Stefan Nožinić

Poređenje pouzdanosti prepoznavanja osobe metodom prepoznavanja lica, metodom prepoznavanja govora i kombinacijom ovih metoda

Poređena je pouzdanost različitih metoda prepoznavanja osobe: prepoznavanje korišćenjem podataka o licu, prepoznavanje korišćenjem podataka o glasu i kombinacija ove dve metode. Testiranje je vršeno na dva različita skupa podataka. Svaki test sadržao je jednu sliku lica i jedan audio zapis govora osobe. Prvi skup testova je imao zanemarljivo malo šuma u ulaznim podacima. Drugi skup testova imao je značajan šum – velike varijacije osvetljenja kod slika lica i značajan nivo pozadinske buke kod audio zapisa. Eksperimentalni rezultati nad testovima sa zanemarljivo malo šuma pokazuju da se najveća pouzdanost dobija upotrebom kombinacije metoda. U skupu testova sa značajnim šumom, utvrđeno je da je metoda prepoznavanja samo lica pokazala najveću pouzdanost. Prepoznavanje je vršeno u dva dela. Prvi deo je bio utvrđivanje da li je osoba već poznata sistemu ili je treba klasifikovati kao nepoznatu osobu. Ovaj deo je nazvan verifikacijom. Drugi deo se obavlja samo ako je osoba poznata sistemu i u ovom delu se vrši precizno određivanje identiteta osoba, odnosno obavlja se identifikacija. Izmerena je osetljivost verifikacije i preciznost identifikacije u zavisnosti od dominantnosti metode.

Uvod

Tokom vremena razvijeni su mnogi pristupi identifikaciji. Neki od pristupa su prepoznavanje lica (Turk i Pentland 1991), govora (Reynolds i Rose 1995) i prepoznavanje po dinamici kucanja (Fabian i Rubin 2000).

U ovom radu se ispituje pouzdanost verifikacije i identifikacije osobe kombinacijom prepoznavanja lica i govora. Pouzdanost kombinacije metoda se poredi sa upotrebom svake metode posebno.

Za prepoznavanje lica korišćena je metoda koja se oslanja na analizu svojstvenih komponenti (Principial component analysis) (Bishop 2006; Turk i Pentland 1991). Za prepoznavanje glasa korišćeni su *Mel-scale filterbank* (Nealand *et al.* 2002) i težinski Gausov model (Reynolds i Rose 1995).

Određivanje parametara za modele glasa i lica je vršeno kroz jedan skup uzoraka namenjen za obuku, a njihovo testiranje je vršeno pomoću skupa uzoraka za testiranje. Za test-uzorak se računa sličnost poznatom skupu, tako što se uzorak koji se testira poredi sa svakim uzorkom iz skupa za obuku. Poređene su pouzdanost identifikacije i osetljivost verifikacije.

Metode

Prepoznavanje lica osobe. Upotrebljena je metoda za prepoznavanje lica koja se oslanja na analizu svojstvenih komponenti (Principial component analysis) (Bishop 2006; Turk i Pentland 1991).

Stefan Nožinić (1997), Šabac. Arhimandrita Stevana Jovanovića 41/1, učenik 2. razreda Tehničke škole Šabac

MENTOR: Milan Gornik, Fakultet tehničkih nauka Univerziteta u Novom Sadu

Skup za obuku čine jedno-kanalne slike (grascale) koje su istih dimenzija. Na ovim slikama se vrši nadovezivanje vrednosti piksela u niz kako bi se dobili vektori. Potom se računa prosečna vrednost dobijenih vektora iz skupa za obuku. Posle izračunavanja prosečnog vektora, računa se matrica kovarijansi. Za ovu matricu se rešava njen svojstveni problem, odnosno određuju se njene svojstvene vrednosti i svojstveni vektori. Svojstveni vektori sa najvećim svojstvenim vrednostima se čuvaju u bazi.

Nakon ovog postupka svaki vektor iz skupa za obuku projektuje se na podprostor koji čine sačuvani svojstveni vektori iz baze.

Prepoznavanje lica se vrši tako što se neidentifikovano lice pretvara u vektor na način kao što je to rađeno sa slikama za obuku, i oduzima od prosečnog lica koje je izračunato u procesu obuke modela. Novodobijeni vektor se projektuje na podprostor određen svojstvenim vektorima sačuvanim u bazi.

Posle projekcije na manji podprostor, računa se ocena pripadnosti za svako lice iz skupa za obuku na sledeći način:

$$F_i = \log(T - d_i) - \log T$$

gde je F_i ocena pripadnosti neidentifikovanog lica za *i*-to lice iz skupa za obuku, d_i euklidska distanca vektora neidentifikovanog lica i vektora *i*-tog lica iz skupa za obuku i T granica osetljivosti. Ako je distanca između vektora veća od ove granice, logaritam nije definisan. Ovo je važno, jer ako nije moguće izračunati ocenu pripadnosti nijednom licu iz skupa za obuku, osoba se smatra nepoznatom, što znači da nije u bazi poznatih osoba.

Uslovi za dobro funkcionisanje ovog metoda jesu da su sve slike iste veličine i da je na njima samo lice, odnosno da samo lice varira dok pozadina i osvetljenje moraju ostati isti. Zbog toga je pre primene prepoznavaja lica ovom metodom potrebno uraditi detekciju lica na slici i normalizaciju slike.

Prepoznavanje glasa osobe. Slično kao kod prepoznavanja lica, i ovaj model je bilo potrebno prvo obučiti, kako bi mogao da funkcioniše kasnije pri identifikaciji. Skup za obuku su činili uzorci ljudskog govora, pri čemu je svaki uzorak trajao 5 sekundi. Uzorsu su bili podeljeni u manje delove od po 20 ms. Na svakom manjem delu je rađen FFT algoritam koji je dao spektar tog dela govora, a iz tog spektra su izvučene specifične frekvencije. Ovih frekvencija bilo je 26, te je tako za svaki deo uzorka dobijen 26-dimenzionalni vektor koji predstavlja opis glasa u tom trenutku. Na taj način je svaki uzorak činio niz od 26-dimenzionalnih vektora. Na osnovu tog niza potrebno je odrediti parametre modela tako da zajedička verovatnoća pojavljivanja dobijenog niza vektora bude maksimalna:

$$p(X \mid \vec{\mu}, \vec{C}, \vec{m}) = \prod_{k=1}^{K} p(x_k \mid \vec{\mu}, \vec{C}, \vec{m})$$

gde je X niz 26-dimenzionalnih vektora koji predstavljaju opis dela govora, a $p(x_k \mid \vec{\mu}, \vec{C}, \vec{m})$ predstavlja linearnu kombinaciju normalnih raspodela za dati vektor iz liste *X,* odnosno:

$$p(X \mid \vec{\mu}, \vec{C}, \vec{m}) = \sum_{q=1}^{Q} \mu_q N(x_k \mid \vec{\mu}, \vec{C}_q, \vec{m}_q)$$

Q je broj različitih raspodela i preciznost je veća kako se Q povećava, ali isto tako raste i količina memorije potrebne za pamćenje parametara. C_q i m_q su matrica kovarijanse i očekivana vrednost za datu raspodelu.

Kada se ovi parametri izračunaju, oni se čuvaju u bazi i pamte se za svaki uzorak iz skupa za obuku.

Prilikom identifikacije osobe, govor osobe se ponovo deli u delove od po 20 ms i na njima se radi FFT i uzimaju se karakteristične frekvencije kao što je to rađeno prilikom obuke. Sada se i za neidentifikovanu osobu formira lista 26-dimenzionalnih vektora X koja predstavlja karakteristične frekvencije u govoru neidentifikovane osobe tokom vremena.

Za svaku osobu iz baze (skupa za obuku) se onda može izračunati ocena pripadnosti po obrascu:

$$S_i = \log p(X \mid \vec{\mu}_i, \vec{C}_i, \vec{m}_i)$$

gde je S_i ocena pripadnosti *i*-te osobe iz skupa za obuku, $p(X \mid \vec{\mu}_i, \vec{C}_i, \vec{m}_i)$ zajednička verovatnoća za listu X po parametrima iz baze izračunatih za *i*-tu osobu iz skupa za obuku.

Kombinacija metoda. Konačna ocena je formirana kao linearna kombinacija posebnih ocena i konačna ocena neidentifikovane osobe je

rađena za svaku osobu iz skupa za obuku, odnosno:

$$U_i = w_1 F_i + w_2 S_i$$

predstavlja konačnu ocenu pripadnosti neidentifikovane osobe i-toj osobi iz skupa za obuku na osnovu već izračunatih ocena S_i i F_i i koeficijenata w_1 i w_2.

Koeficijenti w_1 i w_2 se mogu modifikovati i time se može podešavati dominantnost određene metode u odnosu na drugu.

Kada se izračuna ocena za sličnost neidentifikovane osobe sa svakom osobom iz skupa za obuku, kao konačni rezultat identifikacije se uzima osoba iz skupa za obuku sa kojom neidentifikovana osoba ima maksimalnu ocenu sličnosti.

Ovim postupkom je moguće vršiti određivanje pripadnosti poznatom skupu osoba (verifikacija) na sličan način kao kod prepoznavanja lica, time što bi se uvela granica prolaznosti. Ovaj postupak ipak nije korišćen za samu verifikaciju kako bi se osigurao manji broj potrebnih parametara za izračunavanje i jednostavnost sistema.

Testiranje. Izvršena je obuka modela sa 180 osoba gde svaka ima sliku lica dobrog osvetljenja i snimljen audio zapis govora od 5 sekundi sa zanemarljivom bukom u pozadini. Kod testiranja su korišćena dva skupa. Jedan skup je bio sačinjen od 120 osoba gde nije postojao šum i gde su slike lica imale zanemarljivo male promene u osvetljenju, a snimak govora je bio sa zanemarljivom bukom u pozadini. Drugi skup za testiranje je bio sačinjen od 120 osoba gde su postojale velike promene u osvetljenju slika lica i značajna buka u snimku govora. Svaki skup od ova dva je bio podeljen na osobe koje su poznate sistemu i koje sistem može da identifikuje i osobe koje bi sistem trebalo da proglasi nepoznatim jer nisu korišćene tokom obuke modela. Lica su korišćena iz FEI baze lica koja je preuzeta sa Interneta. Snimci govora su dobijeni sa snimaka TED konferencija.

Za ispitivanje modela prilikom testiranja su uvedeni sledeći parametri:

1. Prosečan odnos ocena lica i govora: ovaj parametar predstavlja prosečnu vrednost količnika ocene lica i govora dobijenih kao maksimalne ocene tokom testiranja za datu test osobu, odnosno:

$$r = \frac{1}{N} \sum_{i=1}^{N} \frac{w_1 F_i}{w_2 S_i}$$

gde r predstavlja prosečan odnos maksimalnih ocena lica i govora, N broj osoba u skupu za testiranje, F_i ocenu pripadnosti lica najpribližnije osobe iz skupa za obuku i-toj osobi u skupu za testiranje i S_i ocenu pripadnosti glasa najpribližnije osobe iz skupa za obuku i-toj osobi u skupu za testiranje.

2. Odnos granice za verifikaciju i maksimalne dužine lica u skupu za obuku: ovaj parametar predstavlja količnik najveće dozvoljene udaljenosti dva lica i maksimalne dužine vektora u skupu za obuku (projektovanog na manji podprostor kog čine izračunati svojstveni vektori) i definisan je kao:

$$k = \frac{T}{\max_{i=1}^{M} |v_i|}$$

gde je k odnos granice za verifikaciju, M broj osoba u skupu za obuku, T maksimalna dozvoljena distanca između dva vektora lica projektovanih na podprostor određen svojstvenim vektorima, a $\max_{i=1}^{M} |v_i|$ najduža veličina vektora iz skupa za obuku projektovanog na svojstveni podprostor. Ako je k suviše malo, sistem ima malu osetljivost i trebalo bi da odobri veliki procenat osoba kao poznate. Sa druge strane, ako je k suviše veliko, sistem treba da odbije veliki procenat osoba i označi ih kao nepoznate.

3. Preciznost: verovatnoća da je poznata osoba pravilno identifikovana i definiše se kao:

$$p = \frac{I}{P}$$

gde je p preciznost, I broj osoba koje su pravilno identifikovane prilikom testiranja, P broj osoba u skupu za testiranje koje se nalaze i u skupu za obuku.

4. Osetljivost: Predstavlja verovatnoću da se poznate osobe prepoznaju kao poznate (ovde se ne uzima u obzir da su pravilno identifikovane) i nepoznate osobe pravilno proglase nepoznatim:

$$q = \frac{B'C'}{BC}$$

gde q predstavlja osetljivost, B' predstavlja broj osoba klasifikovanih kao nepoznate tokom

obuke, B predstavlja broj svih nepoznatih osoba u skupu za obuku, C' predstavlja brojosoba klasifikovanih kao poznate tokom testiranja, a C predstavlja ukupan broj svih poznatih osoba u skupu za testiranje.

Upotrebom ovih parametara, moguće je izmeriti pouzdanost verifikacije i identifikacije. Verifikacija podrazumeva sposobnost sistema da razlikuje poznate osobe od nepoznatih dok je identifikacija zadužena za konkretno prepoznavanje poznatih osoba odnosno klasifikaciju osobe kao poznate osobe iz skupa za obuku.

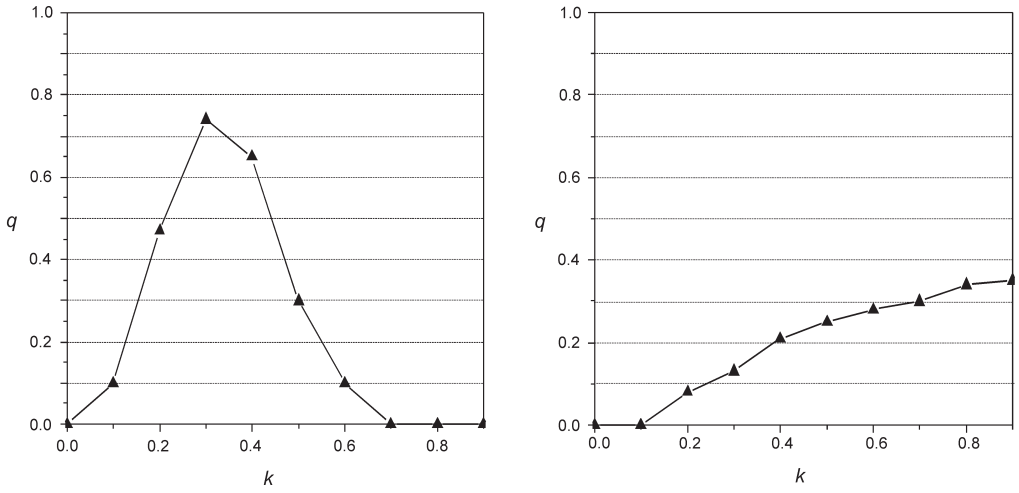
Rezultati i diskusija

Na slici 1 prikazana je zavisnost osetljivosti od parametra k. Na slici 2 se može videti preciznost u odnosu na parametar r za određene vrednosti parametra k (0.3, 0.5 i 0.7).

Iz dobijenih rezultata se može zaključiti da je osetljivost zadovoljavajuća kada je granica za klasifikaciju između poznatog i nepoznatog lica 30% od maksimalne dužine lica u skupu za obuku. Isto tako se parametar r može posmatrati kao parametar koji određuje dominaciju jednog metoda. Na primer, ako je $r = 1$ to znači da su oba metoda jednaka a ako je $r = 10$ to znači da je rezultat prepoznavanja lica 10 puta dominantniji od rezultata prepoznavanja glasa. Primećuje se da za veliku granicu preciznost ima najveću vrednost kod ravnopravne kombinacije metoda. Što je manja granica, to se preciznost smanjuje i postaje manje zavisna od same metode. Isto se tako može primetiti da i kombinacija, ali i pojedinačne metode, dosta gube osetljivost i preciznost kod drugog skupa testova koji je sadržao šum.

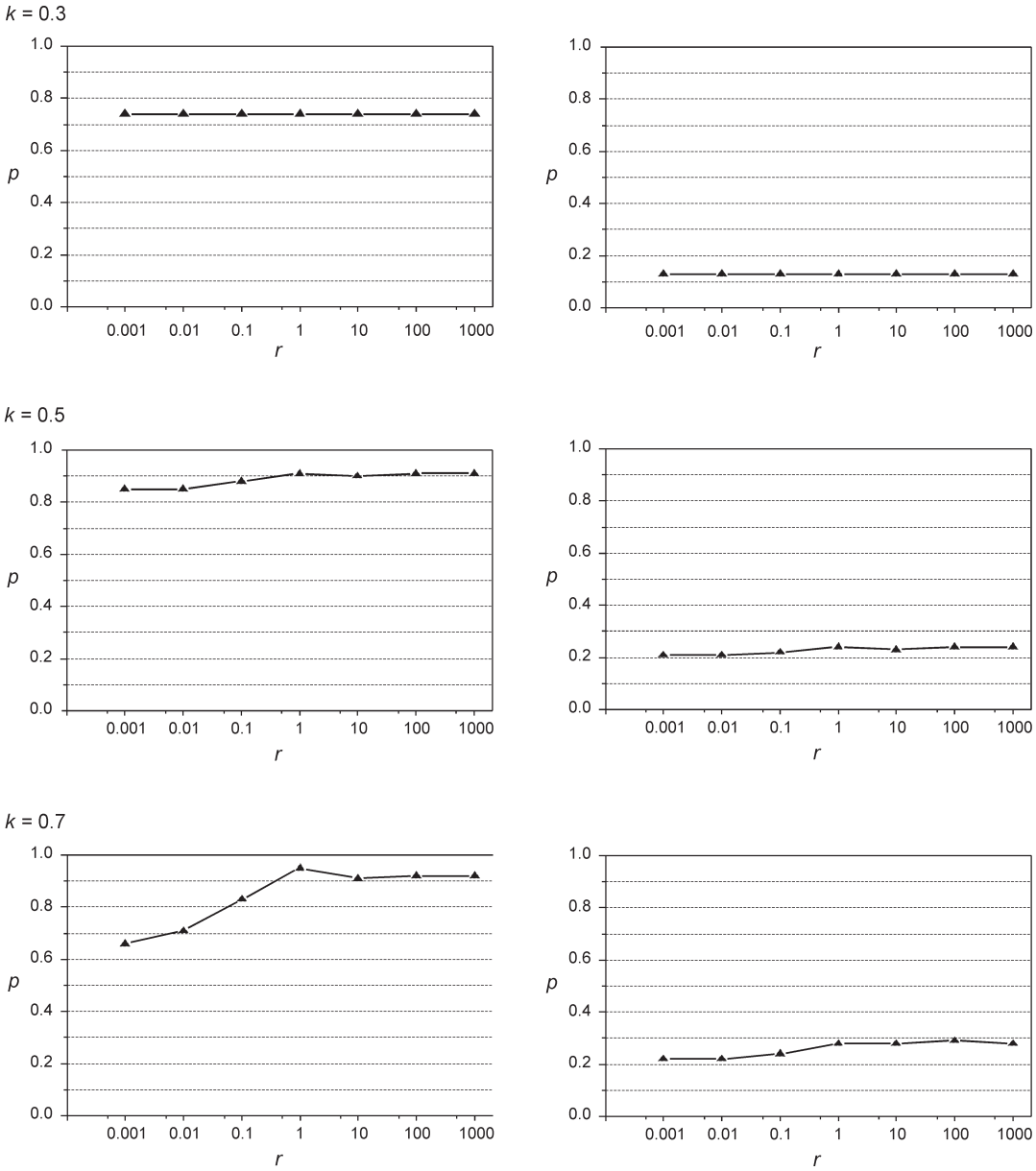
Kada je šum zanemarljivo mali, ravnopravnom kombinacijom se postiže bolja preciznost, nego korišćenjem samo metoda prepoznavanja lica, ili korišćenjem samo metoda prepoznavanja govora. Kada je šum veći, metoda prepoznavanja samo lica pokazuje bolju preciznost nego kombinacija metoda ili samo metoda prepoznavanja glasa. Ovo znači da je metoda prepoznavanja lica manje osetljiva na šum od metode prepoznavanja glasa, čija je osetljivost na šum dominantnija pa time smanjuje pouzdanost kombinacije. Promene granice osetljivosti za verifikaciju osobe menjaju i samu pouzdanost kombinacije, ali i svake metode posebno. Spuštanje granice osetljivosti uzrokuje smanjenje preciznosti kombinacije pa se razlika između preciznosti kombinacije dva metoda i preciznosti svakog metoda posebno smanjuje.

Zaključak



Slika 1. Zavisnost osetljivosti od parametra k za prvi skup testova (levo) i drugi skup testova (desno)

Figure 1. Sensitivity depending on parameter k for the first (left) and the second (right) test sets



Slika 2. Zavisnost preciznosti od parametra r za $k = 0.3$, $k = 0.5$ i $k = 0.7$; levo – prvi skup testova, desno – drugi skup testova.

Figure 2. Precision depending on parameter r for $k = 0.3$, $k = 0.5$ and $k = 0.7$; left – the first test set, right – the second test set.

Prema dobijenim rezultatima, pretpostavka da se kombinacijom prepoznavanja lica i govora može postići bolja pouzdanost verifikacije i identifikacije osobe se pokazala kao tačna samo u slučaju kada je šum manje prisutan i kada ga je moguće zanemariti. U slučaju kada je šum više prisutan, metoda prepoznavanja samo lica pokazuje bolju pouzdanost. Moguća su unapređenja ovog istraživanja kao što su ispitivanje ponašanja kombinacije kada je šum prisutan samo u snimku govora ili samo u slici lica, ili preciznije utvrđivanje pouzdanosti kombinacije u zavisnosti od količine šuma. Na sličan način moguće je izvršiti poređenje kombinacije drugih metoda sa posebnim metodama ili čak poređenje kombinacija različitih metoda.

Literatura

Beigi H. 2011. *Fudamentals Of Speaker Recognition*. Springer

Bishop C. M. 2006. *Pattern Recognition and Machine Learning*. Springer

Fabian M., Rubin A. D. 2000. Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems*, **16**: 351.

Nealand J. H., Bradley A. B., Lech M. 2002. Filterbank Feature Extraction For Gaussian Mixture Model Speaker Recognition. *Proceedings of the 9th Australian International Conference on Speech Science & Technology Melbourne, December 2 to 5, 2002*. Australian Speech Science & Technology Association, str. 415-420.

Reynolds D. A., Rose R. C. 1995. Robust Text-Independent Speaker Identification Using Gaussian Mixture Models. *IEEE Transaction on Speech and Audio Processing*, **3** (1): 72.

Turk M. A., Pentland P. 1991. Face Recognition Using Eigenfaces. *Proceedings CVPR '91.*, str. 586-591.

Stefan Nožinić

Reliability Comparison of Person Recognition Methods: Face Recognition, Speech Recognition and a Combination of These Methods

This paper compares the reliability of different methods for person recognition. The following methods are compared: recognition using a person's facial data, recognition using a person's speech data, and a combination of these two methods. The testing has been done with two distinct sets of data. Every test case contained one face image and one audio recording with a person's speech. The first test set contained a negligible amount of noise in the input data. The second test set contained a significant amount of noise – large variations in face images' brightness and a non-negligible level of background noise in audio data. The experimental results in the first testing set show that a combination of recognition methods outperforms any of the methods alone. In the second testing set with non-negligible noise level, the method which uses only facial data showed the best reliability. Recognition has been done in two stages. The first stage is the decision-making stage where a person is classified as known or unknown. This stage is called verification. The second stage is executed only if a person is classified as known. In this stage, the person is identified as a person from the database of known persons. This stage is called identification. Sensitivity of verification and precision of identification are determined depending on the specific method which is used during the testing phase.