*, Hong Kong University of Science and Technology, Hong Kong, 2004 (Dostopno na: <http://www.cse.ust.hk/svc2004/download.html>).

- [17] F. Vargas, A. Ferrer, M. Travieso, B. Alonso, *Off-line Handwritten Signature GPDS-960 Corpus*, IAPR 9th International Conference on Document Analysis and Recognition, str. 764-768, Curitiba, Brazil, 9. 2007 (Dostopno na: <http://www.gpds.ulpgc.es/download/index.htm>).
- [18] Biometric Research Centre, The Hong Kong Polytechnic University, 2006, (Dostopno na: <http://www4.comp.polyu.edu.hk/biometrics/>).
- [19] H. Hom, *The Hong Kong Polytechnic University (PolyU) Palmprint Database*, Biometric Research Centre (UGC/CRC), The Hong Kong Polytechnic University, 14. 11. 2007, (Dostopno na: <http://rate.pku.edu.cn/RATE/showdatabase?database=1012>).
- [20] T. Tan, *Note on CASIA Palmprint Database*, Center for Biometrics and Security Research National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences (Dostopno na: <http://www.cbsr.ia.ac.cn/english/Databases.asp>).
- [21] Biometrics Research Laboratory, *IIT Delhi Palmprint Image Database version 1.0*, IIT Delhi, 10. 2007 (Dostopno na: <http://web.iitd.ac.in/ajaykr/Databasepalm.htm>).
- [22] M. S. Nixon, *The Southampton Human ID at a Distance Gait Database*, University of Southampton, 15. 2002, (Dostopno na: <http://www.gait.ecs.soton.ac.uk/database/>).
- [23] S. Zheng, *CASIA Gait Database*, National Laboratory of Pattern Recognition, 7. 2005 (Dostopno na: <http://www.cbsr.ia.ac.cn/english/Databases.asp>).
- [24] A. Zavar, S. Nixon, J. Hurley, *On Model-Based Analysis of Ear Biometrics*, IEEE Conference on Biometrics: Theory, Applications and Systems, 9. 2007, Washington (Dostopno na: <http://eprints.ecs.soton.ac.uk/14888/>).
- [25] A. A. Darwish, R. Abd Elghafar, A. Fawzi, *Multimodal Face and Ear Images*, Department of Mathematics, Faculty of Science, Helwan University, Cairo, Egypt, 2009 (Dostopno na: <http://www.scipub.org/fulltext/jcs/jcs55374-379.pdf>).
- [26] *The PASCAL Object Recognition Database Collection*, (Dostopno na: <http://pascallin.ecs.soton.ac.uk/challenges/VOC/databases.html>).

- [27] A. Pronobis, J. Luo, *The KTH-IDOL2 Database*, 2006 (Dostopno na: <http://cogvis.nada.kth.se/IDOL/>).
- [28] D. Omerčević, O. Drbohlav, A. Leonardis, *High-Dimensional Feature Matching: Employing the Concept of Meaningful Nearest Neighbors*, ICCV, Rio de Janeiro, Brazil, 2007 (Dostopno na: <http://vicos.fri.uni-lj.si/downloads/luis34/>).
- [29] D. L. Dimmick, M. D. Garris, *Nist Special Database 6*, National Institute of Standards and Technology, Advanced Systems Division, Image Recognition Group, 16. 9, 1992 (Dostopno na: <http://www.nist.gov/srd/nistsd6.cfm>).
- [30] D. Doermann, J. Liang, H. Li, *Progress in camerabased document image analysis.*, In Proc. of 7th Int. Conf. on Document Analysis and Recognition (ICDAR 2003), zbornik I, str. 606–616, Edinburgh, Scotland, IEEE Press, 8. 2003.
- [31] J. Yang, J. Gao, Y. Zang, X. Chen, A. Waibel, *An automatic sign recognition and translation system*, In Proceedings of the Workshop on Perceptive User Interfaces (PUI'01), 11. 2001.
- [32] N. Ezaki, M. Bulacu, L. Schomaker, *Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons*, Proc. of 17th Int. Conf. on Pattern Recognition (ICPR 2004), IEEE Computer Society, str. 683-686, zbirnik II, Cambridge, UK, 8. 2004.
- [33] M. Cheriet, N. Kharma, C. L. Liu, C. Y. Suen, *Character Recognition Systems: A Guide for Students and Practitioners*, Wiley-Interscience, 2007.