Prepoznavanje znakovnog jezika pomoću tenzora

Petra Sočo, Jelena Zaninović

8. ožujka 2021.

Prepoznavanje gesti rukama

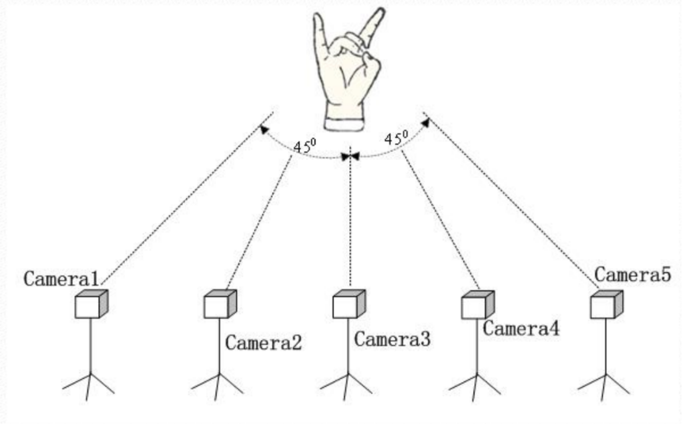
- Računalno prepoznavanje gesti rukama može zamijeniti direktni kontakt s ekranima u javnosti ili u prostorima koji se pokušavaju držati sterilnima (npr. operacijske sale)
- To bi omogućilo sigurniju, ali i intuitivniju interakciju ljudi s računalima (smart uređaji, AR, VR).

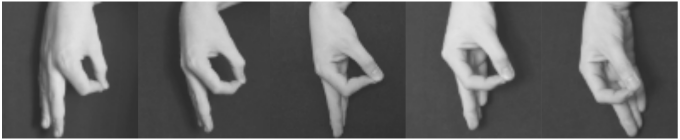
- Jedna od zanimljivijih i korisnijih primjena je kod prepoznavanja znakovnog jezika; naime, računala bi mogla automatski transkribirati znakove u tekst ili govor, što bi moglo olakšati i ubrzati komunikaciju s osobama koje ga koriste.

- Kako bi program što bolje funkcionirao u praksi, moramo razmisliti na koji će način primati informacije - u našem slučaju slike ruku. Osobi nije praktično gestikulirati iz jedne fiksne poze pa ima smisla prilagoditi program da prepoznaje znakove iz različitih kuteva.
- Kad bi računalo (kamera) radilo izvan ambijenta s konzistentnim, umjetnim osvjetljenjem, bilo bi poželjno da mu točnost ne ovisi o dobu dana.

Fotografije

- Fotografije su slikane iz 5 različitih kuteva - od 45° slijeva do 45° zdesna. Fotoaparat je bio u razini prsa osobe i paralelan s podom.





Slika: Slovo P iz 5 različitih kuteva

- Od 26 znakova uspješno samo slikale 24 (nedostaju nam slova 'M' i 'N') te smo u konačnici dobile 120 slika.
- Rezolucija svake slike je 25×25 piksela.

Prepoznavanje ruke na slici

- Piksele dijelimo na 2 klastera: "boja kože" i "ostalo"
- YCbCr model boja osmišljen je tako da Y prati svjetlinu slike, a Cb i Cr informacije o bojama (konkretno, plavoj i crvenoj)
- Sliku iz češće korištenog RGB-a konvertiramo u YCbCr preko sljedeće formule

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

- Iz RGB slika, zbog jasnije teksture, dobile smo greyscale slike, a iz YCbCr slika, gdje je razlika između ruke i podloge značajnija, dobile smo binary slike.
- U ovom smo koraku koristile Octave Forge 'Image' paket: konkretno, funkcije **rgb2gray** i **im2bw**.
- Napomena: **im2bw** koristi određeni "threshold" nad kojim nemamo kontrlou, što može utjecati na konačnu sliku.

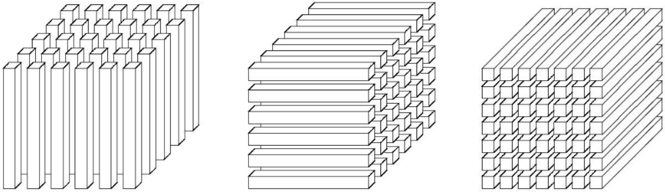
Slika: Slovo A u različitim modelima boja

Definicija

Ako je polje $\mathcal{A}$ određeno s N indeksa, kažemo da je $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ tenzor reda N *nad* $\mathbb{R}$, $\mathcal{A} = (x_{i_1 i_2 \ldots i_{n-1} i_n i_{n+1} \ldots i_N})$, $i_1 = 1, \ldots I_1; \ldots; \; i_N = 1, \ldots I_N$.

Definicija

Neka je $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *tenzor.* I_n*-dimenzionalni vektor dobiven t.d. fiksiramo svaki indeks osim indeksa* i_n *zovemo* nit u modu n*.*



Definicija

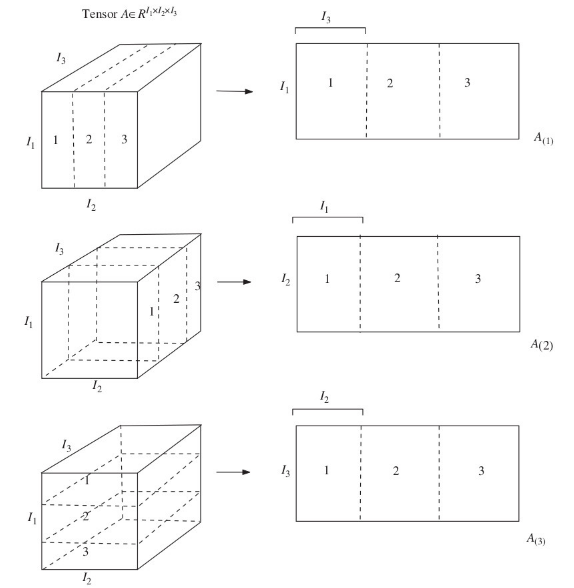
Neka je $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *i neka je* $\mathbf{U} \in \mathbb{R}^{J_n \times I_n}$. Produkt u modu n *tenzora* $\mathcal{A}$ *i matrice* $\mathbf{U}$, s oznakom $\mathcal{A} \times_n \mathbf{U}$, je $(I_1 \times I_2 \times \dots I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N)$-dimenzionalni tenzor zadan s

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 i_2 \dots i_{n-1} j_n i_{n+1} \dots i_N} := \sum_{i_n} a_{i_1 i_2 \dots i_{n-1} i_n i_{n+1} \dots i_N} u_{j_n i_n}.$$

- Produkt u modu n možemo izraziti i na sljedeći način

$$(\mathcal{A} \times_n \mathbf{U})_{(n)} = \mathbf{U} \cdot \mathbf{A}_{(n)},$$

gdje smo s $\mathbf{A}_{(n)}$ označili **matricizaciju** od $\mathcal{A}$ **u n-tom modu**


Tensor $A \in R^{I_1 \times I_2 \times I_3}$
I_3
I_1
I_2
1
2
3
$A_{(1)}$
$A_{(2)}$
$A_{(3)}$

Teorem

Neka je $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. *Za* $\mathrm{A} \in \mathbb{R}^{I_m \times J_m}$, $\mathrm{B} \in \mathbb{R}^{I_n \times J_n}$, $m \neq n$ *vrijedi*

$$\mathcal{A} \times_m \mathrm{A} \times_n \mathrm{B} = (\mathcal{A} \times_m \mathrm{A}) \times_n \mathrm{B} = (\mathcal{A} \times_n \mathrm{B}) \times_m \mathrm{A}.$$

Teorem

Neka je $\mathcal{A} \in \mathbb{R}^{l \times m \times n}$. *Tenzor* $\mathcal{A}$ *možemo zapisati kao*

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}$$

gdje su $\mathrm{U}_1 \in \mathbb{R}^{l \times l}, \mathrm{U}_2 \in \mathbb{R}^{m \times m}, \mathrm{U}_3 \in \mathbb{R}^{n \times n}$ *ortonormalne matrice, a* jezgreni tenzor $\mathcal{S}$ *je istih dimenzija kao* $\mathcal{A}$ *te ima svojstvo potpune ortogonalnosti. Vrijedi i da je*

$$\mathcal{S} = \mathcal{A} \times_1 (\mathbf{U}^{(1)})^T \times_2 (\mathbf{U}^{(2)})^T \times_3 (\mathbf{U}^{(3)})^T.$$

Definicija

Neka je $\mathrm{A} \in \mathbb{R}^{m\times n}$*. Kažemo da je matrica* $\mathrm{A}^+ \in \mathbb{R}^{n\times m}$ Moore-Penroseov pseudoinverz *matrice* A *ako zadovoljava sljedeće uvjete:*

- $\mathrm{A}\mathrm{A}^+\mathrm{A} = \mathrm{A}$
- $\mathrm{A}^+\mathrm{A}\mathrm{A}^+ = \mathrm{A}^+$
- $(\mathrm{A}\mathrm{A}^+)^* = \mathrm{A}\mathrm{A}^+$
- $(\mathrm{A}^+\mathrm{A})^* = \mathrm{A}^+\mathrm{A}$

Reprezentacija podataka tenzorom

- I_l, I_v i I_{pix} ... broj slova, perspektiva i piksela.
- Ulazni podaci ... tenzor oblika $\mathcal{A} \in \mathbb{R}^{I_l \times I_v \times I_{pix}}$.
- Informacije o slovima i perspektivama faktoriziramo pomoću HOSVD-a

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}_l \times_2 \mathbf{U}_v \times_3 \mathbf{U}_{pix}, \tag{1}$$

  gdje su $\mathbf{U}_l, \mathbf{U}_v$ i $\mathbf{U}_{pix}$ ortogonalne matrice reda I_l, I_v i I_{pix}, a $\mathcal{S} \in \mathbb{R}^{I_l \times I_v \times I_{pix}}$ **jezgreni tenzor**.
- Za sliku slova i u pogledu j $(i = 1, \ldots, I_l,\ j = 1, \ldots, I_v)$ imamo

$$\mathcal{A}^{(i,j)} = \mathcal{S} \times_1 \mathbf{u}_i \times_2 \mathbf{u}_j \times_3 \mathbf{U}_{pix}$$

  gdje su $\mathbf{u}_i$ i $\mathbf{u}_j$ redom retci matrica $\mathbf{U}_l$ i $\mathbf{U}_v$, $\mathcal{A}^{(i,j)} \in \mathbb{R}^{1 \times 1 \times I_{pix}}$.

Prepoznavanje znakovnog jezika tenzorom

- Naći najbolju aproksimaciju među već dostupnim podacima
- Testna slika $\mathcal{A}_{test} \in \mathbb{R}^{1\times1\times I_{pix}}$ ima rastav

$$\mathcal{A}_{test} = \mathcal{S} \times_1 \mathbf{u}_i \times_2 \mathbf{u}_j \times_3 \mathbf{U}_{pix}$$

  za neke nepoznate $\mathbf{u}_i$ i $\mathbf{u}_j$.
- Problem glasi

$$\underset{\mathbf{u}_i,\mathbf{u}_j}{\arg\min} \|\mathcal{A}_{test} - \mathcal{S} \times_1 \mathbf{u}_i \times_2 \mathbf{u}_j \times_3 \mathbf{U}_{pix}\|_2.$$

- Rezultat: aproksimacije vektora $\mathbf{u}_i$ i $\mathbf{u}_j$
- Računamo tako da $\mathbf{u}_j$ budu generirani iz konačnog skupa.

Konkretno, neka je $\mathbf{u}_j \in \{\mathbf{U}_v(k, 1 : I_v) : k = 1, \ldots, I_v\}$. Problem sada glasi:

$$\underset{\mathbf{u}_i}{\arg\min} \, \|\mathcal{A}_{test} - \mathcal{S} \times_1 \mathbf{u}_i \times_2 \mathbf{u}_j \times_3 \mathbf{U}_{pix}\|_2 \tag{2}$$

Korištenjem svojstava množenja tenzora i matrice u modu 1:

$$\underset{\mathbf{u}_i}{\arg\min} \, \|\mathcal{A}_{test} - \mathbf{u}_i \times (\mathcal{S} \times_2 \mathbf{u}_j \times_3 \mathbf{U}_{pix})_{(1)}\|_2 \tag{3}$$

pa imamo kandidata za aproksimaciju:

$$\mathbf{u}_i \; = \; \mathcal{A}_{test} \times (\mathcal{S} \times_2 \mathbf{u}_j \times_3 \mathbf{U}_{pix})^{+}_{(1)} \tag{4}$$

gdje s '+' označavamo Moore-Penroseov pseudoinverz.

- I_v kandidata za aproksimaciju $\{\mathbf{u}_i : i = 1, \ldots, I_v\}$ koje uspoređujemo sa retcima matrice $\mathbf{U}_I \longrightarrow$ tražimo $k \in \{1, \ldots, I_I\}$ koji minimizira sljedeći izraz

$$\|\mathbf{u}_i - \mathbf{U}_I(k, 1 : I_I)\|_2.$$

- Pomoću kosinusa kuta 2 vektora $\mathbf{u}_i$ i $\mathbf{U}_I(k, 1 : I_I)$ možemo računati:

$$\arg\max_{i,k} \frac{\langle \mathbf{u}_i, \mathbf{U}_I(k, 1 : I_I)\rangle}{\|\mathbf{u}_i\| \|\mathbf{U}_I(k, 1 : I_I)\|} \tag{5}$$

ALGORITAM

IN : $\mathcal{A}_{test}$

- Za svaki $\mathbf{u}_j \in \{\mathbf{U}_v(k, 1 : I_v) : k = 1, \ldots, I_v\}$ izračunaj

$$\mathbf{u}_i \;=\; \mathcal{A}_{test} \times (\mathcal{S} \times_2 \mathbf{u}_j \times_3 \mathbf{U}_{pix})^+_{(1)}$$

- Nađi $k \in \{1, \ldots, I_l\}$ koji minimizira sljedeći izraz

$$\|\mathbf{u}_i - \mathbf{U}_l(k, 1 : I_l)\|_2.$$

OUT: $k \in \{1, \ldots, I_l\}, i \in \{1, \ldots, I_v\}$
RJEŠENJE: slovo $=$ k

Testiranje

- Za svaki $j = 1, \ldots, 5$ iz tenzora izoliramo j-tu perspektivu

$$\mathcal{A}(:, j, :)$$

$\longrightarrow$ slova $\mathcal{A}(i, j, :), i = 1, \ldots, 24$ su *testni primjeri*.

- Ostatak tenzora ($\mathcal{A}(:, k, :)$, $k \neq j$) je tretiran kao *training set*.
- 7 slova (A, L, H, P, R, U i Y), kojima se siluete međusobno značajno razlikuju

7 slova

Perspektiva	Grey-scale	Binary
Kut 1	6	5
Kut 2	7	7
Kut 3	7	7
Kut 4	7	7
Kut 5	7	5
Prosjek	6.8	6.2

Tablica: Točnost po perspektivi (7 slova)

Perspektiva	Grey-scale	Binary
A, L, R, U	5	5
H	3	3
P	5	4
Y	5	4

Tablica: Broj pogodaka po slovu (7 slova)

24 slova

Perspektiva	Grey-scale	Binary
Kut 1	10	11
Kut 2	14	13
Kut 3	17	16
Kut 4	13	11
Kut 5	13	14
Prosjek	13.4	13

Tablica: Točnost po perspektivi (sva slova)

Perspektiva	Grey-scale	Binary
A, J, P	5	5
R, W, Y	4	4
D, K, S, Z	3	3
C, E, O, T, U	2	2
B, Q	1	1
H	0	0
F, I	2	1
G	5	4
V	0	1
X	3	4

Tablica: Broj pogodaka po slovu (sva slova)