

FACE PARAMETRIZATION FROM VIDEO USING FACE RECOGNITION METHODS

Pavol Lieskovský

Master Degree Programme (2), FEEC BUT

E-mail: xliesk01@stud.feec.vutbr.cz

Supervised by: Martin Rajnoha

E-mail: martin.rajnoha@vutbr.cz

Abstract: In the frame of this article, a program has been created that can extract persons face features from the video of an speaking person into a structured time series form suitable for further processing. Such a program utilizes machine learning and face recognition methods and it represents a solution to the general problem of face parametrization from video that can find its applications across many different fields.

Keywords: face recognition, face landmark estimation, face parametrization, time series

1 ÚVOD

V posledných desaťročiach napredujú Informačné technológie (IT) nezastaviteľným tempom. Pravidelne vznikajú nové technológie, ktoré nachádzajú svoje uplatnenie v čoraz viac odvetviach a na miestach, kde to pred pár rokmi nebolo možné. Mnohé procesy sa dajú vďaka novým technológiám automatizovať a tým pádom dochádza k úspore času. Jedným z rýchlo sa rozvíjajúcich odvetí IT je aj počítačové videnie. V dnešnej dobe sa s pojmom počítačové videnie stretávame na takmer pravidelnej báze. Využíva sa v takmer každom jednom odvetví, od poľnohospodárstva cez bankovnícky sektor až po medicínu [1]. V rámci tohoto článku bolo využité počítačové videnie na spracovanie videa rozprávajúcej osoby do formátu zachytávajúceho najdôležitejšie črty tváre a emócie osoby na každej snímke videa. Vytvorí sa tak časová rada s parametrami danej tváre. Využívajú sa k tomu práve metódy slúžiace pre identifikáciu osôb [2]. Nakoľko sa jedná o riešenie všeobecného problému, využiteľnosť tohoto riešenia má široké spektrum v rôznych oboroch. Medzi možnosťami využitia patrí napríklad rozpoznávanie štádia choroby pacienta trpiaceho niektorou z neurodegeneratívnych chorôb [3], vytvorenie datasetu časovej rady pre rozpoznávanie textu z hovorenej reči alebo aj identifikácia toho, či daná osoba povedala určitý text. Článok je rozdelený nasledovne: 2. časť obsahuje popis metód ktoré boli využité pri riešení a taktiež obsahuje popis samotného experimentu. 3. časť obsahuje popis dosiahnutých výsledkov a práca je ukončená zhodnotením v záveru v 4. časti.

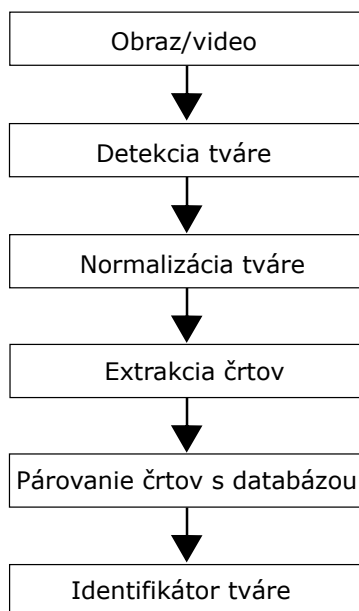
2 METODIKA

Samotné metódy počítačového videnia sú rozdelené do viacerých podskupín na základe toho, aký problém riešia. V súčasnosti je takmer štandardným postupom u obrazového spracovania dát využitie efektivity hlbokého učenia a konvolučných neurónových sietí [4]. V rámci riešenia boli použité konkrétne metódy na rozpoznávanie tváre z projektu facenet [2], keďže tie sa ukázali ako najvhodnejšie pre parametrizáciu tváre a sú úzko späté s problematikou.

2.1 ROZPOZNÁVANIE TVÁRE

Pokroky v automatizovanej analýze tváre, párovaní vzorov a počítačovom učení viedli k možnosti vytvorenia systémov automatického rozpoznania tváre. Proces rozpoznania tváre je pre človeka pri-

rodzený a automaticky vykonávaný proces, avšak problém spočíva v zapamätaní si veľkého množstva neznámych tvárí. V oblasti počítačového videnia sa stále jedná o jeden z obtiažných a zložitých procesov, no aj napriek tomu už existujú rôzne metódy, ktoré zvládajú identifikáciu tváre s rozsiahlymi databázami. Tieto metódy nachádzajú svoje využitie napríklad v oblastiach zaoberajúcich sa bezpečnosťou. Systém vykonávajúci počítačovú identifikáciu tváre sa väčšinou skladá zo šiestich prvkov zobrazených na Obr. 1. Jedným z kľúčových problémov detekcie tváre je výkonnosť detekcie. Je potrebná extrakcia kľúčových črtov tváre a konštrukcia robustných klasifikátorov. To všetko s ohľadom na výkon. Týmto problémom sa zaoberá čoraz viac odborníkov [5].



Obr. 1: Typické prvky systému rozpoznávania tváre [5].

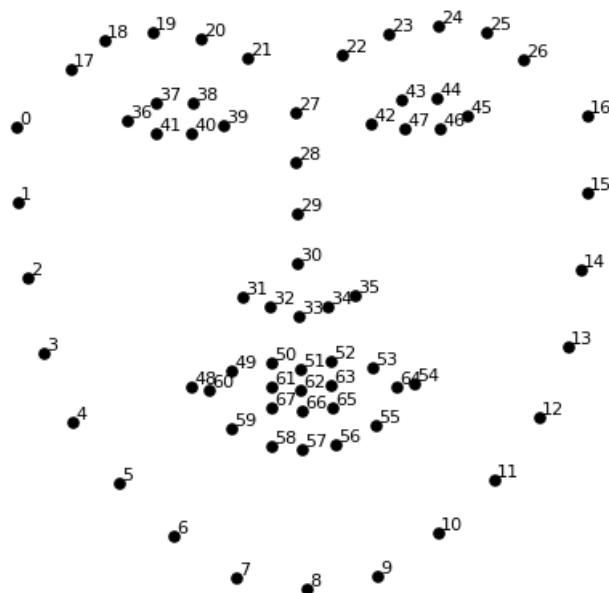
2.2 ODHAD BODOV TVÁRE

Odhad bodov tváre (face landmark estimation) je jednou z metód slúžiacich na rozpoznávanie tváre. Je definovaná ako detekcia a lokalizácia určitých kľúčových bodov tváre v obrázku. Hoci sa môže zdať tento problém pomerne ľahko riešiteľný ukázal sa ako mimoriadne náročný vďaka veľkému množstvu faktorov ovplyvňujúcich výpočet bodov. Medzi ne patria napríklad póza, osvetlenie, výraz tváre alebo aj to, či daná osoba má fúzy alebo okuliare [6].

Algoritmy na detekciu bodov tváre sa delia na tri hlavné kategórie: holistické, obmedzeného lokálneho modelu (Constrained local model–CLM) a metódy založené na regresii. Rozdielne sú v spôsobe akým modelujú vzhľad tváre a vzory tvaru črtov tváre. Vzhľad tváre predstavuje rozdielnosť v intenzite pixelov okolo kľúčových bodov tváre alebo v rámci celej tváre. Vzory tvaru črtov tváre sa vzťahujú na vzory tvarov črtov tváre v miestach kde sa detekujú kľúčové body. Holistické metódy explicitne modelujú holistický vzhľad tváre a globálne vzory tvaru črtov tváre. CLM metódy sa spoliehajú na explicitný lokálny vzhľad tváre a explicitné globálne vzory tvaru črtov tváre. Metódy založené na regresii používajú holistické alebo lokálne informácie o vzhľade a môžu zahŕňať globálne vzory tvaru črtov tváre implicitne pre spoločnú detekciu orientačných bodov. Vo všeobecnosti metódy založené na regresii vykazujú lepší výkon [7].

2.3 PARAMETRIZÁCIA TVÁRE DO ČASOVEJ RADY

Riešenie je realizované v programovacom jazyku Python. Predovšetkým je využitá knižnica `face_recognition` založená na implementácii `facenet` [2], ktorá je využitá pre detekciu bodov tváre. Myšlienka tohoto riešenia spočíva v sledovaní viacerých parametrov tváre, ako napríklad šírka a výška úst, šírka a výška tváre, vzdialenosť medzi nosom a bradou apod. Pri detekcii črtov tváre sú najprv s využitím metódy `face_landmarks` knižnice `face_recognition` získané body popisujúce tvár. Tieto body sú zobrazené na Obr. 2.



Obr. 2: Body popisujúce tvár získané metódou `face_landmarks` [8].

Z týchto bodov sú pomocou matematických operácií vyrátané výsledné hodnoty, ktoré sú následne naformátované do objektu typu JSON popisujúceho tvár na aktuálnom snímku. Zároveň sa medzi parametre ukladá aj odhad emócie z tváre.

Odhad emócií je vykonávaný na základe modelu natrénovaného na datasete `fer2013` [9]. Jedná sa o dáta skladajúce sa z 35887 obrázkov tvárí v stupnici šedej o rozmeroch 48x48. Každý z obrázkov je označený emóciou, ktorú tvár na danom obrázku vystihuje. Celkovo sa jedná o sedem typov emócií: hnev, znechutenie, strach, radosť, smútok, prekvapenie a neutrálna. Snímka je pred odhadom emócie transformovaná do rozmerov 48x48 pixelov a následne je na nej prevedený odhad na základe modelu. Výsledkom odhadu je vektor siedmich číselných hodnôt, ktoré predstavujú percentuálne zastúpenie každej jednej odhadovanej emócie v snímke, pričom poradie čísla v zozname udáva o akú emóciu sa jedná. Na Obr. 3 sú zobrazené dáta snímky tváre, v ktorej bola odhadnutá na 93% emócia prekvapenia, 6% strach, 1% smútok atď.

Tieto parametre sú vypočítané a následne zaznamenané v každom snímku videa. Výstupom skriptu sú dáta vo formáte JSON¹ uložené do súboru, ktorý je zadaný ako argument skriptu.

¹JSON (JavaScript Object Notation) je jedným z často využívaných formátov na výmenu a formátovanie dát. Využíva človekom čitateľný text, ktorý sa skladá z dvojíc kľúč-hodnota[10].

Samotné spustenie skriptu vyžaduje určité argumenty:

- -i cesta k súboru s videom,
- -w cesta k súboru s váhami pre odhad emócií,
- -o cesta k súboru do ktorého bude uložený výstup,
- -v zvýšenie verbozity skriptu,
- -g zapnutie grafického výstupu pre ukážku a ladenie.

Prvé tri z argumentov sú povinné a bez ich definovania sa skript nepustí. Zvyšné dva argumenty sú voliteľné.

Všetky vyrátané hodnoty popisujúce črty tváre aj s ukážkou výsledného formátu JSON-u sú popísané v kapitole 3. V prípade zapnutého grafického výstupu táto metóda ešte vykreslí jednotlivé detekované črty do aktuálneho snímku a zobrazí ho užívateľovi. Ukážka je na Obr. 4, taktiež v nasledujúcej kapitole.

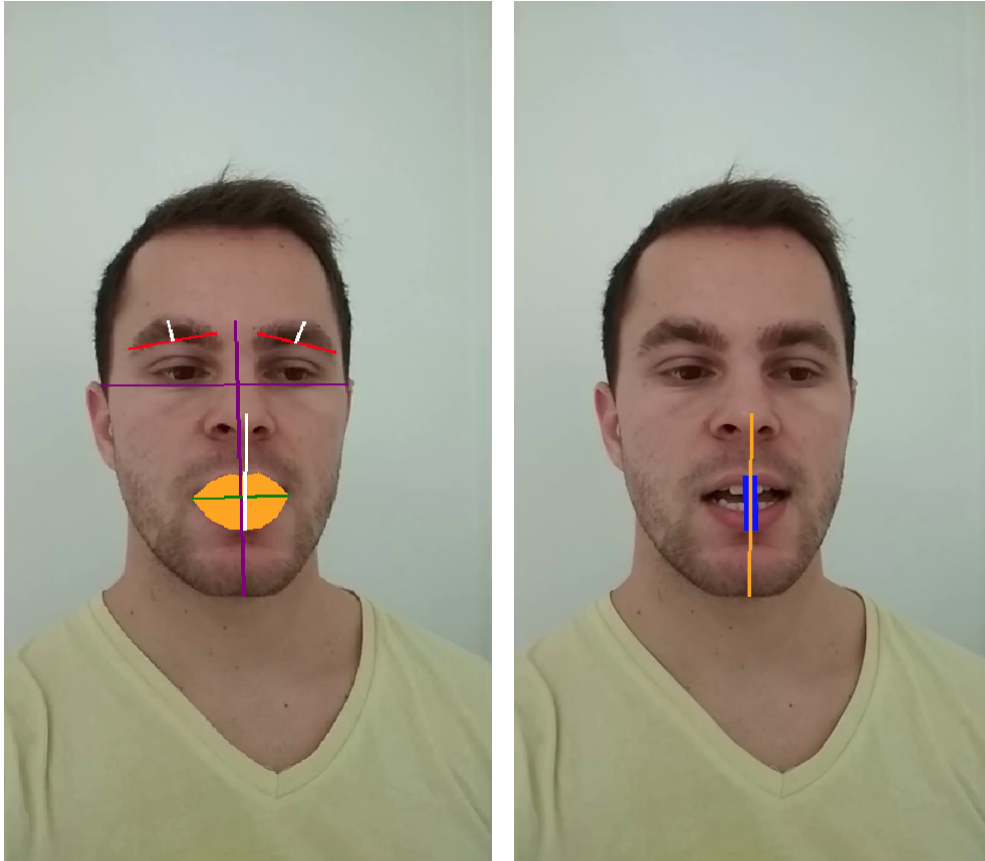
3 VÝSLEDKY

Výsledkom práce sú dáta popisujúce črty a emócie tváre zobrazenej na každom snímku spracovaného videa v sekvenčnom pradí. Formát dát je zobrazený na Obr. 3.

```
{
  " video ":{
    " file name ":" videos\\video . avi " ,
    " frames count ":" 1 " ,
    " fps ":" 25 " ,
    " frames ":[
      {
        " frame index ":" 1 " ,
        " mouth area ":" 2647.500 " ,
        " mouth width ":" 80.056 " ,
        " mouth height ":" 45.011 " ,
        " face width ":" 205.010 " ,
        " face height ":" 228.592 " ,
        " left eyebrow width ":" 73.164 " ,
        " left eyebrow height ":" 16.978 " ,
        " right eyebrow width ":" 66.940 " ,
        " right eyebrow height ":" 18.200 " ,
        " dist nose mouth ":" 97.005 " ,
        " dist nose chin ":" 152.013 " ,
        " emotions ":" {0.0026075528 , 9.9731415e-06 , 0.058610495 ,
          4.5851215e-05 , 0.013636396 , 0.9250778 ,
          1.2037376e-05}"
      }
    ]
  }
}
```

Obr. 3: Ukážka JSON formátu výsledných dát (pre 1 snímok).

Z formátu dát je možné vidieť, že prvým kľúčom je „video“. Jemu náležiaci hodnota je objekt popisujúci video. Ten obsahuje popis názvu videa, počet celkových snímok videa, počet snímokov za sekundu a pole objektov popisujúcich jednotlivé snímky. Každý jeden snímok je popísaný svojím číslom v poradí videa, a črtami detekovanými z tváre osoby. Tieto črty sú: plocha úst, šírka úst, výška úst, šírka tváre, výška tváre, šírka ľavého obočia, výška ľavého obočia, šírka pravého obočia, výška pravého obočia, vzdialenosť nosu a úst, vzdialenosť nosu a brady a hodnoty odhadu emócií. Na Obr. 4 sú rôznymi farbami zakreslené jednotlivé detekované črty (z dôvodu prekryvania sú zakreslené do dvoch obrázkov).



Obr. 4: Grafický výstup detekcie črtov tváre.

4 ZÁVER

V rámci článku bola predstavená metóda na extrakciu významných črtov tváre a následnú parametrizáciu tváre do štrukturovanej formy v časovej rade. Metóda využíva techniky strojového učenia a rozpoznávania tváre, konkrétne face landmark estimation, na základe ktorých sú počítané základné parametre tváre. Ďalej využíva už natrénovaný model pre odhad emócií, vytvárajúci pravdepodobnostný vektor. Medzi sledované parametre tváre patrí napríklad vzdialenosť nos-brada alebo šírka a výška úst apod. V 3. kapitole sú popísané dosiahnuté výsledky práce, vrátane ukázkového formátu dát. Výsledné dáta sú vo formáte JSON, ktorý je široko podporovaný a vhodný pre strojové spracovanie. Takto pripravené dáta sú vhodné pre akékoľvek počítačové spracovanie. Nakoľko sa jedná o dáta popisujúce črty a emócie tváre osoby na videu v čase, využiteľnosť tohoto riešenia má široké spektrum v rôznych oboroch.

LITERATÚRA

- [1] SZIELSKI, R. (2010). *Computer Vision: Algorithms and Applications - electronic draft*.
- [2] SCHROFF, F., KALANICHENKO, D., PHILBIN, J. *Facenet: A unified embedding for face recognition and clustering*. Proceedings of the IEEE conference on computer vision and pattern recognition, s. 815–823, 2015.
- [3] REKTOROVÁ, Irena. *Současné možnosti diagnostiky a terapie Parkinsonovy nemoci*. *Neurológia pre prax: Supplement 2*. 2009, 10(S2). ISSN 1335-9592. 27. 10.1142/S0218001413560053.
- [4] CHEN, Y., LIN, Z., ZHAO, X., WANG, G., GU, Z. *Deep Learning-Based Classification of Hyperspectral Data*. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 7, č. 6, s. 2094-2107, 2014.
- [5] WÓJCIK, W., GROMASZEK, K., JUNISBEKOV, M. *Face Recognition: Issues, Methods and Alternative Applications*[online]. *Face Recognition IntechOpen*, 6.7.2016. DOI: 10.5772/62950. [cit. 21.10.2018]. Dostupné z URL:<https://www.intechopen.com/books/face-recognition-semisupervised-classification-subspace-projection-and-evaluation-methods/face-recognition-issues-methods-and-alternative-applications>.
- [6] OUANAN, Hamid & OUANAN, Mohammed & AKSASSE, B. (2016). *Facial landmark localization: Past, present and future*. 487-493. 10.1109/CIST.2016.7805097.
- [7] WU, Yue & JI, Qiang. (2018). *Facial Landmark Detection: A Literature Survey*. *International Journal of Computer Vision*. 10.1007/s11263-018-1097-z.
- [8] SHAKIROV, N. *Face recognition: realtime masks development*[online]. Dostupné z URL:<https://becominghuman.ai/face-recognition-realtime-masks-development-9d3a399b4c3>.
- [9] KINLI, F. *[Deep Learning Lab] Episode-3: fer2013*[online]. Dostupné z URL:<https://medium.com/deep-learning-turkey/deep-learning-lab-episode-3-fer2013-c38f2e052280>.
- [10] Wikipedia contributors. *JSON* [online]. *Wikipedia, The Free Encyclopedia*. 10.10.2018 [cit. 22.10.2018]. Dostupné z URL:<https://en.wikipedia.org/wiki/JSON>.