

Tensor - Train dekompozicija za raspoznavanje slika

MTMAP - 2. seminarski rad

Doris Đivanović, Karlo Gjogolović, Vana Glumac

Akadska godina: 2024./2025.

1 Uvod

Na početku ovog rada sažeto ćemo definirati i izvesti neke osnovne činjenice o tenzorima i njihovim operacijama koje ćemo kasnije koristiti pri pisanju.

Neka je $\mathcal{A} \in \mathbb{R}^{n_i \times n_e \times n_p}$ tenzor koji predstavlja kolekciju slika s n_i piksela, n_p osoba te za svaku osobu n_e ekspresija lica. Dani skup slika možemo spremiti i u n_p disjunktne matrice $A_p \in \mathbb{R}^{n_i \times n_e}$, $p = 1, \dots, n_p$.

Idućom definicijom uvodimo pojam množenja tenzora s matricom.

Definicija: Za dani tenzor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ definiramo njegovo *množenje u modu n* sa matricom $U \in \mathbb{R}^{J \times I_n}$ u oznaci $\mathcal{A} \times_n U \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N}$ gdje su elementi od \mathcal{A} $(\mathcal{A} \times_n U)_{i_1, \dots, i_{n-1}, j, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} a_{i_1, i_2, \dots, i_n, \dots, i_N} u_{j, i_n}$

Propozicija: Za $n \neq m$ vrijedi $\mathcal{A} \times_n U \times_m V = \mathcal{A} \times_m V \times_n U$, tj. vrijedi komutativnost množenja modova ukoliko su oni različiti i ulančani.

Za $n = m$, tj. ako su modovi jednaki, vrijedi i više; $\mathcal{A} \times_n U \times_n V = \mathcal{A} \times_n (VU)$.

Idućom definicijom uvodimo pojam množenja tenzora s tenzorom.

Definicija: Za dani tenzor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ definiramo $\left(\frac{m}{n}\right)$ -produkt tenzora \mathcal{A} s tenzorom $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \times \dots \times J_M}$, (pri čemu vrijedi $I_n = J_m$) kao $C = \mathcal{A} \times_n^m \mathcal{B}$, gdje je $C \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N \times J_1 \times \dots \times J_{m-1} \times J_{m+1} \times \dots \times J_M}$.

Elementi od C su jednaki: $C(i_1, \dots, i_{n-1}, i_{n+1}, \dots, i_N, j_1, \dots, j_{m-1}, j_{m+1}, \dots, j_M) = \sum_{i_n=1}^{I_n} \mathcal{A}(i_1, \dots, i_{n-1}, i_n, i_{n+1}, \dots, i_N) \mathcal{B}(j_1, \dots, j_{m-1}, i_n, j_{m+1}, \dots, j_M)$.

Definicija: Za dani tenzor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ definiramo *proces matricizacije* (engl. *unfolding*). k -ta unfold matrica $A_k \in \mathbb{R}^{(I_1 I_2 \dots I_k) \times (I_{k+1} I_{k+2} \dots I_N)}$ definirana je kao:

$$A_k(i_1 * \dots * i_k, i_{k+1} * \dots * i_N) = \mathcal{A}(i_1, \dots, i_k, i_{k+1}, \dots, i_N)$$

Glavni cilj našeg rada je danu sliku z nepoznate osobe s nepoznatom ekspresijom lica, predstavljenu vektorom u \mathbb{R}^{n_i} pridružiti nekoj od n_p osoba kojoj je slika z po izgledu najbliža. U tu svrhu definirat ćemo udaljenost $dist(z, \mathcal{A}(:, \cdot, p))$ između z i svake osobe p iz naše baze podataka, za $p = 1, \dots, n_p$ koja će nam biti od pomoći pri implementaciji algoritama za raspoznavanje osobe.

2 Tensor-Train dekompozicija

Glavni cilj Tensor-Train dekompozicije je izbjegavanje ovisnosti o dimenziji. Tenzor reda N želimo zapisati kao produkt tenzora reda 3.

Definicija: Neka je $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_N}$ tenzor reda N . Kažemo da je \mathcal{A} u TT-formatu ako se njegovi elementi mogu zapisati u obliku:

$$\mathcal{A}(i_1, \dots, i_N) = G(i_1, :) \mathcal{G}_2(:, i_2, :) \mathcal{G}_3(:, i_3, :) \dots \mathcal{G}_{N-1}(:, i_{N-1}, :) G_N(i_N, :)$$

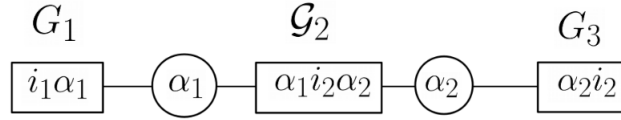
gdje su:

- $G_1 \in \mathbb{R}^{r_1 \times I_1}$, $G_N \in \mathbb{R}^{I_N \times r_N}$ i $\mathcal{G}_k \in \mathbb{R}^{r_k \times I_k \times r_{k+1}}$ za $k = 1, \dots, N$ **TT-jezgre**.
- r_k za $k = 1, \dots, N$ **TT-rangovi** pri čemu su $r_0 = r_N = 1$.
- $r = \max_{1 \leq k \leq N} r_k$ **maksimalni TT-rang**.

Izraz iz prethodne definicije možemo zapisati i kao

$$\mathcal{A}(i_1, \dots, i_N) = \sum_{\alpha_1, \dots, \alpha_{N-1}} G_1(i_1, \alpha_1) \mathcal{G}_2(\alpha_1, i_2, \alpha_2) \dots G_N(\alpha_{N-1}, i_N),$$

Na idućoj slici dan je primjer za vizualizaciju Tensor-Train dekompozicije koji koristi indekse tenzora reda 3. Sa slike je vidljivo da pravokutnici sadrže indekse izvornog



tenzora te barem jedan pomoćni indeks, a kružići sadrže samo pomoćni indeks. Pravokutnike povezujemo ako i samo ako imaju zajednički α_k između njih. Ukoliko ovakvo množenje interpretiramo kao *vagone*, dolazimo do obrazloženja zašto se ova dekompozicija naziva *Tensor-Train dekompozicija*.

Koristeći se definicijom $\left(\frac{m}{n}\right)$ -produkta tenzora \mathcal{A} gornju definiciju možemo također zapisati i kao

$$\mathcal{A} = G_1 \times_2^1 \mathcal{G}_2 \times_3^1 \mathcal{G}_3 \times_3^1 \dots \times_3^1 G_N$$

Ukoliko je \mathcal{A} tenzor reda 3, zbog

$$\mathcal{A}(i_1, i_2, i_3) = G_1(i_1, :) \times_2^1 \mathcal{G}_2(:, i_2, :) \times_3^1 G_3(:, i_3) \quad (1)$$

$$= \sum_{\alpha_1, \alpha_2} G_1(i_1, \alpha_1) \mathcal{G}_2(\alpha_1, i_2, \alpha_2) G_3(\alpha_2, i_3) \quad (2)$$

$$= \sum_{\alpha_2} (\mathcal{G}_2 \times_1 G_1)(i_1, i_2, \alpha_2) G_3(\alpha_2, i_3) \quad (3)$$

$$= (\mathcal{G}_2 \times_1 G_1 \times_3 G_3^T)(i_1, i_2, i_3) \quad (4)$$

slijedi da gornju dekompoziciju možemo zapisati i kao $\mathcal{A} = \mathcal{G}_2 \times_1 G_1 \times_3 G_3^T$. Kako bi izračuni s tenzorima bili računski prihvatljiviji, uobičajeno se koristi redukcija ranga, te dalje postupamo slično kao u klasičnom SVD postupku za matrice. Komentirajmo reduciranu verziju TT-SVD dekompozicije tenzora \mathcal{A} , koristeći rastav na singularne vrijednosti matrice A_1 . Prilikom redukcije dimenzije zadržavamo prvih k singularnih vrijednosti od A_1 , čiji zbroj čini $\pi 100\%$ ukupnog zbroja singularnih vrijednosti. Zbog toga uzimamo samo prvih k stupaca matrice G_1 , te definiramo konačnu matricu \hat{G}_1 . Posljedično, također definiramo i $\hat{\mathcal{G}}_2 \in \mathbb{R}^{k \times n_e \times n_p}$ kao tenzor prvih k elemenata prvog moda od \mathcal{G}_2 . Tako skraćena TT-SVD dekompozicija poprima oblik $\mathcal{A} = \hat{\mathcal{G}}_2 \times_1 \hat{G}_1 \times_3 G_3^T$.

3 Klasifikacija pomoću najmanjih kvadrata

Ova metoda klasifikacije fokusira se na matrice dobivene usporedbom sa svim mogućim slikama u danoj ekspresiji. Preciznije, za svaku ekspresiju lica e neka vrijedi $A^e := \mathcal{A}(:, e, :)$, pri čemu su stupci matrice A^e skalirani u jediničnoj normi. Promatramo sljedeći problem najmanjih kvadrata: $\min_x \|z - A^e x\|_2$. Neka je x_e rješenje gornjeg problema. Definirajmo je $r_e = z - A^e x_e$ kao pripadni rezidual. Udaljenost od z do kolekcije svih osoba n_p koje imaju ekspresiju lica e definiramo kao $\|r_e\|_2$. Tada možemo izračunati $\hat{e} = \arg \min_e \|r_e\|_2$ i klasificirati sliku z kao osobu \hat{p} , gdje je $\hat{p} = \arg \max_p |x_{\hat{e}}(p)|$, tj. \hat{p} odgovara najvećoj komponenti od $x_{\hat{e}}$, pri čemu $x_{\hat{e}}$ predstavlja linearnu kombinaciju svih osoba s najbližom ekspresijom lica. Prethodne rezultate možemo spojiti u idući algoritam koji ćemo implementirati i opisati njegove rezultate u posljednjem poglavlju.

Algorithm 1: Algoritam najmanjih kvadrata

Input: testna slika $z \in \mathbb{R}^{n_i}$ i tenzor \mathcal{A}

- 1: **for** $e = 1, \dots, n_e$ **do**:
 Riješi problem $\min_{x_e} \|z - A^e x_e\|_2$ te izračunaj $r_e = z - A^e x_e$
 end
 - 2: $\hat{e} = \arg \min_e (\|r_e\|_2)$
 $\hat{p} = \arg \max_p |x_{\hat{e}}(p)|$
 Output: Slika z pridružena je osobi \hat{p}
-

4 HOSVD

Koristeći HOSVD, tenzor $\mathcal{A} \in \mathbb{R}^{n_i \times n_e \times n_p}$ možemo zapisati kao $\mathcal{A} = S \times_i F \times_e G \times_p H$, gdje su x_i, x_e, x_p odgovarajuća mod 1, mod 2 i mod 3 množenja te $G \in \mathbb{R}^{n_e \times n_e}$, $H \in \mathbb{R}^{n_p \times n_p}$, te ukoliko vrijedi $n_i > n_e n_p$, tada je $S \in \mathbb{R}^{n_e n_p \times n_e n_p}$, $F \in \mathbb{R}^{n_i \times n_e n_p}$, a inače $S \in \mathbb{R}^{n_i \times n_e \times n_p}$ i $F \in \mathbb{R}^{n_i \times n_i}$.

Definirajmo $C = S \times_i F \times_e G$ za fiksnu ekspresiju lica e . Tada slijedi:

$$A^e = C^e H^T, \quad e = 1, 2, \dots, n_e,$$

gdje su A^e i C^e matrice dobivene fiksiranjem drugog indeksa \mathcal{A} i C na e . Primijetimo da je matrica H^T dimenzije $n_p \times n_p$ ortogonalna te da je C^e matrica dimenzije $n_i \times n_p$. Stupac matrice A^e možemo zapisati i kao $a_p^{(e)} = C^e h_p^T$, pri čemu je h_p^T p -ti stupac matrice H^T .

Neka je sada $z \in \mathbb{R}^{n_i}$ nova slika koju želimo klasificirati. Koordinate slike z sa ekspresijom lica e mogu se pronaći rješavanjem problema najmanjih kvadrata:

$$\hat{\alpha}_e = \arg \min_{\alpha_e} \|C^e \alpha_e - z\|_2, \quad e = 1, \dots, n_e.$$

Zatim, za svaki $e = 1, \dots, n_e$ i za svaki $p = 1, \dots, n_p$, računamo

$$D_{HO}(e, p) = \|\hat{\alpha}_e - h_p^T\|_2$$

Ako je C^e punog stupčenog ranga, tada problem najmanjih kvadrata ima jedinstveno rješenje. U suprotnom rješenje nije jedinstveno, ali se može pronaći s minimalnom normom pomoću SVD-a matrice C^e .

Propozicija: Neka je x_e rješenje $\min_x \|z - A^e x\|_2$ u metodi najmanjih kvadrata te neka je $\hat{\alpha}_e$ rješenje $\hat{\alpha}_e = \arg \min_{\alpha_e} \|C^e \alpha_e - z\|_2$ u metodi HOSVD. Tada vrijedi $\hat{\alpha}_e = H^T x_e$. Definiramo li s $I_{:,p}$ p -ti stupac kanonske baze, tada vrijedi

$$D_{HO}(e, p) = \|\hat{\alpha}_e - h_p^T\|_2 = \|x_e - I_{:,p}\|_2.$$

5 3D klasifikacijski algoritam pomoću Tensor-Train dekompozicije

Zapišimo prvo tenzor podataka \mathcal{A} za svaku ekspresiju lica e , a zatim ćemo definirati pripadnu udaljenost koju želimo minimizirati.

Uočimo, $\mathcal{A}(:, e, :)$ može se zapisati kao

$$\mathcal{A}(:, e, :) = G_2^e \times_i G_1 \times_p G_3^T, \quad \forall e = 1, \dots, n_e,$$

pri čemu je $G_2^e = \mathcal{G}_2(:, e, :)$ matrica.

Dakle, slika osobe p s ekspresijom lica e se računa kao

$$\mathcal{A}(:, e, p) = G_1 G_2^e g_3^p, \quad \text{gdje je } g_3^p = G_3(:, p).$$

Kako bismo riješili problem klasifikacije, stavimo $C = \mathcal{G}_2 \times_i G_1$. Uvrstimo li to u prethodni izraz, dobivamo

$$\mathcal{A}(:, e, p) = C^e g_3^p, \quad \text{gdje} \quad C^e = C(:, e, :) = G_1 G_2^e.$$

Objasnimo prethodni izraz. Stupci matrice C^e čine bazu za prostor osoba n_p sa istom ekspresijom lica e , dok izraz g_3^p sadrži koordinate svih slika osobe p s različitim izrazima lica.

Za danu sliku z nepoznate osobe n_p u nepoznatom izrazu e želimo pronaći koordinate α_e od z u svim n_e bazama $\{C^e\}_{e=1, \dots, n_e}$ te usporediti vektore α_e za $e = 1, \dots, n_e$ s koordinatama svih n_p osoba u istoj bazi e , koje su predstavljene stupcima matrice G_3 . Drugim riječima, za svaki $e = 1, \dots, n_e$ računamo $\min_{\alpha_e} \|C^e \alpha_e - z\|_2$, te nakon , za svaki $e = 1, \dots, n_e$ i za svaki $p = 1, \dots, n_p$, udaljenost $D_{TT}(e, p) := \|\hat{\alpha}_e - g_3^p\|_2$.

Dakle, udaljenost između slike z i osobe p dana je s

$$\text{dist}(z, \mathcal{A}(:, \cdot, p)) = \min_e D_{TT}(e, p).$$

Računanje $\min_{\alpha_e} \|C^e \alpha_e - z\|_2$, može se izvesti pomoću reducirane QR dekompozicije $G_2^e = Q^e R^e$. Koordinate z u prostoru osoba n_p sa istom ekspresijom lica e tada su dane s $\hat{\alpha}_e = R^{(e)-1} Q^{(e)T} G_1^T z$, pa tako iz izraza $D_{TT}(e, p) := \|\hat{\alpha}_e - g_3^p\|_2$ dobivamo

$$D_{TT}(e, p) = \|R^{(e)-1} Q^{(e)T} G_1^T z - g_3^p\|_2, \quad p = 1, \dots, n_p,$$

gdje je $g_3^p = G_3(:, p)$.

Implementacija je izrađena pomoću dolje definiranog Algoritma TT za 3D.

Algorithm 2: 3D TT algoritam

Input: testna slika z , matrice i tenzor G_1 , \mathcal{G}_2 i G_3

- 1: Izračunaj $\hat{z} = G_1^T z$
- 2: **for** $e = 1, \dots, n_e$ **do**:
 - Riješi sustav jednadžbi $G_2^e \alpha_e = \hat{z}$ po α_e
 - for** $p = 1, \dots, n_p$ **do**
 - $D_{TT}(e, p) = \|\alpha_e - g_3^p\|_2$, pri čemu je $g_3^p = G_3(:, p)$
 - end**
- end**
- 3: $(\hat{e}, \hat{p}) = \underset{e, p}{\operatorname{argmin}}(D_{TT})$

Output: Slika z pridružena je osobi \hat{p}

6 4D klasifikacijski algoritam pomoću Tensor-Train dekompozicije

Klasifikaciju slika možemo poboljšati promatrajući ih kao matrice $I \in \mathbb{R}^{n_1 \times n_2}$ umjesto vektora, čime problemu dodajemo dodatnu dimenziju. Stoga se baza podataka s n_p osoba sa n_e ekspresija reprezentira četverodimenzionalnim tenzorom $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_e \times n_p}$, zapisanom u TT-obliku kao $\mathcal{A} = G_1 \times_2^1 \mathcal{G}_2 \times_3^1 \mathcal{G}_3 \times_4^1 G_4$, gdje su $G_1 \in \mathbb{R}^{n_1 \times n_2}$, $\mathcal{G}_2 \in \mathbb{R}^{n_1 \times n_2 \times n_e \times n_p}$, $\mathcal{G}_3 \in \mathbb{R}^{n_e \times n_p \times n_e \times n_p}$ i $G_4 \in \mathbb{R}^{n_p \times n_p}$.

Dakle, slika osobe p sa ekspresijom e dana je kao

$$\mathcal{A}(:, :, e, p) = \mathcal{G}_2 \times_1 G_1 \times_3 (G_3^{(e)})^T \times_4 g_4^{(p)},$$

pri čemu je $G_3^{(e)} = \mathcal{G}_3(:, :, e, :)$ i $g_4^{(p)} = G_4(:, p) \in \mathbb{R}^{n_p}$.

Neka je $C^{(e)} = \mathcal{G}_2 \times_1 G_1 \times_3 (G_3^{(e)})^T$. Tada iz gornjeg izraza dobivamo

$$\mathcal{A}(:, :, e, p) = C^{(e)} \times_4 g_4^{(p)}.$$

Nadalje, možemo zaključiti da je sada strategija klasifikacije analogna je onoj za tro-dimenzionalni slučaj. Konkretno, za danu sliku $z \in \mathbb{R}^{n_1 \times n_2}$ neke nepoznate osobe sa nepoznatom ekspresijom lica, definiramo udaljenost od z od osobe p kao

$$D_{TT4}(e, p) = \|\hat{\alpha}_e - g_4^{(p)}\|_2,$$

pri čemu definiramo $\hat{\alpha}_e$ kao $\hat{\alpha}_e = \underset{\alpha_e}{\operatorname{argmin}} \|\operatorname{unfold}(C^{(e)})\alpha_e - z\|_2$.

Implementacija je izrađena pomoću dolje definiranog Algoritma TT za 4D.

Algorithm 3: 4D TT algoritam

Input: testna slika z , matrice i tenzori $G_1, \mathcal{G}_2, \mathcal{G}_3$ i G_4

- 1: Izračunaj $\mathcal{G}_{12} = \mathcal{G}_2 \times_1 G_1$
- 2: **for** $e = 1, \dots, n_e$ **do**:
 - Izračunaj $C^{(e)} = \mathcal{G}_{12} \times_3 (G_3^{(e)})^T$, pri čemu je $G_3 = \mathcal{G}_3(:, :, e, :)$
 - Izračunaj $\min_{\alpha_e} \|\operatorname{unfold}(C^{(e)})\alpha_e - z\|_2$ po α_e
 - for** $p = 1, \dots, n_p$ **do**
 - $D_{TT}(e, p) = \|\alpha_e - g_4^{(p)}\|_2$, pri čemu je $g_4^{(p)} = G_4(:, p)$
 - end**
- end**
- 3: $(\hat{e}, \hat{p}) = \underset{e, p}{\operatorname{argmin}}(D_{TT})$

Output: Slika z pridružena je osobi \hat{p}

TT format proširujemo i na višedimenzionalni slučaj. Na primjer, pretpostavimo da je \mathcal{A} tenzor N -tog reda koji predstavlja naše podatke, te pretpostavimo da su prva dva moda pikseli (kao u TT4D) i da je posljednji mod osoba. Tada prateći prethodne zapise, \mathcal{A} možemo zapisati kao

$$\mathcal{A} = G_1 \times_2^1 \mathcal{G}_2 \times_3^1 \mathcal{G}_3 \times_3^1 \cdots \times_3^1 \mathcal{G}_{N-1} \times_3^1 G_N.$$

Dakle, osobu p u svim mogućim kombinacijama $(N-3)$ značajki možemo zapisati kao

$$\mathcal{A}(:, :, i_3, \dots, i_{N-1}, p) = G_1 \times_2^1 \mathcal{G}_2 \times_3^1 \mathcal{G}_3^{(i_3)} \times_3^1 \cdots \times_3^1 \mathcal{G}_{N-1}^{(i_{N-1})} \times_3^1 G_N(:, p).$$

Za $C^{(i_3, \dots, i_{N-1})} = G_1 \times_2^1 \mathcal{G}_2 \times_3^1 \mathcal{G}_3^{(i_3)} \times_3^1 \cdots \times_3^1 \mathcal{G}_{N-1}^{(i_{N-1})}$ dobivamo da je gornji izraz moguće zapisati kao

$$\mathcal{A}(:, :, i_3, \dots, i_{N-1}, p) = C^{(i_3, \dots, i_{N-1})} \times_3^{(p)} g_N^{(p)},$$

gdje je $g_N^{(p)} = G_N(:, p)$.

S obzirom na sliku $z \in \mathbb{R}^{n_1 n_2}$ nepoznate osobe u nepoznatoj kombinaciji $(N-3)$ značajki, definiramo udaljenost od z od p kao

$$D_{TTN}(i_3, \dots, i_{N-1}, p) = \|\hat{\alpha}_{i_3, \dots, i_{N-1}} - g_N^{(p)}\| \quad \text{gdje je}$$

$$\hat{\alpha}_{i_3, \dots, i_{N-1}} = \arg \min_{\alpha_{i_3, \dots, i_{N-1}}} \|\text{unfold}_n(C^{(i_3, \dots, i_{N-1})}) \alpha_{i_3, \dots, i_{N-1}} - z\|.$$

7 Rezultati implementiranih algoritama

Testirali smo implementirane algoritme na bazama Orl i COIL20. Orl je baza koja sadrži po 10 slika različitih facijalnih ekspresija za 40 ljudi dimenzije 92x112. COIL20 je baza koja sadrži slike 20 igračaka slikanih iz 72 različita kuta dimenzije 128x128. U tablici su dane preciznosti klasifikacije koje smo računali kao $\frac{TK}{TK+NK}$ gdje je TK broj točno klasificiranih lica/igračaka, a NK broj netočno klasificiranih lica/igračaka.

Baza	TT 3D 2	TT 4D
Orl	90.00%	97.50%
COIL20	99.28%	99.64%

Tablica 1: Preciznost

Možemo vidjeti kako dobivamo značajno bolji rezultat za TT 4D primijenjen na bazu Orl u odnosu na TT 3D. Nema značajne razlike prilikom primjene na COIL20.

8 Literatura

- [1] D.Brandoni i V.Simoncini, TENSOR-TRAIN DECOMPOSITION FOR IMAGE RE-COGNITION (2019.)
- [2] D.Brandoni, Tensor decompositions for Face Recognition (2017.)
- [3] Ning Hao, Misha E Kilmer, Karen Baman, and Randy C Hoover. Facial recognition using tensor-tensor decompositions. SIAM Journal on Imaging Sciences (2013.)
- [4] Slajdovi s predavanja