

Primena skrivenog Markovljevog modela u prepoznavanju znakovnog jezika

Problem prepoznavanja znakovnog jezika kojeg koriste gluvoneme osobe predstavlja dugogodišnji izazov. U cilju rešavanja tog problema razvijene su raznovrsne metode bazirane na veštačkoj inteligenciji. Skriven Markovljev model predstavlja jedan od najčešće korišćenih načina modelovanja kontinualnog signala u procesu automatskog prepoznavanja. Ovaj model je posebno pogodan zato što kontinualni signal nije potrebno prethodno segmentirati (izdvojiti značajne celine, npr. reči), već se segmentacija vrši u procesu prepoznavanja. U ovom radu se ispituje uticaj broja stanja skrivenog Markovljevog modela kojim se modeluju reči znakovnog jezika na uspešnost prepoznavanja. Dobijeni rezultati ukazuju da je malo verovatno pronaći zajednički optimalan broj stanja koji važi za sve korišćene reči već da se svaka reč mora modelovati njoj odgovarajućim brojem stanja. Zaključuje se i da postoji granica broj stanja posle koje svaki veći broj stanja modela nema uticaj na uspešnost prepoznavanja.

Uvod

Veštačko prepoznavanje ljudskih gestova igra važnu ulogu u unapređivanju interakcije između mašine i čoveka. Više puta je pristupano problemu segmentacije značajnih delova tela (šake, glave, itd.) i praćenja položaja istih (Moeslund *et al.* 2006). Razvijena je tehnologija koja ovakvo praćenje vrši sa relativno velikim nivoom uspešnosti (Shotton *et al.* 2011). Na osnovu algoritma opisanog u (Shotton *et al.* 2011), od strane Microsoft korporacije proizveden je senzor „Kinect”. Ovaj uređaj predstavlja značajan tehnološki napredak koji, premda napravljen da služi kao

kontroler u igrama, omogućava istraživačima razvoj kvalitetnijih sistema za prepoznavanje gestova.

Jedan od problema koji spada pod prepoznavanje gestova je prepoznavanje znakovnog jezika koji koriste gluvoneme osobe. U znakovnom jeziku se informacija prenosi putem gestova ruku i stanja šake (orijentacija šake i položaj prstiju). Dva su načina prenošenja informacije:

- gestom ruku i stanjem šake se daje znak koji nosi informaciju o jednoj reči (slika 1, levo)
- stanjem šake se daje znak koji nosi informaciju o jednom slovu iz znakovne azbuke (slika 1, desno)

Za prepoznavanja slova znakovne azbuke su razvijeni funkcionalni sistemi koji snimanjem položaja šake i prstiju realizuju prepoznavanje putem različitih metoda (Ricco i Tomasi 2009) gde je uspešnost prepoznavanja dostizala 99.7% (Birk *et al.* 1997).

Sveobuhvatan pregled problema prepoznavanja znakovnog jezika izneo je Starner (Starner 1995). Govornim jezikom se prenosi isti tip informacije kao i znakovnim i usled toga je pretpostavljeno da će slične metode prepoznavanja, primenjene na oba tipa signala, dati slične rezultate. Skriveni Markovljev model (eng. Hidden Markov Model, dalje SMM) je našao uspešnu primenu u prepoznavanju govornog signala (Rabiner 1989; Juang i Rabiner 1991). Starner pokazuje da je moguće uspešno primeniti SMM na prepoznavanje znakovnog jezika, potvrđujući navedenu pretpostavku.

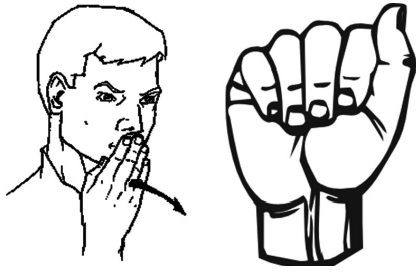
U ovom radu je istraživana zavisnost tačnosti prepoznavanja od parametara SMM-a. Poseban akcenat je stavljen na ispitivanje karakterističnog broja stanja koji odgovara svakoj reči. Istraživana je mogućnost postojanja zajedničkog optimalnog broja stanja koji se uspešno mogu modelovati sve reči zna-

Miloš Stojanović (1993), Beograd, Bulevar Arsenjia Čarnojevića 183/13, učenik 3. razreda Devete gimnazije „M. P. Alas” u Beogradu

MENTORI:

Ana Anastasijević, student Elektrotehničkog fakulteta Univerziteta u Beogradu

Đorđe Nijemčević, Microsoft Development Center, Beograd



Slika 1. Znak za reč „hvala” (levo) i znak za slovo „A” (desno)

Figure 1. Sign for the words “thank you” (left), and sign for letter “A” (right)

kovnog jezika. Cilj ovog rada je ispitivanje uticaja broja stanja SMM-a kojim se modeluju gestovi znakovnog jezika na uspešnost u veštačkom prepoznavanju istih.

Skriveni Markovljev model

Problem prepoznavanja znakovnog jezika

Rečenica na znakovnom jeziku, kao i u govoru, predstavlja kontinualni signal koji se sastoji od više reči koje poštujući gramatiku prenose određenu informaciju. Signal znakovnog jezika predstavlja prostornu promenu položaja ruku tokom vremena. Za praćenje ovakve promene potreban je sistem koji uspešno razdvaja značajne delove tela i određuje njihov položaj u prostoru. Navedene mogućnosti sadrži Microsoft-ov senzor Kinect koji je korišćen u ovom istraživanju.

Teorija Markovljevih modela

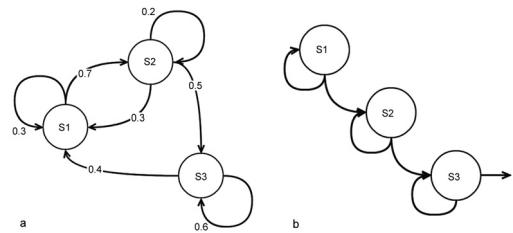
Markovljeva osobina. Dat je sistem čije se stanje u proizvoljnom trenutku može modelovati elementima iz konačnog skupa stanja. Neka je taj sistem do diskretnog vremenskog trenutka t prošao kroz n stanja. Ako kod takvog sistema važi da buduće stanje u trenutku $t + 1$: S_{n+1} zavisi isključivo od trenutnog stanja S_n i ni od jednog drugog prethodnog stanja, onda se kaže da taj sistem ima Markovljevu osobinu.

Markovljev proces je nasumičan proces kod koga se sistem u proizvoljnom vremenskom trenutku može nalaziti u tačno jednom od n mogućih stanja i gde tokom vremena sa određenom verovatnoćom može doći do prelaza u neko drugo stanje. Slika 2a

grafički predstavlja primer Markovljevog procesa koji se sastoji od tri stanja i sedam mogućih prelaza između njih. Ivice grafa označavaju verovatnoće prelaza između odgovarajućih stanja. Te se verovatnoće mogu predstaviti i u matricnom obliku. Za navedeni primer matricni oblik je:

$$a_{ij} = \begin{bmatrix} 0.3 & 0.7 & 0.0 \\ 0.3 & 0.2 & 0.5 \\ 0.4 & 0.0 & 0.6 \end{bmatrix}$$

Poseban slučaj Markovljevog procesa je Markovljev lanac prvog reda (slika 2b) gde sistem iz stanja S_i može preći isključivo u stanje S_{i+1} ili ostati u istom stanju S_i .



Slika 2. a) Markovljev proces sa prikazanim verovatnoćama tranzicija; b) Markovljev lanac prvog reda

Figure 2. a) Markov proces with transition probabilities; b) First order Markov chain

Skriveni Markovljev model. Ponekad stanja modela nisu direktno merljiva, već su merljivi samo izlazi sistema, koji su u određenoj vezi sa njegovim unutrašnjim stanjem. Navedene dve grupe signala (unutrašnja stanja i izlazna stanja) nije uvek moguće povezati jednoznačnim preslikavanjem. Kod skrivenog Markovljevog modela svakom stanju S_i se dodeljuje funkcija gustine verovatnoće $b_i(O_t)$ koja određuje verovatnoću da stanje S_i generiše izlaz (opservaciju) O_t u trenutku t . Model se naziva skriveni jer je nepoznato koje stanje odgovara kojoj opservaciji (Melnikoff *et al.* 2001).

Problem implementacije skrivenog Markovljevog modela

Data je sekvenca opservacija O koja se sastoji od T uzoraka signala snimljenih u određenim fiksnim vremenskim intervalima, tj. $O = O_0, O_1, O_2, \dots, O_{T-1}$.

Za svaku reč koja se modeluje moguće je dizajnirati poseban skriven Markovljev model M koji se sastoji od određenog broja stanja. Cilj prepoznavanja je pronaći model M koji najbolje odgovara datoj sekvenci opservacija O , tj. pronaći model koji ima najveću verovatnoću generisanja sekvence O .

Pre samog prepoznavanja nepoznatog signala potrebno je svaki model zasebno trenirati. To je proces prilagođavanja parametara modela tako da se sa što većim uspehom prepoznaju sekvence iz datog, unapred određenog i poznatog skupa opservacija (Rabiner 1989).

Pri implementaciji SMM u prepoznavanju gestova (a i u prepoznavanju bilo kakvog kontinualnog signala) od posebnog je značaja broj stanja SMM-a koji se koristi za prepoznavanje. Pogrešno pretpostavljen optimalan broj stanja SMM-a može dovesti do značajnog pada uspešnosti prepoznavanja (Rabiner 1989).

Metod

Vektor karakteristika

Uočeno je da se pri davanju znaka informacija prenosi putem sledećih parametara:

- međusobnog položaja ruku,
- pomeraja ruku,
- oblasti davanja znaka (posmatra se u smislu položaja ruku u odnosu na glavu),
- stanja šake.

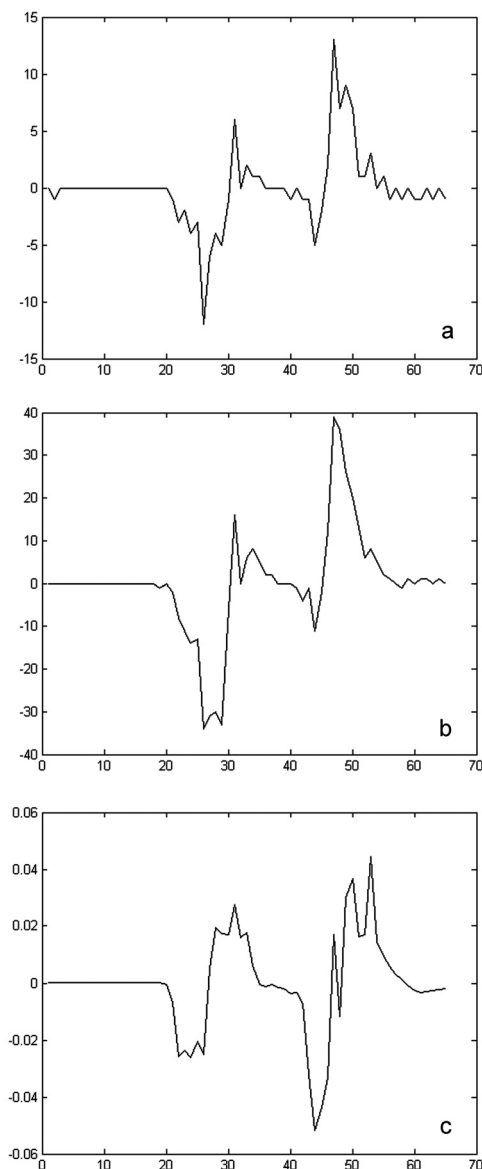
Neophodno je dizajnirati vektor karakteristika – način predstavljanja izmerenih vrednosti od interesa (Starner 1995) dobija zadovoljavajuće rezultate prepoznavanja i ako ne koristi parametar stanja šake, odakle se može zaključiti da su od najvećeg značaja za prepoznavanje prva tri parametra.

Podaci koji opisuju određen gest se prikupljaju putem Kinect senzora (Campbell *et al.* 1996) koji daje informaciju o položaju značajnih delova tela. Od interesa za ovo istraživanje su podaci koji predstavljaju položaje sledećih delova tela:

- zglobovi leve i desne šake,
- levi i desni lakat,
- glava.

Za svaki od navedenih delova tela merene su sledeće vrednosti:

- x i y koordinate piksela slike koje odgovaraju položaju datog dela tela,



Grafik 3. Sirov snimljeni signal pomeraja desne šake jednog od ispitanika za reč „ručak”; a) promena x koordinate, b) promena y koordinate, c) promena z koordinate kroz diskretne vremenske intervale. Na x osi je prikazan broj frejma snimka, a na y osi promene odgovarajuće koordinate.

Figure 3. Raw hand movement signal recorded from one subject for the word “lunch”; a) x -coordinate delta b) y -coordinate delta c) z -coordinate delta through discrete time intervals. Frame number is plotted on x -axis and delta of particular coordinate is plotted on y -axis.

- podatak o udaljenosti datog dela tela od Kinect senzora (z koordinata).

Ovi podaci predstavljaju osnovni set sirovih podataka od kojih se konstruišu vektori karakteristika kojima se vrši treniranje i testiranje. Slika 3 prikazuje jedan od korišćenih signala (za reč „ručak“). Signal predstavlja pomeraje po određenim koordinatima kroz diskretne vremenske intervale.

Korišćen je vektor karakteristika gde se za svaki praćeni deo tela koriste promene njegove koordinate u Dekartovom koordinatnom sistemu. Koordinatni početak ovog sistema nalazi se na mestu senzora. Orijetisan je tako da z osa bude usmerena od senzora ka ispitaniku, a pozitivne strane x i y ose polaze iz gornjeg levog ugla slike snimljene kamerom senzora (dx, dy, dz) (Ricco i Tomasi 2009). U vektor nije uvršten položaj glave.

U ovom radu je korišćen vektor karakteristika koji sadrže podatke o položaju potrebnih zglobova, oblika:

$$[(dx, dy, dz)_{desne\ šake}, (dx, dy, dz)_{leve\ šake}, (dx, dy, dz)_{desnog\ lakta}, (dx, dy, dz)_{levog\ lakta}]$$

Svi podaci su normalizovani u odnosu na rastojanje između glave i zgloba desne šake, kako bi se kompenzovao uticaj telesnih dimenzija ispitanika na apsolutne koordinate pokreta.

Primena SMM-a

SMM je implementiran putem Hidden Markov Model Toolkit-a (HTK). Ovaj alat sadrži implementacije algoritama za treniranje i testiranje SMM-a.

Podatke koje koristi HTK pri treniranju potrebno je snimiti u definisan format koji predstavlja vektor karakteristika. Za svaku reč iz rečnika je definisan poseban SMM sa određenim brojem stanja korišćenjem jezika za definiciju SMM-a koji koristi HTK. Na osnovu seta podataka za treniranje je izvršena inicijalizacija (početno podešavanje parametara), a potom prilagođavanje parametara svih definisanih modela. Modeli su zatim na osnovu predefinisane gramatike spojeni u mrežu korišćenu za prepoznavanje. Prepoznavanja se vrši tako što se za svaki model, koji na osnovu gramatike predstavlja kandidata za reč, testira data opservacija, a kao rezultat prepoznavanja se dobija reč opisana modelom koji najbolje odgovara opservaciji.

Pronalaženje modela koji najbolje odgovara opservaciji postiže se time što se za svaki model pronalazi putanja kroz stanja modela tako da ona generišu

datu opservaciju. Potom se Viterbijevim algoritmom (Rabiner 1989) računa optimalna putanja koja odgovara toj opservaciji. Za svaki frejm (izvršeno merenje u diskretnom vremenskom trenutku) koji je opisan odgovarajućim vektorom karakteristika vrši se izračunavanje verovatnoće prelaska u neko drugo stanje SMM-a.

Variranje parametara

Posmatrano je kako rezultat prepoznavanja zavisi od promene broja stanja skriveneog Markovljevog modela. Obrađene su dve situacije promene broja stanja:

1. Sve reči su modelovane jednakim brojem stanja i tokom testiranja se za svaku reč menja broj stanja za istu vrednost.
2. Pojedinačne reči se modeluju određenim brojem stanja koji se varira samo za jednu reč koja se u datom trenutku ispituje. Posmatra se optimalan broj stanja SMM-a koji opisuje datu reč nezavisno od celokupnog rečnika.

Rezultati

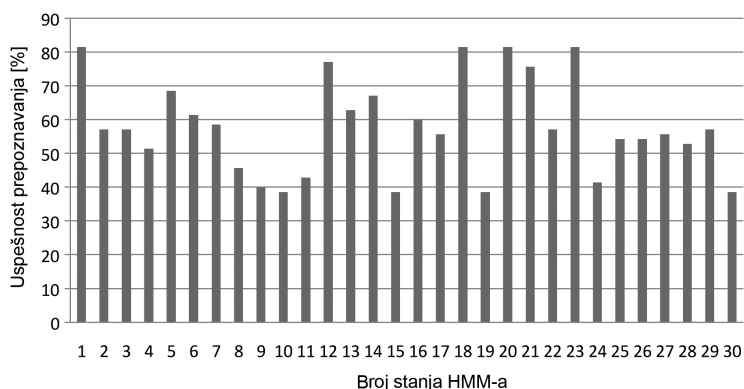
Rečnik se sastoji od 10 različitih pojedinačnih reči (tabela 1). Za prikupljanje podataka je korišćena grupa od 20 subjekata gde je svaki subjekat svaku reč pokazivao po dva puta, što čini set od 40 snimljenih primera za svaku reč.

Tabela 1. Korišćene reči

Dobro	Ne
Država	U
Diskoteka	Razumem
Gladan	Ručak
Morati	Ti

Usled ljudskog faktora je dolazilo do greške u merenju pa je broj korisnih primera putem naknadne manualne analize signala sveden na 34 koliko ih je i korišćeno pri treniranju i testiranju modela.

Za trening modela je korišćeno 24 od 34 primera što je približno 70% ukupnog broja primera, dok je za testiranje korišćeno 10 od 34 primera tj. približno 30% ukupnog broja primera. Broj aktivnih stanja modela je variran na opsegu od 1 do 30. Model svake reči je u datoj iteraciji imao podjednak broj stanja, a



Slika 4. Uticaj broja stanja SMM-a na uspešnost prepoznavanja celokupnog rečnika

Figure 4. HMM state number impact on success rate of whole vocabulary recognition

uspešnost prepoznavanja je računata u odnosu na ceo rečnik. Dobijeni rezultati su prikazani na slici 4.

Uočeni su maksimumi za brojeve aktivnih stanja: 1, 18, 20 i 23 od 81% uspešnosti.

Tabela 2. Optimalan broj stanja SMM-a i uspešnost prepoznavanja pojedinačnih reči.

Reč	Optimalan broj stanja SMM-a	Uspešnost prepoznavanja (%)
Dobro	13	84
Država	8, 9, 11	77
Diskoteka	3, 4, 5	84
Gladan	1	81
Morati	9, 11, 12	78
Ne	1, 2	81
U	10	82
Razumem	1, 2, 3	95
Ručak	3	87
Ti	1	90

Testiran je i optimalan broj stanja SMM-a za svaku pojedinačnu reč (tabela 2). Za testiranu reč varira se broj aktivnih stanja SMM-a koji je opisuje, dok je za ostale reči iz sistema taj broj konstantan i iznosi tri.

Diskusija

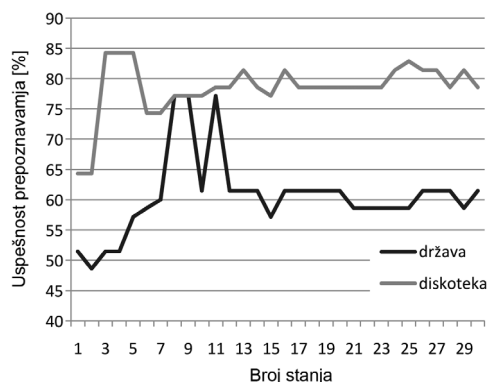
Ukoliko bi postojao optimalan broj stanja modela koji je isti za sve modelovane reči slika 4 bi bio oblika Gausove raspodele sa lako uočljivim jednim maksimumom. Ovako ponašanje modela se moglo

očekivati usled toga što modeli opisuju signal istog tipa – gest znakovnog jezika.

Slika 4 pokazuje da ne postoji optimalan broj stanja modela koji bi bio isti za sve modelovane reči. Pretpostavlja se da za svaku pojedinačnu reč postoji njoj odgovarajući optimalan broj stanja SMM-a kojim se ona najbolje opisuje.

Izdvajanje određenih maksimuma na slici 4 se može tumačiti time što dat broj stanja modela za te maksimume približno odgovara optimalnom broju stanja većeg dela reči iz rečnika i usled toga je dobijena veća uspešnost prepoznavanja za taj broj stanja.

Merenjem broja optimalnih stanja za svaku pojedinačnu reč je potvrđena pretpostavka da postoji zaseban, za svaku pojedinačnu reč određen, broj stanja SMM-a koji daje najbolju uspešnost pri prepozna-



Slika 5. Zavisnost uspešnosti prepoznavanja od broja stanja za reči „država” i „diskoteka”

Figure 5. Success rate dependency of state number for words “country” and “disco”

vanju date reči. Uočeno je da se kod određenih reči javlja više različitih brojeva stanja SMM-a koji daju maksimalnu uspešnost. Usled ovoga se može govoriti o optimalnim opsezima broja stanja (nasuprot optimalnom broju stanja) u kojima je uspešnost prepoznavanja maksimalna, kao što je to npr. opseg 3-5 kod reči „država“ (slika 5). Moguće je da se pri testiranju na većem broju primera iz optimalnog opsega izdvoji optimalan broj stanja.

Posmatranjem grafika zavisnosti uspešnosti prepoznavanja od broja stanja za pojedinačne reči (slika 5) uvida se da broj stanja ima značajan uticaj na uspešnost prepoznavanja samo pri manjem broju stanja. Za veći broj stanja (od 15 i više) uspešnost prepoznavanja postaje konstantna i uvek manja od maksimalne za datu reč.

Zaključak

Na osnovu navedenog ispitivanja uticaja broja stanja SMM-a kojim se modeluju gestovi znakovnog jezika na uspešnost u veštačkom prepoznavanju istih zaključuje se da je za svaku reč potrebno koristiti zaseban broj stanja čime bi se ta, pojedinačna, reč najbolje opisala.

Optimalan broj stanja postoji u određenom opsegu karakterističnom za svaku reč ali se uglavnom nalazi na nižem broju stanja modela (od 15 i manje). Broj stanja prestaje da ima uticaj na uspešnost prepoznavanja kada isti ima veliku vrednost, u ovom istraživanju od 20 i više.

Skriveni Markovljev model primenjen na prepoznavanje znakovnog jezika posle određenog, visokog broja stanja, daje konstantnu uspešnost prepoznavanja za svaki veći broj stanja i ta uspešnost je uvek manja od maksimalne postignute na ospegu nižeg broja stanja.

U daljim radovima potrebno je istražiti od kojih karakteristika signala zavisi optimalan broj stanja modela kojim se on opisuje. Na osnovu izgleda signala se može pretpostaviti da je broj prelaza iz negativnog u pozitivan pomeraj i obrnuto od značajnog uticaja na optimalan broj stanja SMM-a kojim se opisuje dati signal. Takođe najverovatnije postoji i uticaj dužina vremenskog intervala u kome signal predstavlja pozitivan ili negativan pomeraj po odgovarajućoj koordinati.

Zahvalnost. Autor se zahvaljuje mentorima Ani Anastasijević i Đorđu Nijemčeviću koji su značajno

doprineli u stvaranju konačne forme rada. Posebnu zahvalnost autor duguje Ivanu Razumeniću, Darku Todoroviću i Aleksandri Trivunac kao i ostalim saradnicima seminara primenjene fizike i elektronike bez čijih sugestija realizacija ovog projekta ne bi bila moguća.

Literatura

Birk H., Moeslund T., Madsen C. 1997. Real-Time Recognition of Hand Alphabet Gestures Using Principal Component Analysis. U *10th Scandinavian Conference on Image Analysis*, str. 261–268.

Campbell L.W. Becker D. A., Azarbayejani A., Bobick A. F., Pentland A. 1996. Invariant features for 3-D gesture recognition. *M. I. T Media Laboratory Perceptual Computing Section Technical Report*, 379: 1.

Drake A.W. 1988. *Fundamental of applied probability theory*. McGraw-Hill

Dreuw P., Keysers D., Deselaers T., Ney H. 2006. Gesture Recognition Using Image Comparison Methods. U *Gesture Workshop 2005*, LNAI 3881 (ur. S. Gibet, N. Courty, i J.-F. Kamp). Springer, str. 124–128.

Juang B. H., Rabiner L. R. 1991. Hidden Markov Models for Speech Recognition. *Technometrics*, 33 (3): 251.

Makhoul J., Starner T.E., Schwartz R., Chou G. 1994. On-line Curisive Handwriting Recognition Using Hidden Markov Models and Statistical Grammars. U *Human Language Technology* (Proceedings of a Workshop held at Plainsboro, New Jersey, USA, March 8-11, 1994). Morgan Kaufmann, str. 432.

Melnikoff S. J., Quigley S. F., Russell M. J. 2001. Implementing a Hidden Markov Model Speech Recognition System in Programmable Logic. *Lecture Notes in Computer Science*, 2147: 81.

Messing L. S., Erenshteyn R., Foulds R., Galuska S., Stern G. American Sign Language Computer Recognition: Its Present and its Promise. U *Actes de ISAAC'94*. Maastricht, str. 289-291.

Moeslund T. B., Hilton A., Kruger V. 2006. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, **104**: 90.

Rabiner L. R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, **77**: 257.

Ricco S., Tomasi C. 2009. Fingerspelling recognition through classification of letter-to-letter transitions. Dostupno na: http://rd.springer.com/chapter/10.1007/978-3-642-12297-2_21

Shotton J., Fitzgibbon A., Cook M., Sharp T., Finocchio M., Moore R. et al. 2011. Real-time human pose recognition in parts from a single depth image. U *CVPR, IEEE*, June 2011

Starner T. E. 1995. *Visual Recognition of American Sign Language Using Hidden Markov Models*. Massachusetts Institute of Tehnology

Wang R. Y., Popović J. 2009. Real-time hand-tracking with a color glove. *ACM Transactions on Graphics*, **28** (3): 63:1-63:8.

Young S., Evermann G., Gales M., Hain T., Kershaw D., Liu X. (A.) et al. 2009. *The HTK Book*. Cambridge University

Hidden Markov Model Toolkit (HTK), Machine Intelligence Laboratory, Cambridge University Engineering Department. Dostupno na: <http://htk.eng.cam.ac.uk/>

Miloš Stojanović

Applying the Hidden Markov Model in Sign Language Recognition

The problem in sign language recognition is a long-standing scientific challenge. In order to solve this problem, various methods have been developed based on artificial intelligence.

Hidden Markov model is one of the most frequently used methods for modeling a continuous signal in the process of automatic recognition. This model is particularly suitable because prior segmentation (extraction of meaningful units, i.e. words) of the continuous signal is not required, since the segmentation is performed in the process of recognition.

This paper examines the impact of the number of states in Hidden Markov Model, used to model words of sign language, on the effectiveness of recognition. The results indicate that it is unlikely to find the general optimal number of states that apply to all used words, but rather that each word has its own corresponding optimal number of model states. It is concluded that there is a plateau in the number of states and that increasing this number after the plateau is reached has no impact on the efficiency of recognition. 