

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Projekt iz predmeta Raspoznavanje uzoraka

Ak. god. 2014/15

Raspoznavanje aktivnosti osoba na temelju siluete

Kuman Stipe

Meštrović Stjepan

Mujić Azzaro

Mužar Irena

Novković Igor

Skukan Luka

Venanzoni Andrea

Zagreb, siječanj 2017.

SADRŽAJ

1. Projektni zadatak	1
1.1. Opis projektnog zadatka	1
1.2. Pregled i opis srodnih rješenja	1
1.3. Konceptualno rješenje zadatka	3
1.3.1. Baza	3
1.3.2. Arhitektura sustava	5
2. Postupak rješavanja zadatka	8
2.1. Izdvajanje pozadine	8
2.1.1. Transformacija slike u boji u sliku intenziteta	8
2.1.2. Statistički model izdvajanja pozadine	8
2.2. Reprezentacija siluete	9
2.3. Izlučivanje vektora značajki	9
2.4. Klasifikacija	11
2.5. Vremensko zaglađivanje	12
3. Ispitivanje rješenja	13
3.1. Ispitna baza	13
3.2. Rezultati učenja i ispitivanja	14
3.3. Analiza rezultata	15
4. Opis programske implementacije rješenja	16
5. Zaključak	17
6. Literatura	18

1. Projektni zadatak

(do 10 stranica)

1.1. Opis projektnog zadatka

Opis problema koji zadatak obuhvaća. Što su ulazni podaci, a što zahtjevani izlaz?
Koncepti/algoritmi koji se obavezno moraju upotrijebiti?

1.2. Pregled i opis srodnih rješenja

U nastavku slijedi opis radova srodnih tema koji će se koristiti kao materijal za učenje.

Andrea Bottino, Aldo Laurentini

A silhouette based technique for the reconstruction of human movement

Invazivni senzorni sustavi obično se temelje na magnetskom i optičkom praćenju pokreta i podrazumijevaju veće ili manje objekte pričvršćene na tijelu izvođača i pritom ga mogu ometati. Za neke primjene, kao što je npr. analiza sportske izvedbe, to može imati velik negativni utjecaj. U ovom se radu opisuje neinvanzivna tehnika kojom se može rekonstruirati prirodni pokret i kojom se nastoji riješiti navedeni problem.

Tehnikom presjeka volumena obavlja se 3D rekonstrukcija pokreta iz slika snimljenih običnom kamerom sa različitih gledišta. Podatci o pokretu prikupljaju se projiciranjem modela na rekonstruirani volumen.

Rekonstruiranje 3D oblika iz 2D silueta je popularan pristup u računalnom vidu. Model se sastoji od dvije komponente: prikaza kostura i prikaza tijela koje ga okružuje. Kostur ima 15 segmenata koji su povezani zglobovima u obliku sfere. Tijelo oko kostura je prikazano sa približno 600 trokuta.

Siluate se izlučuju oduzimanjem pozadine. Presjek volumena algoritam je primjenjiv za različite rezolucije i daje granice rekonstruiranog volumena. Određivanje posture tijela temelji se na pretraživanju 32-dimenzionalnog prostora parametara, i podrazumijeva pronalaženje poze modela koji najbolje aproksimira stvarni pokret. Sustav je testiran u virtualnom okruženju i na stvarnom slijedu slika.

Meghna Singh, Anup Basu, Mrinal Kr. Mandal

Human activity recognition based on silhouette directionality

I ovaj rad, također, daje opis neinvanzivne tehnike kojom se rekonstruira pokret. Koristi adaptivnu tehniku odvajanja pozadine za izlučivanje informacija i generiranje silueta iz snimke. Iz kontura siluete se, tada, određuju usmjereni vektori i, za grupiranje i raspoznavanje, koristi se distribucija podataka vektora smjera. Iskorištavaju se dinamičke karakteristike ljudskog pokreta da bi se zagladile odluke i smanjile pogreške u prepoznavanju pokreta.

Za izdvajanje siluete koristi se statistički model pozadine, računanje očekivanja i varijance intenziteta svakog slikovnog elementa na slici na kojoj se nalazi samo pozadina. Ta je metoda otpornija na šum, sjenu i promjenu svjetla nego oduzimanje pozadine.

Silueta je obično pod utjecajem šuma i može se sastojati od nepovezanih komponentata. Da bi se dobila „čista“ silueta, koriste se morfološke operacije i analiza povezanih komponenti. Kontura siluete prikazana je lančanim kodom iz kojeg se onda određuju vektori smjera. Dobiveni vektori se normaliziraju.

Pretpostavka je da vrijedi:

1. Normalizirani vektori smjera, za različite izvođače za iste aktivnosti na jednakoj udaljenosti od kamere, imaju malu varijancu.
2. Normalizirani vektori smjera, za različite izvođače za iste aktivnosti na različitim udaljenostima od kamere i s promjenjivom pozadinom, imaju malu varijancu.
3. Varijanca normaliziranih vektora smjera za različite aktivnosti je velika.

Vektori se grupiraju k-centroid algoritmom grupiranja (eng. *k-means clustering*). Zbog mogućeg lošeg izdvajanja siluete, može doći do velikih varijanci vektora smjera susjednih okvira i uzrokovati pogrešne odluke te se, stoga, koristi zaglađivanje. Točnost algoritma varira između 85% i 99% za osam aktivnosti, kada se ne koristi zaglađivanje. Kada se koristi zaglađivanje, točnost raste na 100%. Ovaj algoritam dobro radi i

za periodični i za neperiodični pokret, ne koristi uspoređivanje predložaka i ne zahtjeva praćenje povijesti za raspoznavanje. Algoritam se koristi za određivanje pet osnovnih aktivnosti: uspravno stajanje, sjedenje, čučanje, ležanje i pokazivanje.

Liang Wang, Weiming Hu, Tieniu Tan

Recent developments in human motion analysis

U ovom se radu daje opsežan opis radova sa područja ljudskog pokreta od 1989. do danas. Fokusira se na općeniti pregled aktivnosti koje uključuje analiza ljudskog pokreta: detektiranje, praćenje čovjeka te razumijevanje ljudskog ponašanja.

Detekcija čovjeka podrazumijeva odvajanje segmenta slike koji se odnosi na čovjeka od ostatka slike. Oduzimanje pozadine jedna je od popularnijih metoda segmentacije pokreta, posebice u slučajevima relativno statične pozadine. S druge strane, statističke metode koriste slikovni element ili grupe slikovnih elemenata za konstruiranje naprednijih pozadinskih modela, a statistika pozadine se može dinamički ažurirati.

Cilj klasifikacije objekta u pokretu je odvojiti područje koje odgovara čovjeku od svih ostalih objekata navedenim metodama segmentacije. Dvije su glavne kategorije: klasifikacija na temelju oblika i klasifikacija na temelju pokreta.

Praćenje pokreta služi za pripremu procjene poze i raspoznavanje akcije. Može se podijeliti na četiri kategorije: praćenje na temelju modela (model kostura, 2D konture, volumni modeli), praćenje na temelju područja, praćenje na temelju aktivne konture te praćenje na temelju značajki.

Razumijevanje ljudskog pokreta uključuje raspoznavanje akcija te njihov opis. Opće tehnike koje se koriste za usporedbu vremenski zavisnih podataka su: dinamično savijanje vremena (eng. *Dynamic time-warping*), skriveni markovljevi modeli i neuronske mreže. Metode za raspoznavanje akcija su: uspoređivanje predložaka, pristup prostoru stanja i semantički opis.

1.3. Konceptualno rješenje zadatka

1.3.1. Baza

Baza koju ćemo koristiti će biti Weizman baza¹. Baza se sastoji od 90 snimaka podijeljenih u 10 aktivnosti, znači sadrži 9 snimaka za svaku aktivnost. Aktivnosti sadržane u bazi koje se mogu prepoznavati su:

¹Dostupna na <http://www.wisdom.weizmann.ac.il/vision/SpaceTimeActions.html>

1. hodanje✱,
2. trčanje✱,
3. skakanje,
4. hodanje u stranu,
5. savijanje✱,
6. mahanje jednom rukom✱,
7. mahanje s obje ruke,
8. skakanje na mjestu✱,
9. skakanje s ispruženim rukama,
10. preskakivanje

U sklopu ovog projekta korišteno je samo 5 od 10 aktivnosti, odnosno one aktivnosti označene zvijezdicom (✱) u prethodnoj listi.

Snimke su snimane statičnom kamerom rezolucije 180 x 144 piksela. Na snimkama je prisutan šum u obliku hodanja s psom ili nošenja torbe. Snimke sadrže homogenu pozadinu radi lakšeg izdvajanja pozadine od siluete.



Slika 1.1: Primjer hodanja i pripadne siluete

Primjer hodanja i pripadne siluete U aplikaciji nećemo koristiti sve aktivnosti za prepoznavanje, već ćemo se, radi jednostavnosti, ograničiti na dvije ili tri aktivnosti. Na slikama 1.1 i 1.2 mogu se vidjeti primjeri snimaka hodanja i trčanja i pripadnih silueta.



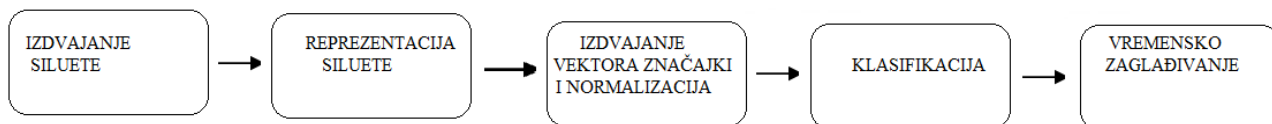
Slika 1.2: Primjer trčanja i pripadne siluete

1.3.2. Arhitektura sustava

Sustav možemo podijeliti na sljedećih 5 modula:

- izdvajanje siluete,
- reprezentacija siluete,
- izdvajanje vektora značajki i normalizacija,
- klasifikacija,
- vremensko zaglađivanje

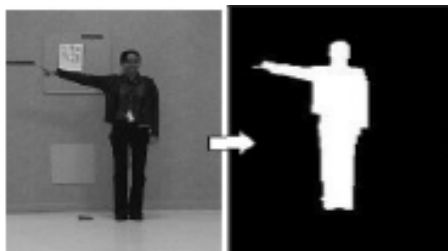
Slika 1.3 prikazuje blok shemu sustava.



Slika 1.3: Blok shema sustava

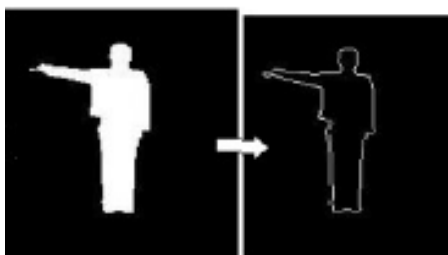
Ulaz u sustav je snimka na kojoj osoba obavlja određenu aktivnost. Na svakoj slici snimke se izdvaja silueta od pozadine. Prije izdvajanja siluete slika se iz RGB sustava prebacuje u sustav intenziteta sive boje postupkom određivanja prosjeka R, G i B komponente. Silueta se izdvaja od pozadine statističkim modelom pozadine, gdje se računa očekivanje i varijanca svakog piksela pozadine početnih N slika, na kojima osoba nije prisutna. Ova metoda je robusnija od jednostavnijih metoda, poput oduzimanja pozadine. Za svaki piksel se računa prag kao maksimalna razlika očekivanja i intenziteta sive boje kod početnih N slika. Pripada li piksel silueti se određuje na sljedeći način:

- ako je razlika intenziteta piksela i očekivanja intenziteta piksela manja od praga, piksel pripada pozadini
- inače piksel pripada silueti



Slika 1.4: Izdvajanje siluete od pozadine

Zatim se izlučuje kontura siluete primjenom Gaussovog filtra. Izlaz prvog modula je slika koja sadrži konturu siluete.



Slika 1.5: Izlučivanje konture siluete

Drugi modul radi reprezentaciju konture siluete pomoću lančanog koda. Lančani kod govori u koju stranu se moramo pomaknuti da od jednog piksela konture dođemo do sljedećeg piksela konture. Na slici 1.6 prikazan je lančani kod sa sljedećim oznakama:

- kod 0 ima piksel koji je iznad trenutnog piksela,
- kod 1 ima piksel koji je gore-desno od trenutnog piksela,
- kod 2 ima piksel koji je desno od trenutnog piksela,
- kod 3 ima piksel koji je dolje-desno od trenutnog piksela,
- kod 4 ima piksel koji je ispod trenutnog piksela,
- kod 5 ima piksel koji je dolje-lijevo od trenutnog piksela,
- kod 6 ima piksel koji je lijevo od trenutnog piksela,
- kod 7 ima piksel koji je gore-lijevo od trenutnog piksela

Obilaženjem cijele konture siluete dobiva se vektor lančanog koda koji je izlaz ovog modula. Sljedeći modul grupira vektor lančanog koda u 8-dimenzionalni usmjereni vektor. Grupacija se radi na temelju frekvencije pojavljivanja koda. Radi postizanja robusnosti usmjereni vektori se normaliziraju. Za svaki par slika, određuje se



Slika 1.6: Prikaz lančanog koda

veličina kuta između normaliziranih vektora tih slika, te se ti kutevi šalju kao ulaz u sljedeći modul.

Za svrhu klasificiranja koristimo k-centroidni algoritam grupiranja (eng. *k-means clustering algorithm*). Pošto algoritam grupiranja spada pod model nenadziranog učenja, prvo ćemo odrediti centroeide snimaka na kojima znamo koja se aktivnost odvija. Zatim će se na nepoznatim primjerima klasifikacija vršiti uzimanjem minimalne udaljenosti centroida.

Ako se između dvije slike aktivnost naglo promijeni, moguće je da je riječ o pogrešci. Pogreška se može dogoditi ako se loše izdvojila silueta od pozadine. Tada se uspoređuje slika na kojoj se pogreška dogodila sa slikama prije i nakon nje.

Izračunavaju se nove vrijednosti piksela slike kao prosjek vrijednosti piksela slika prije i nakon nje. Na takvoj slici sa izračunatim novim vrijednostima piksela se vrši ponovna klasifikacija.

2. Postupak rješavanja zadatka

(do 10 stranica)

Navesti numerirani slijed koraka rješavanja. Npr.: 1. Dobivanje binarne slike iz slike u boji, 2. Segmentacija objekata na slici, 3. Nalaženje rubova u slici ...

2.1. Izdvajanje pozadine

Prvi korak postupka prepoznavanje aktivnosti osobe je izdvajanje siluete osobe od pozadine. Ulaz ovog modula je snimka rezolucije 180 x 144, a izlaz modula je snimka na kojoj je silueta označena bijelom bojom a pozadina crnom bojom

2.1.1. Transformacija slike u boji u sliku intenziteta

Prvi korak izdvajanja pozadine je pretvorba slike u boji u sliku intenziteta sive boje. Na taj način se postize manja razlika vrijednosti piksela te se odvajanje siluete od pozadine može provesti uspješnije. Transformacija se provodi upotrebom CCIR 601 standarda. Standard CCIR 601 prilikom transformacije sustava uzima u obzir da ljudsko oko ne vidi svaku boju jednako, neke boje vidi intenzivnije a neke slabije. Ljudsko oko zelenu boju percipira najintenzivnije, potom crvenu pa plavu. U skladu s tim transformacija se vrši otežanom sumom vrijednosti intenziteta crvene, zelene i plave boje na sljedeći način:

$$Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B. \quad (2.1)$$

2.1.2. Statistički model izdvajanja pozadine

Nakon dobivanja slike u intenzitetu sive boje, nad slikom se provodi postupak izdvajanje pozadine statističkim modelom. Statistički model se definira izračunom varijance i očekivanja intenziteta svakog piksela. Pretpostavka je da u početnih N slika

snimke će se prikazivati samo pozadina, te se varijanca i očekivanje računaju na tih N početnih slika na sljedeći način:

$$u(x, y) = \frac{1}{N} \cdot \sum (p(x, y; i)) \quad (2.2)$$

Pri tome je $u(x, y)$ očekivanje intenziteta piksela na poziciji (x, y) , $p(x, y; i)$ je vrijednost intenziteta piksela na poziciji (x, y) na slici i . Time se dobiva robusnost i točnije prepoznavanje u odnosu na jednostavnije modele oduzimanja pozadine. Nakon određivanja varijance i očekivanja intenziteta svakog piksela određuje se prag intenziteta kojeg mora preći piksel na novoj slici da bi se smatrao pikselom siluete. Prag se određuje kao maksimalna razlika očekivanja intenziteta pojedinog piksela i vrijednosti piksela na početnim slikama na temelju kojih se računala varijanca i očekivanje.

$$T(x, y) = \max |u(x, y) - p(x, y; i)|, 1 \leq i \leq N \quad (2.3)$$

Gdje je $T(x, y)$ prag intenziteta piksela (x, y) .

Potom se na slikama na kojima je objekt prisutan pripadnost piksela klasificira na sljedeći način: Pikel pripada silueti samo ukoliko vrijedi jednačba 2.4.

$$|p(x, y; k - u(x, y))| < T(x, y) \quad (2.4)$$

Prednost ovog načina izdvajanja pozadine je već spomenuta robusnost zbog mogućnosti izgradnje razdiobe za svaki piksel slike. No u slučaju da je objekt prisutan od prve slike, moguće su pogreške u izdvajanju pozadine u prvih par slika jer nije određena razdioba intenziteta.

2.2. Reprezentacija siluete

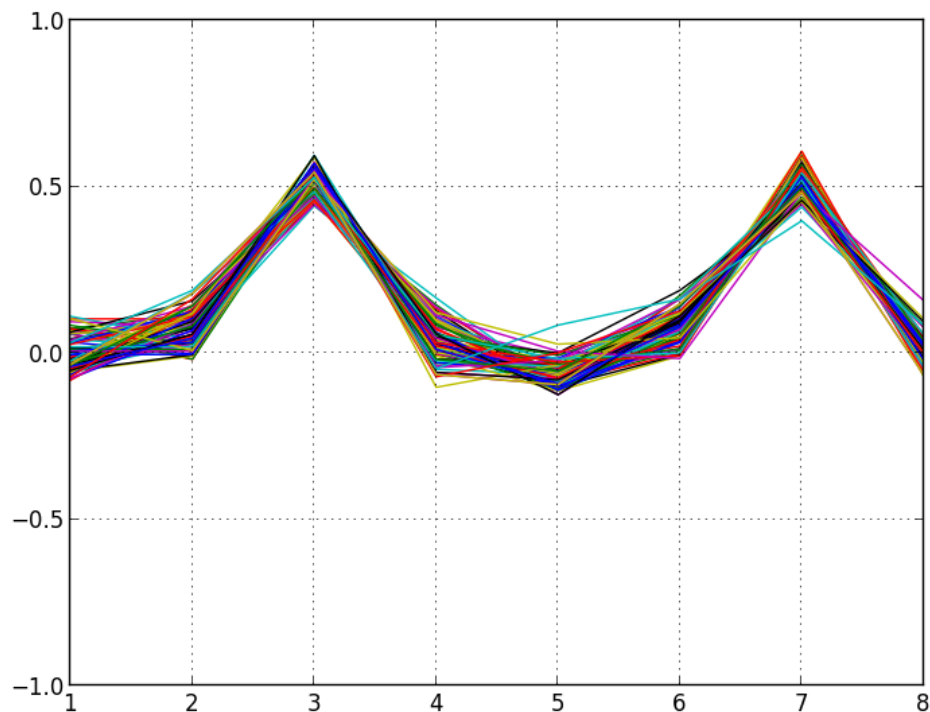
2.3. Izlučivanje vektora značajki

U ovom koraku izlučuju se korisne značajke iz chain koda siluete. Ulaz u modul opisan u ovom odjeljku je jedan ili više chain kodova, a izlaz jednak broj vektora značajki izvučenih iz tih chain kodova. Kao relevantne značajke u ovom projektu koriste se frekvencije pojavljivanja svih vrijednosti vektora smjera. To znači da će modul iz ulaznog chain koda stvoriti histogram s 8 vrijednosti (vektor smjera može poprimiti 8 vrijednosti) koje predstavljaju normalizirane vrijednosti frekvencije pojavljivanja vektora smjera.

Prvi korak modula je jednostavno prebrojavanje pojavljivanja svih vrijednosti vektora smjera. Potom se te vrijednosti normaliziraju. Normalizacijom srednja vrijednost vektora izlučenih značajki postaje nula, a vrijednosti vektora će sada biti u intervalu $[-\frac{1}{8}, 1 - \frac{1}{8}]$. Normalizacija se vrši po formuli:

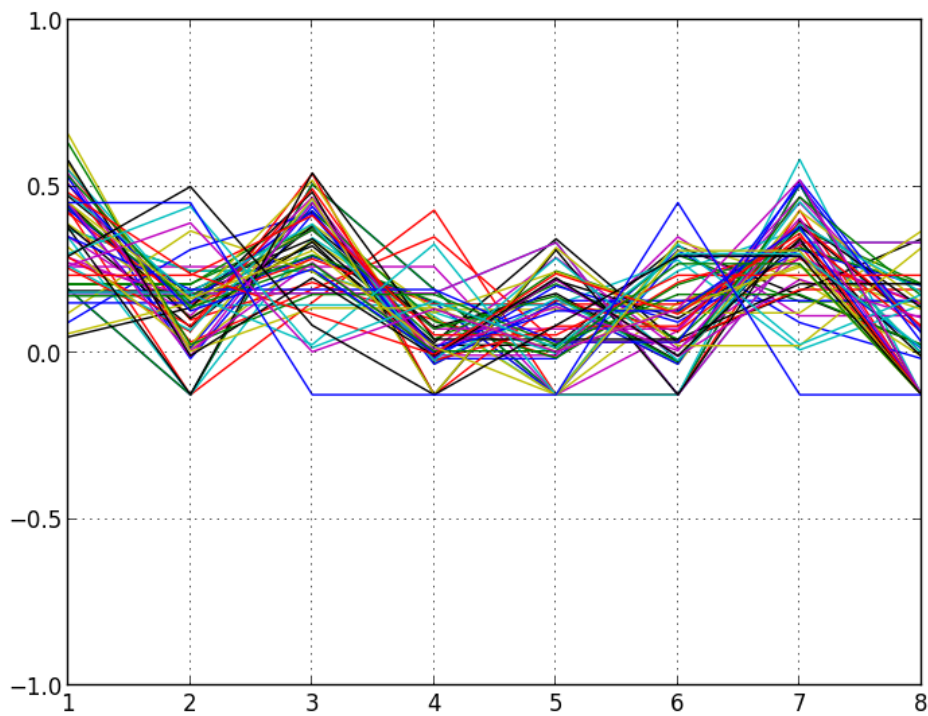
$$\phi(i) = \frac{|f(i)|}{\sum(f(i))} - \frac{1}{8} \quad (2.5)$$

$\phi(i)$ predstavlja normaliziranu frekvenciju pojavljivanja jednog vektora smjera, a $f(i)$ standardnu frekvenciju dobivenu prebrojavanjem pojavljivanja. Ta frekvencija se dijeli sa sumom frekvencija pojavljivanja svih vektora smjera i tako se svodi na vrijednost u intervalu $[0, 1]$. Dodatnim oduzimanjem vrijednosti $\frac{1}{K}$ (u ovom slučaju je $K = 8$ pa je ta vrijednost $\frac{1}{8}$) se srednja vrijednost vektora izlučenih značajki izjednačava s nulom. Normalizacijom smo postigli to da naš program ne ovisi o tome koliko je osoba koja radi neku radnju udaljena od kamere jer je sustav postao otporan na razlike u veličini chain koda. Drugim riječima, razlika između dva vektora koji predstavljaju istu radnju na različitim udaljenostima od kamere će biti malena, dok će razlika između 2 vektora koji predstavljaju dvije različite radnje biti daleko veća.



Slika 2.1: Vektorski prikaz videa hoda

Razlike između histograma koji predstavljaju različite radnje su lako vidljive, kao što i vidimo na priloženim slikama 2.1, koja predstavlja video hodanja, te 2.2, koji je vektor videa saginjanja. Te razlike će biti dovoljne za klasificiranje različitih radnji, iako, primjerice, baš dane dvije akcije imaju slične vrijednosti grafova.



Slika 2.2: Vektorski prikaz videa saginjanja

2.4. Klasifikacija

Za konačni proces klasifikacije koristimo normalizirane vektorske reprezentacije silueta dobivene u koraku 2.3. Nad tim vektorima vršimo algoritam grupiranja k-centroida (eng. *k-means clustering*). Kao mjera udaljenosti koristi se kutna udaljenost između vektora, kao što je definirana na 2.6, pri čemu je θ kut između vektora.

$$\Omega_{\vec{a}, \vec{b}} = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|} = 1 - \cos(\theta) \quad (2.6)$$

Budući da su vektori normalizirani, jednačba 2.6 svodi se na 2.7.

$$\Omega_{\hat{a}, \hat{b}} = 1 - \hat{a} \cdot \hat{b} \quad (2.7)$$

Pri tome se dobiva k centroida, pri čemu je u našem slučaju $k = 8$. Budući da je poznato da je algoritam nerobustan u ovisnosti o početnim centroidima, ovaj se proces pokreće 10 puta, te su zadržava najbolji rezultat. Kao početni centroidi uzete su nasumično odabrane vrijednosti u granicama vrijednosti koje se pojavljaju u danim vektorima.

Nakon ponavljanja procesa, svakoj grupi centroida pridružuje se oznaka. Ona se dobiva uspoređivanjem svakog centroida sa centroidima klasificiranih primjera (odnosno centroidima videa u skupu za učenje). Pri tome više centroida može dobiti istu oznaku.

2.5. Vremensko zaglađivanje

3. Ispitivanje rješenja

3.1. Ispitna baza

U ovom projektu koristili smo Weizman bazu koja se sastoji od 90 snimaka podijeljenih u 10 aktivnosti (9 snimaka za svaku aktivnost). Aktivnosti iz baze koje se mogu prepoznavati našim programom su:

- hodanje,
- trčanje,
- skakanje na mjestu,
- mahanje jednom rukom,
- saginjanje.

Snimke su snimane statičnom kamerom rezolucije 180 x 144 piksela. Na snimkama je prisutan šum u obliku hodanja sa psom ili nošenja torbe. Snimke sadrže homogenu pozadinu radi lakšeg izdvajanja pozadine od siluete.



Slika 3.1: Kadar iz videa – akcija trčanja

Pri klasifikaciji smo 9 ponuđenih snimaka za svaku akciju razdijelili na 2 skupa: skup za treniranje i skup za testiranje. Skup za treniranje sadrži 8 snimaka, a skup za testiranje jednu.

Dva primjera kadrova iz korištenih videa dani su na slikama 3.2, te 3.1.



Slika 3.2: Kadar iz videa – akcija hodanja

3.2. Rezultati učenja i ispitivanja

Svaka grupa snimaka iste akcije podijeljena je na dva dijela – skup za učenje, te skup za testiranje. Optimizacija hiperparametara vršena je validacijom izostavljanja jedne snimke (eng. *leave-one-out validation*). Odnosno, treniranje je ponavljano n puta (gdje $n = 8$, odnosno broj snimaka u skupu za učenje), uz izostavljanje jedne od snimaka pri svakom koraku. Ostatak je korišten za učenje, dok je provjera vršena na zadnjoj od tih snimaka. Nakon ponavljanja, uzet je prosjek greške kao ocjena uspješnosti.

Pri tome su optimirani parametri broj grupa u koje se vrši grupiranje, te parametri predobrade (zaglađivanja) slike.

Konačno ispitivanje vršilo se nad skupom od jedne snimke za svaku akciju. Kao mjera gubitka korišten je 0-1 gubitak (eng. *zero-one loss*). Kako metoda grupiranja (k-means) čak i uz 10 ponavljanja daje razlike u rezultatima, postotak točnosti nije uvijek identičan, već ima malu varijancu ($\approx 1.5\%$).

Konačni iznosi grešaka dani su u tablici 3.1. Ti su rezultati dobiveni uprosječivanjem na 30 izvođenja, a precizni su do na jednu decimalu.

Greška, Učenje (%)		Greška, Ispitivanje (%)	
μ	σ	μ	σ
64.3	1.6	60.1	1.3

Tablica 3.1: Greške (30 izvođenja)

3.3. Analiza rezultata

Analizirati uzroke rezultata ispitivanja, povezati sa uzorcima u bazi i algoritmima korištenim u rješenju. Raspraviti moguća poboljšanja.

4. Opis programske implementacije rješenja

(do 5 stranice)

Opisati sučelje programske implementacije i način korištenja implementacije.

5. Zaključak

(do 2 stranice)

Ocijeniti uspješnost implementacije, navesti budući rad u smislu potrebnih poboljšanja.

6. Literatura

1. Ime i prezime autora: Naziv časopisa vol. br. godina izdanja, pp od-do (npr. pp 486-492)/knjige/članka/web resursa (s linkom i datumom pristupa web resursu) ...
. . DVD/CD . kompletan tekst projekta izvorni kod projekta exe verzija readme file –
upute za korištenje i pokretanje programa . baze slika (sve koje su korištene) E-oblik
članaka koji su korišteni za izradu projekta primjeri obrade ..

Raspoznavanje aktivnosti osoba na temelju siluete

Sažetak

Sažetak na hrvatskom jeziku.

Ključne riječi: Ključne riječi, odvojene zarezima.

Title

Abstract

Abstract.

Keywords: Keywords.