



SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE FAKULTA ELEKTROTECHNIKY  
A INFORMATIKY

**Ing. Filip Csóka**

**Autoreferát dizertačnej práce  
Rozpoznávanie sekvencií znakov prstovej abecedy**

**na získanie akademického titulu:** „doktor“ („philosophiae doctor“, v skratke „PhD.“)

**v doktorandskom študijnom programe:** Telekomunikácie

**v študijnom odbore:** 5.2.15. Telekomunikácie

**Forma štúdia:** denná

**Miesto a dátum:** Bratislava 27.5.2019

**Dizertačná práca bola vypracovaná na:**

Ústav multimediálnych informačných a komunikačných technológií  
Fakulta elektrotechniky a informatiky  
Slovenská technická univerzita v Bratislave

**Predkladateľ:**

Ing. Filip Csóka  
Slovenská technická univerzita v Bratislave  
Fakulta elektrotechniky a informatiky  
Ústav multimediálnych informačných a komunikačných technológií  
Ilkovičova 3, 812 19 Bratislava 1

**Školiteľ:**

prof. Ing. Jaroslav Polec, PhD.  
Slovenská technická univerzita v Bratislave  
Fakulta elektrotechniky a informatiky  
Ústav multimediálnych informačných a komunikačných technológií  
Ilkovičova 3, 812 19 Bratislava 1

**Oponenti:**

doc. Ing. Libor Hargaš, PhD.  
Žilinská univerzita,  
Elektrotechnická fakulta,  
Katedra mechatroniky a elektroniky,  
Univerzitná 8215/1, 010 26 Žilina

doc. Ing. Vanda Benešová, PhD.  
Slovenská technická univerzita v Bratislave  
Fakulta informatiky a informačných technológií  
Ústav počítačového inžinierstva a aplikovanej informatiky  
Ilkovičova 2, 842 16 Bratislava 4

**Autoreferát bol rozoslaný:.....**

**Obhajoba dizertačnej práce sa bude konať dňa 27.8 2019 o 11:00 h.**

**Na Ústave multimediálnych informačných a komunikačných technológií FEI STU v BA**

.....  
**prof. Dr. Ing. Miloš Oravec**  
**dekan fakulty**

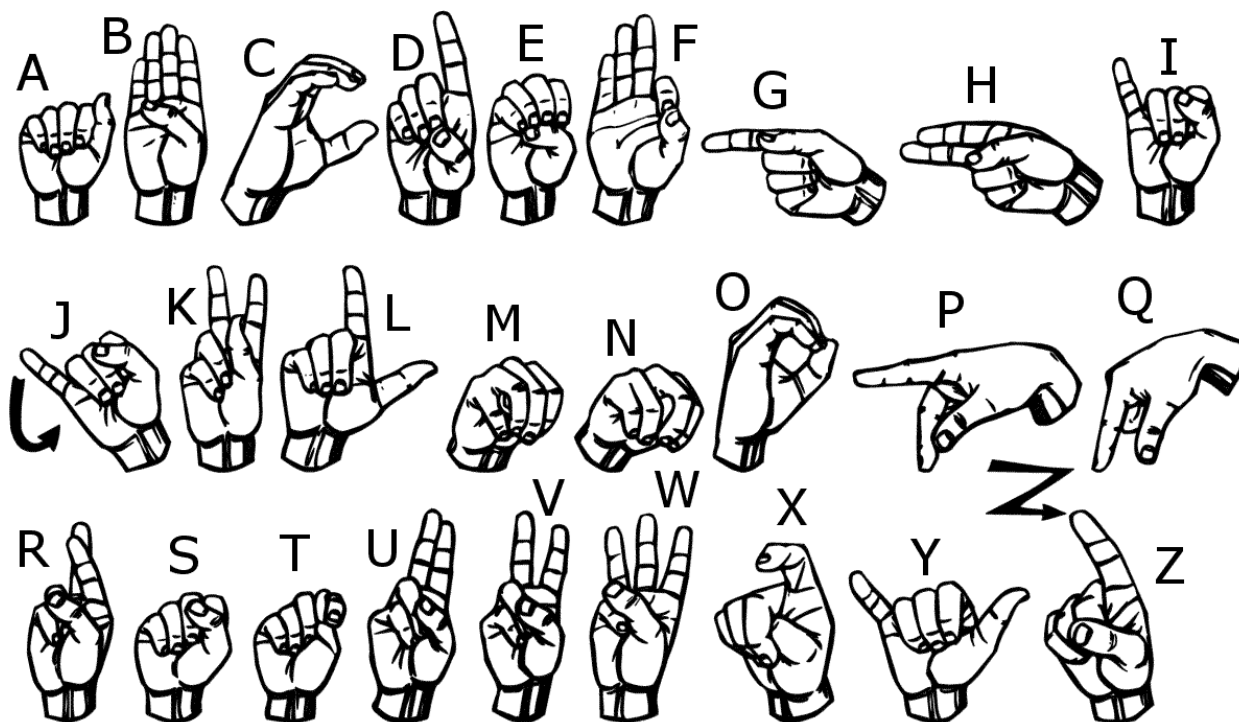
## Obsah

|       |                                                  |    |
|-------|--------------------------------------------------|----|
| 1     | Úvod .....                                       | 4  |
| 2     | Stručný prehľad problematiky .....               | 5  |
| 3     | Ciele dizertačnej práce .....                    | 7  |
| 4     | Finálna metóda riešenia dizertačnej práce .....  | 8  |
| 4.1   | Orientácia kamery a sledovanie pohybu ruky ..... | 8  |
| 4.2   | Predspracovanie obrazov .....                    | 9  |
| 4.3   | Extrakcia charakteristických príznakov .....     | 11 |
| 4.3.1 | Filter Laplacian of Gaussian .....               | 11 |
| 4.3.2 | Algoritmus hľadania lokálnych extrémov .....     | 13 |
| 4.4   | Klasifikácia obrazov .....                       | 16 |
| 4.4.1 | Štruktúra neurónovej siete a jej tréning .....   | 16 |
| 5     | Výsledky a zhodnotenie .....                     | 19 |
| 6     | Záver .....                                      | 21 |
| 7     | Resume .....                                     | 23 |
| 8     | Prínosy dizertačnej práce .....                  | 25 |
| 9     | Vlastná publikačná činnosť .....                 | 26 |
| 9.1   | Publikácie týkajúce sa dizertačnej práce .....   | 26 |
| 9.2   | Ostatné publikácie .....                         | 27 |
| 10    | Zoznam použitej literatúry .....                 | 29 |

# 1 Úvod

Možnosti zaradenia sa do spoločnosti sú pre ľudí so sluchovým postihnutím značne obmedzené. Často dokážu s okolitým svetom komunikovať iba pomocou posunkového jazyka. Aj preto je veľmi potrebné, aby existovali systémy, ktoré dokážu na základe snímania posunkového jazyka vygenerovať čitateľný text v reálnom čase. V dnešnej dobe značná časť populácie pravidelne nosí so sebou inteligentné mobilné telefóny, ktoré majú dostatočný výpočtový výkon a dostatočné rozlíšenie kamery na to, aby takúto funkciu teoreticky mohli plniť. V súčasnosti sa taktiež rozširuje používanie inteligentných okuliarov, ktoré majú v sebe implementovanú snímacu a zobrazovaciu jednotku, takže zachytené sekvencie posunkového jazyka by mohli rozpoznávať a zobrazovať v podobe bežného textu používateľovi rovno na obrazovku [1].

Slovenský posunkový jazyk má dva typy znakov. V prvom type reprezentuje jeden znak (posunok) celé slovo alebo frázu. V druhom type reprezentuje znak písmeno abecedy. Druhý typ sa volá prstová abeceda a používajú ju nepočujúci v komunikácii, keď chcú označiť niečo, čo nemá v prvej verzii ekvivalentné slovo, ako sú napríklad mená, názvy ulíc a miest a pod [1].



Obr. 1. Grafické znázornenie prstovej abecedy [1]

## 2 Stručný prehľad problematiky

Rozpoznávanie obrazu a počítačové videnie sú jednou z aktuálne najskúmanejších vedných oblastí. Výskumné inštitúcie a korporácie po celom svete sa predbiehajú v tom, kto navrhne lepší algoritmus, ktorý by dokázal plniť funkciu podobnú ľudskému neurovizuálnemu systému.

Viacero štatistických modelov, ktoré boli používané na klasifikáciu dát, nenašlo uplatnenie v klasifikácii obrazu. Modely spomínané v [47], ale aj tradičné modely ako lineárna regresia, Gaussov zmesový model, lineárna klasifikácia, skryté Markovove modely, rozhodovacie stromy, ale aj mnoho iných, zlyháva pri nespracovaných obrazových dátach zo skutočného sveta (fotografiách a video snímkach). Z praktického hľadiska sa ukázali nepoužiteľné a jedinou aktuálne preukázateľne fungujúcou metódou na rozpoznávanie obrazov sú neurónové siete, a to najmä konvolučné neurónové siete [48 - 51]. Dôvod na to, prečo bežné štatistické modely nefungujú, je nasledujúci: „ľudský neurovizuálny systém je príliš komplikovaný na to, aby mohol byť exaktne modelovaný a aktuálne nie je ani úplne preskúmaný“ [52 - 53].

Ak má byť navrhnutý deterministický model, ktorý presne opisuje, ako rozpoznať objekty tak, ako by to urobil človek, potom musí tento model zohľadňovať všetky faktory, ktoré ovplyvňujú to, ako rozpoznáva objekty človek. Tieto faktory nielen že sú neúplne zmapované, ale aj tie, ktoré sú zmapované, sa týkajú veľkého množstva vedných disciplín. Od fyzikálnych princípov, akými dopadá svetlo do ľudského oka, biologických princípov, akými sa daný stimul šíri nervovou sústavou, neuropsychologických princípov, akými nervový systém dané stimuly spracuje, až po psychologické vplyvy, ktoré taktiež značne ovplyvňujú ľudskú schopnosť rozoznávať objekty.

Systém schopný rozpoznávať obrazovú informáciu tak, ako by to urobil človek, by mal do čia najväčšej možnej miery aproximovať spôsob, akým fyzicky vidí človek. Spôsob, akým človek vníma stimuly, ktoré sa mu dostanú do mozgu, možno najlepšie označiť ako vnímanie hodnôt [54]. Aktuálne štúdie naznačujú, že ľudský mozog najprv spracuje vizuálne podnety a priradzuje im hodnotu, a až keď je tá hodnota dostatočná, človek ten objekt fyzicky „uvidí“ [52 - 54]. Keď analyzujeme, ako fungujú naše zmysly, musíme brať do úvahy, prečo vznikli. U živočíchov sa prvotne zrak a ani sluch nevyvinuli za účelom komunikácie [53], ale ľudské oči sú prispôbené spoločenskému životu v kolektíve jednotlivcov a plnia i komunikačnú úlohu [53]. Aj z tohto dôvodu sa u ľudí pravdepodobne vyvinuli očné bielka, vďaka ktorým ostatní ľudia vedia, kam sa kto díva, čo je pre ľudskú komunikáciu a medziľudskú interakciu veľmi dôležité [53].

Ľudské oči sú podvedome riadené orientačným reflexom a obraz, ktorý je v nich zachytený, je predspracovaný vizuálnym kortexom v záhlavnom laloku ľudského mozgu [54]. Oči zachytávajú dopadajúce svetlo približne v 120 stupňoch a z toho človek ostro vníma len 5° až 7°. Zvyšok tvorí periférne videnie. Je to spôsobené nerovnomerným rozložením tyčiniek a čapíkov na ľudskej sietnici, ktoré sa z veľkej časti sústreďujú na malej ploche. Toto vyvoláva aj nutnosť hýbať očami pri čítaní kníh aj napriek tomu, že sa celá kniha nachádza v zornom poli čítajúceho. Priemerný zdravý človek vidí na knihe vzdialenej od neho 1 meter naraz ostro 4 slová [54].

Fyziológia oka spôsobuje, že človek musí pri akejkol'vek činnosti, kde pracuje s vizuálnymi podnetmi (teda takmer pri každej činnosti), hýbať očami. Tento pohyb je plne riadený vizuálnym

kortexom podobne, ako je fungovanie iných orgánov riadené autonómnou nervovou sústavou. Z časti je riadenie tohto pohybu inštinktívne a z časti naučené. Biologických a psychologických podnetov, ktoré ovplyvňujú tento pohyb, je veľa a nie všetky sú dodnes známe [53-55]. Je však známe, že aj pri najsústredenejšom, úmyselnom zameriavaní sa oči na nejaké miesto, napríklad pri pozorovaní objektu v diaľke, existujú vizuálne stimuly, ktoré dokážu “donútiť” pozorovateľa, aby zameral svoj pohľad niekam inam, aj napriek svojej vôli a koncentrácii.

Jeden z takýchto stimulov je napríklad vlnivý pohyb hada v periférnom videní [53]. Ide o biologický pud podobný otočeniu sa za zdrojom náhlych intenzívnych zvukov (napríklad výbuchy, spadnuté predmety, rozbité veci, ...). Tento pud človek zdieľa s väčšinou primátov, až na pár druhov, ktoré sa vyvinuli v prostredí, kde sa z evolučného hľadiska dlhodobo nevyskytovali hady [53]. Vývoj v prostredí s hadmi bol podľa evolučných biológov jeden z najvýznamnejších vplyvov, ktorý donútil primáty vyvinúť si zrak, ktorý dnes majú [53]. Dlhodobý výskum ukázal, že opice, ktoré sa vyvinuli na izolovaných ostrovoch, na ktorých sa nevyskytujú hady, majú aj podstatne horší zrak ako porovnateľné primáty z iných oblastí [53]. Iné výskumy dokázali, že pri vykonávaní činnosti, ktorej neoddeliteľnou súčasťou je sledovanie niečoho (práca s textom za počítačom, práca s malými predmetmi, sledovanie lopty pri športoch, sústredené sledovanie niekoho), pozorovateľ nevníma a nevidí iné objekty v jeho zornom poli [55], ak mu svojou prítomnosťou nebránia v činnosti, ktorú vykonáva, neohrozujú ho alebo v ňom neaktivujú nejakú inštinktívnu pudovú reakciu [54].

Obzvlášť známe boli výskumy a pokusy opisované v [55], ktoré sa snažili zistiť, čo všetko si človek nevšimá, aj keď mu to priamo prejde zorným poľom a miestami, na ktoré sa sústreďí. V [54] sa venovali skúmaniu objektov, ktoré človek nevidí aj napriek tomu, že ich mal v zornom poli, zameral sa očami na ne alebo cez ne pomaly prešiel očami. Obidve tieto štúdie skúmali značne rozličné situácie rôznou metodikou a prišli k podobným výsledkom. Človek pri vykonávaní akejkoľvek aktivity podvedome a neúmyselne odfiltruje viac ako polovicu vizuálnej informácie. Pri niektorých situáciách a aktivitách človek nevidí viac ako 90 percent toho, na čo sa pozerá [54, 55].

Moderná neuropsychológia predpokladá, že ľudský mozog priradzuje vizuálnym stimulom hodnotu pred tým, ako ich detailne rozpoznáva [54], preto je možné dívať sa na objekty a nevidieť (nevnímať) ich. Ak je objekt v danej situácii považovaný za bezvýznamný, tak ho človek jednoducho ignoruje. Útes človek nevníma rovnako ako útes, ale mozog mu priradí hodnotu ako nebezpečné miesto, z ktorého sa dá spadnúť, a až potom človek spozná samotný útes.

Podobná situácia nastáva, keď človek hľadá mobilný telefón alebo kľúče na zapratanom stole. Nevidí objekty, ktoré sú od toho veľmi odlišné, jedine ak by ich vnímal ako prekážku, pod ktorou by ten hľadaný predmet mohol byť. Tento princíp platí aj naopak. Ak je nejaký objekt v situácii priveľmi dôležitý a človek ho sleduje, tak je možné, že si nevšimne iné trochu menej dôležité objekty, ktoré by si za iných okolností všimol [55]. Spôsob, akým človek vidí a čo človek reálne vidí, závisí z veľkej časti od situácie, v akej sa nachádza. V súčasnej dobe neexistuje model, ktorý by presne opísal, ako funguje ľudské videnie, aké objekty a prečo človek rozpoznáva a čo filtruje. Ľudský zrak je ovplyvňovaný aj inými zmyslami [54] a je možné, že bez namodelovania celej nervovej sústavy je vytvorenie modelu ľudského videnia nerealizovateľné.

### 3 Ciele dizertačnej práce

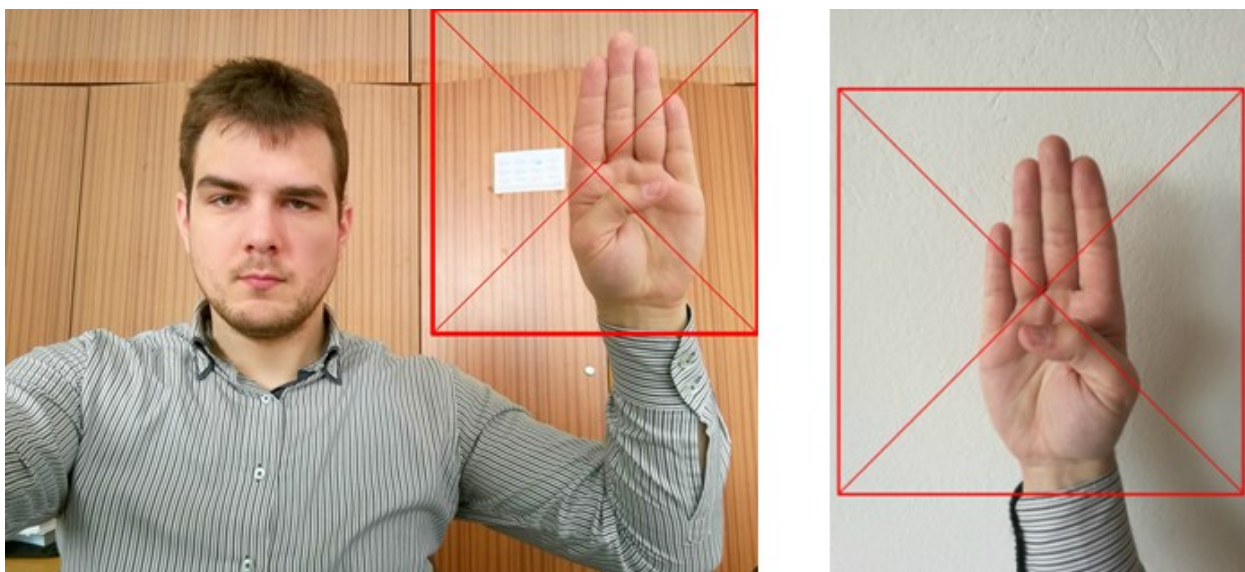
1. Cieľom dizertačnej práce je návrh metodiky na rozpoznávanie snímok prstovej abecedy s prihliadnutím na rýchly rozvoj moderných technológií, konkrétne rozlíšenie snímacích čipov kamier a narastajúci výpočtový výkon procesorov a paralelizácia výpočtov na viacerých výpočtových jednotkách.
2. V prvom kroku bude obraz ruky segmentovaný pomocou farebných segmentácií a morfológických operácií, čím bude oddelený obraz ruky a pozadia. Tento krok je nutný na zvýšenie úspešnosti rozpoznávania konkrétnych znakov. Elimináciou informácií z pozadia sa eliminuje aj ich vplyv na výslednú klasifikáciu.
3. V ďalšom kroku sa zo segmentovaného obrazu extrahujú charakteristické príznaky. Na túto extrakciu bude použitý adaptívny Laplaceov diskretný operátor. Ten bude použitý na odfiltrovanie redundantnej informácie z obrazu a čiastočnú dekoreláciu informácie v obraze.
4. Samotná extrakcia príznakov bude prebiehať pomocou hľadania lokálnych extrémov v Laplaceových obrazoch tak, aby tieto príznaky predstavovali hrany a charakteristické črty ruky pre jednotlivé znaky prstovej abecedy.
5. Následne bude zjednotená dimenzia získaných príznakov do jednotného rozmeru, aby mohli byť klasifikované obrazy s rôznym vstupným rozlíšením. Jednotný rozmer bude vybraný tak, aby mohli byť klasifikované aj obrazy s veľkým vstupným rozlíšením v reálnom čase (pre prstovú abecedu to znamená 3 znaky za sekundu).
6. Nakoniec budú príznaky klasifikované pomocou vhodne navrhutej konvolučnej neurónovej siete. Jej štruktúra bude zvolená tak, aby mohli byť znaky klasifikované v reálnom čase s čo najväčšou úspešnosťou.

## 4 Finálna metóda riešenia dizertačnej práce

Počas môjho doktorandského štúdia sme navrhli a otestovali niekoľko riešení zvolenej témy. 3 z nich sa ukázali ako použiteľné v praxi a boli aj publikované. Ako najlepšia sa ukázala finálna metóda, ktorá pri klasifikácii brala do úvahy poznatky o ľudskom videní. Nasledujúce podkapitoly opisujú práve túto metódu.

### 4.1 *Orientácia kamery a sledovanie pohybu ruky*

Ako prvý musí byť vyriešený problém, čo bude danú kameru smerovať na ruky človeka komunikujúceho znakovou rečou alebo v prípade snímania veľkej plochy, akú časť obrazu vyrezať tak, aby sa na nej nachádzali celé ruky a podľa možnosti nič iné. Pri finálnom riešení bola zvolená metóda využívajúca ľudský orientačný reflex. Metóda je navrhnutá tak, aby mohla fungovať v dvoch módoch. V prvom má ukazujúci človek pred sebou kameru a display, na ktorom sa vidí a sám umiestni ruku do oblasti, v ktorej bude prebiehať proces rozpoznávania. V druhom móde bude kameru na ruky ukazujúceho mieriť niekto iný, napríklad zadnú kameru svojho smart telefónu.



**Obr. 2.** Metóda, pri ktorej si človek sám vloží ruku do oblasti v ktorej prebieha rozpoznávanie a metóda, pri ktorej na ruku namieri kameru niekto iný

Využitie metódy, pri ktorej si človek sám umiestni ruku do vopred definovanej oblasti, vyžaduje prítomnosť zobrazovacieho zariadenia (monitor, TV), na ktorom sa môže subjekt vidieť, a teda aj upraviť polohu, natočenie svojej ruky. Metóda, pri ktorej kameru na ruku človeka namieri niekto iný, vyžaduje prítomnosť ďalšieho človeka, ktorý bude kamerou sledovať pohyb ruky ukazujúceho jednotlivca. Táto metóda by mohla byť implementovaná napríklad do smart telefónov na zjednodušenie komunikácie s nepočujúcimi. Výber snímky, ktorá má byť rozpoznávaná, môže byť vykonávaný priamo užívateľom, alebo algoritmom na detekciu kľúčových snímok [AFC01] .



## 4.2 Predspracovanie obrazov

Na napodobenie fungovania ľudského zraku je vhodné aproximovať proces, ktorý vykonáva ľudský neuro-vizuálny systém. Ten priradzuje videným objektom hodnotu a na základe toho ich rozpoznáva. Objekty, ktorých hodnota je dostatočne nízka, sú odignorované. Ľudský systém hodnôt je veľmi komplexný a plynulo sa v priebehu času mení. Tejto téme sa značne venovali Sigmund Freud, Carl Gustav Jung, Jean Piaget a Jordan Peterson. Neexistuje kompletný model ľudského vnímania hodnôt a v súčasnosti sa ani nedá namodelovať. Vieme ale, ako má v niektorých špecifických prípadoch vyzerat' jeho vplyv. V prípade dvoch ľudí komunikujúcich prstovou abecedou (ak si odmyslíme všetky možné stimuly, ktoré by mohli mať vyššiu hodnotu, ako sú prudké hlasné zvuky, ostré blikajúce svetlá, objekty rútiace sa ich smerom, ľudské pudy...) budú títo jedinci reálne vidieť iba znaky, ktoré ten druhý ukazuje a okolitý svet z veľkej miery nevnímajú.

Uvedený proces môžeme nasimulovať vymazaním všetkej informácie, ktorá by bola považovaná za nepodstatnú, z obrazu. Tým eliminujeme vplyv objektov, ktoré sa nachádzajú v pozadí alebo nie sú relevantné a nebudú vplývať na proces klasifikácie.

Na snímací čip fotoaparátu/kamery dopadá všetko svetlo z oblasti, ktorá je zameraná. Na elimináciu ostatných objektov bol zvolený adaptívny farebný filter. Adaptívny farebný filter deteguje objekty farby kože a eliminuje objekty ostatných farieb z obrazu.



**Obr. 3.** Vstupný obraz ruky

Zo vstupného obrazu ruky sa vyrobí binárna, čierno-biela mapa. Tá bude mať na miestach, kde sú farby kože, hodnotu 1 a na miestach, kde sú všetky ostatné farby, hodnotu 0. Rozsah farieb kože bol určený experimentálne ako priemer hodnôt zo stredu obrazu (berie sa do úvahy, že kamera je na ruku nejako zameraná)  $\pm 35$  z rozsahu 0 až 255 pre každú farbu. Výsledok takejto filtrácie je mapa na obr. 4.



**Obr. 4.** Mapa po farebnej filtrácii a následnom morfológickom spracovaní

Následne sa vykoná operácia morfológické uzavretie, čím sa uzatvoria všetky malé biele objekty. Potom sa začnú morfológicky vyplňať biele objekty od najmenšieho po najväčší čiernou farbou, až kým nezostane iba jeden. Uvedený postup spôsobí, že vo výslednom obraze zostane len jeden objekt farby ruky a za predpokladu, že kamera je zameraná práve na ruku, bude týmto objektom ruka. Pôvodná mapa a mapa po týchto úpravách je zobrazená na obr. 4. Keď je mapa predspracovaná, použije sa na vyoknovanie pôvodnej snímky operáciou násobenia bod po bode s pôvodným vstupným obrazom. Tým vznikne obraz ruky bez pozadia.

Posledným krokom v predspracovaní je zmenšenie obrazu na najmenší možný štvorec orezávaním nulových bodov. To zaručí, že pre ďalšie spracovanie budú všetky fotky, s ktorými budeme pracovať, štvorcovej veľkosti a ruka bude tvoriť ich majoritnú časť.



**Obr. 5.** Orezaný obraz po farebnej filtrácii

## 4.3 Extrakcia charakteristických príznakov

### 4.3.1 Filter Laplacian of Gaussian

Na extrakciu charakteristických príznakov bol zvolený hranový filter LoG (Laplacian of Gaussian – Laplaceov operátor z Gaussovej funkcie). Tento diferenciálny operátor je definovaný ako divergencia gradientu daného skalárneho alebo tenzorového poľa.

$$\Delta = \nabla \cdot \nabla = \text{divergencia gradientu} . \quad (1)$$

Laplaceov operátor sa označuje ako  $\Delta$  a je invariantný voči zámene súradníc. V  $n$  rozmernom priestore je definovaný ako [73]:

$$\Delta = \frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2} + \dots + \frac{\delta^2}{\delta n^2} . \quad (2)$$

Dvojrozmerná Gaussova funkcia so strednou hodnotou  $\mu = 0$  je definovaná ako [73]:

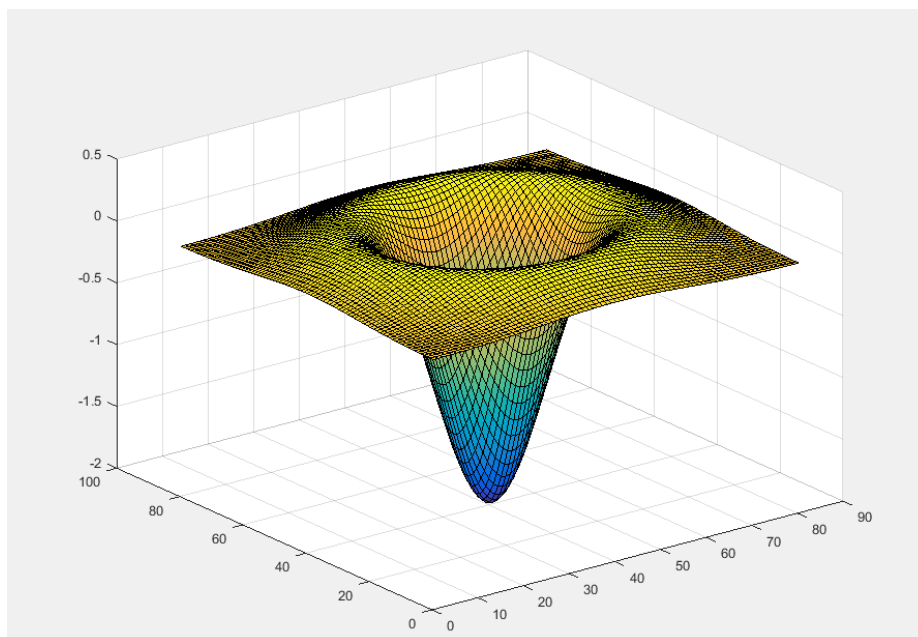
$$G(x, z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) . \quad (3)$$

Po aplikácii Laplaceovho diskrétného operátora na Gaussovú funkciu dostaneme LoG danú ako [73]:

$$LoG = \Delta G(x, z) = \frac{\delta^2 G(x, z)}{\delta x^2} + \frac{\delta^2 G(x, z)}{\delta y^2} . \quad (4)$$

$$LoG = -\frac{1}{\pi\sigma^4} \left(1 - \frac{x^2+y^2}{2\sigma^2}\right) \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) . \quad (5)$$

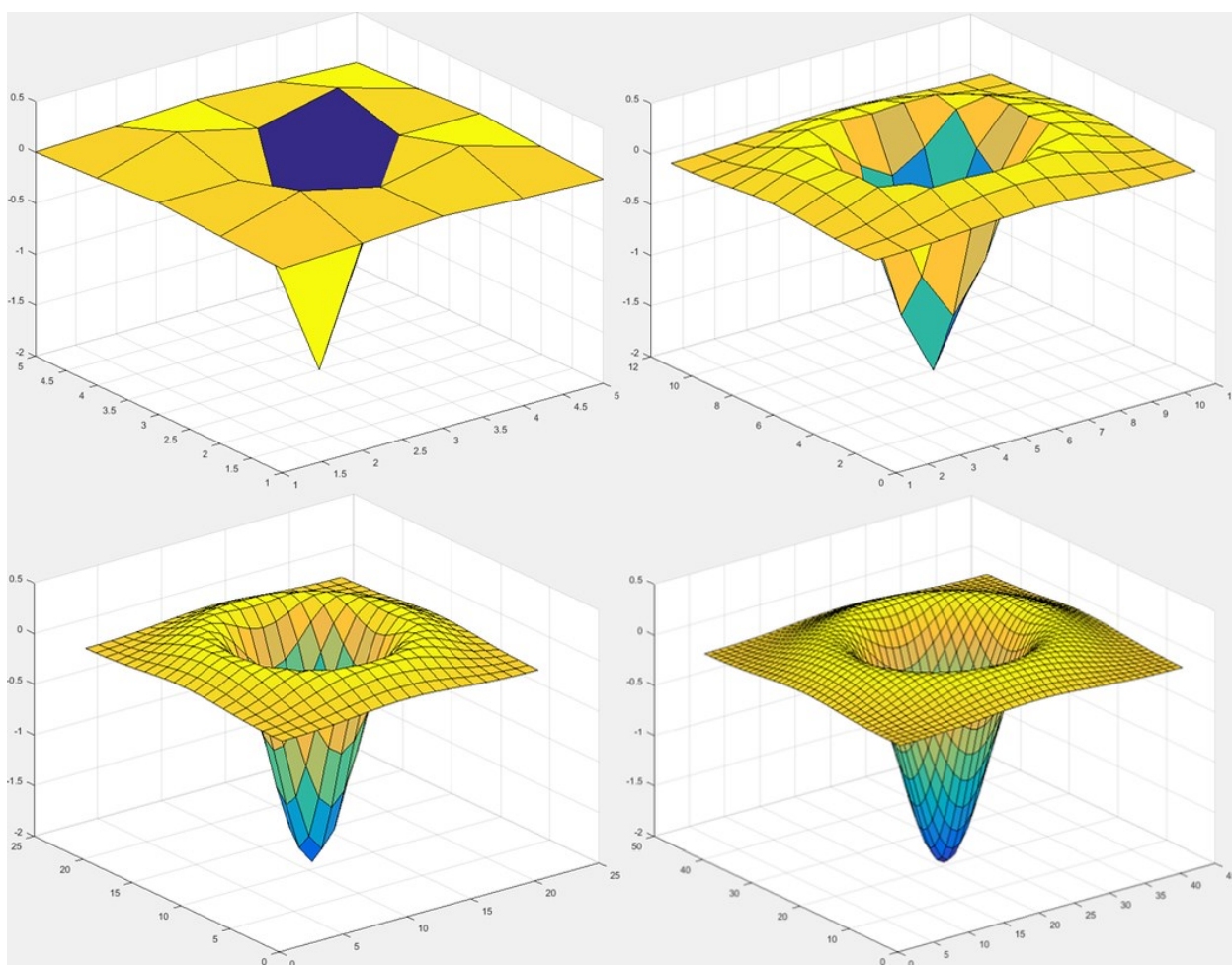
Vizualizácia LoG filtra je znázornená na nasledujúcom obrázku.



Obr. 6. 2D LoG filter

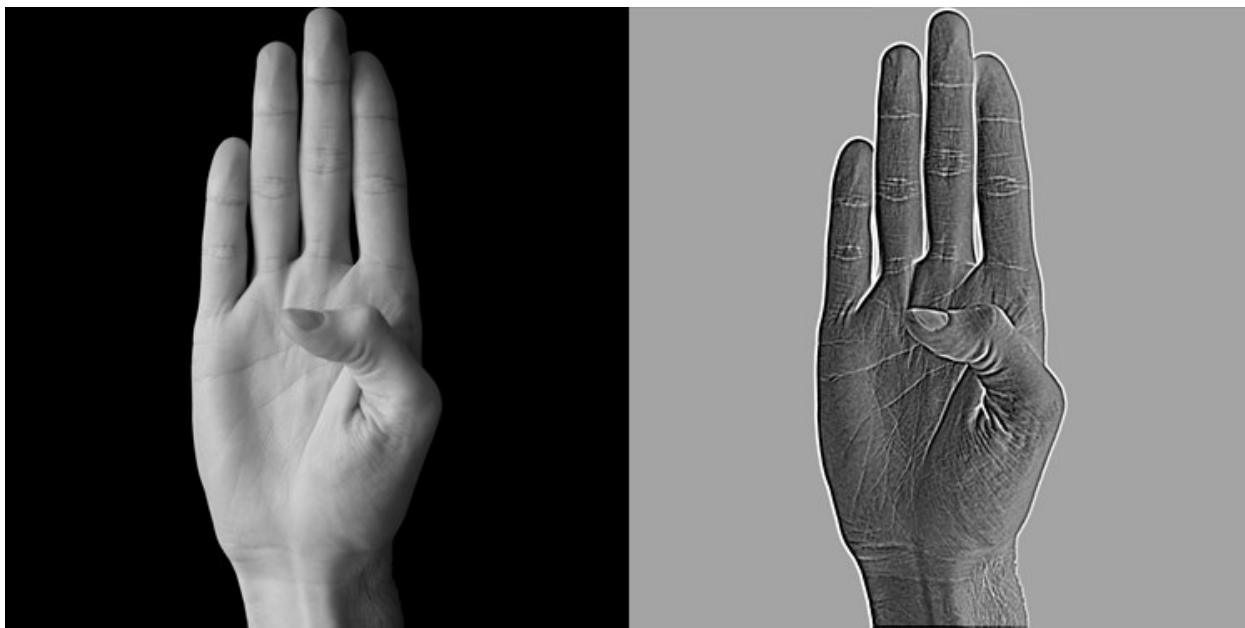
LoG je podobný Marrovmu waveletu (mexický klobúk), ale je invertovaný. Ide o multidimenzionálne zovšeobecnenie tohto waveletu. LoG je izotropický operátor (aplikuje sa s rovnakou intenzitou vo všetkých smeroch). Po aplikácii na obraz zvýrazňuje regióny s prudkou zmenou intenzity. V praxi sa používa hlavne na detekciu hrán a obrysov. Ako filter plní funkciu pásmovej priepusti, ale na rozdiel od iných hranových filtrov (ako napríklad Sobelov, Prewittov, Cannyho...) nepotláča úplne všetky nízkofrekvenčné zložky. LoG filter zachováva aj zhluky.

Získaná LoG funkcia zo vzorca 5 bola rovnomerne navzorkovaná v intervale  $-3 \leq x \leq 3$  a  $-3 \leq y \leq 3$  tak, aby sme získali matice, ktoré môžeme použiť na konvolvovanie so vstupnými obrazmi. Experimentálne boli určené kroky vzorkovania na 1.5, 0.6, 0.3 a 0.15. Pri tomto vzorkovaní boli vygenerované matice veľkostí 5x5, 11x11, 21x21 a 41x41. Matica 5x5 je určená na konvolvovanie s obrazmi s malým rozlíšením (približne 512x512 až 1024x1024), na ktorých majú hrany hrúbku len jeden pixel, matica 11x11 na obrazy, v ktorých sa nachádzajú hrany s hrúbkami 3 pixely (zväčša obrazy s rozlíšením 1024x1024 až 2048x2048), matica 21x21 na obrazy s hranami hrúbky približne 6 pixelov (rozlíšenia 2048x2048 až 4036x4036) a matica 41x41 na obrazy s hranami hrúbky 12 a viac pixelov (tými boli buď obrazy s rozlíšením 4036x4036 a väčšie alebo obrazy s nie dokonale zaostrenými hranami). Výhodou takto zvoleného filtra je, že je plne škálovateľný a vie byť prispôbený rozmeru vstupných obrazov, s ktorými ho konvolvujeme. Na rozdiel od bežných filtrov používaných v číslicovom spracovaní obrazov, tieto filtre nemajú sumu matice 1. Neslúžia na úpravu obrazu, na zhodnocovanie človekom a je nám jedno, že dynamicky menia rozsah jasu.



**Obr. 7.** LoG filter 5x5, 11x11, 21x21, 41x41

Sú dva spôsoby, akými môžu byť tieto filtre aplikované na vstupné obrázky. Prvý je na farebné v pôvodnej farebnej oblasti a druhý je na upravené obrázky (do iných farebných priestorov, napríklad šedotónové). Vo finálnej aplikácii boli LoG filtre použité na šedotónové obrázky.

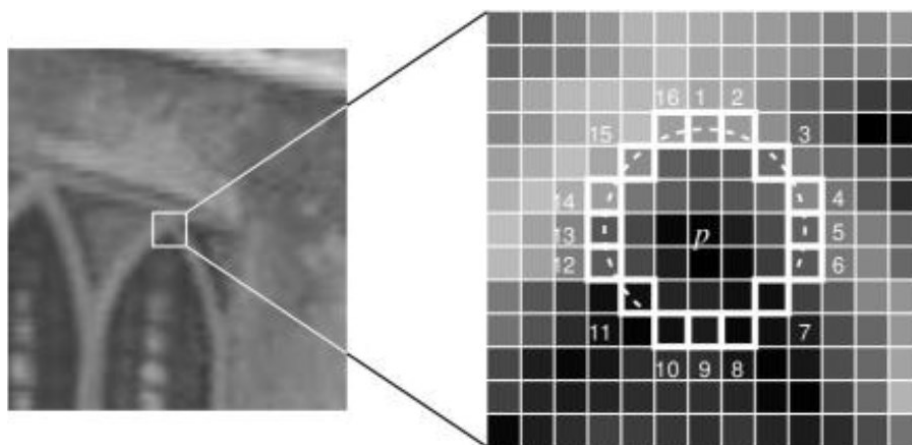


**Obr. 8.** Obrázok po aplikácii LoG filtra

#### 4.3.2 Algoritmus hľadania lokálnych extrémov

Extrakcia príznakov je následne založená na hľadaní lokálnych extrémov v LoG vyfiltrovaných obrazoch. Lokálne extrémny predstavujú prudké zmeny jasnosti spôsobené prítomnosťou hrán a iných vysokofrekvenčných črt na ruke.

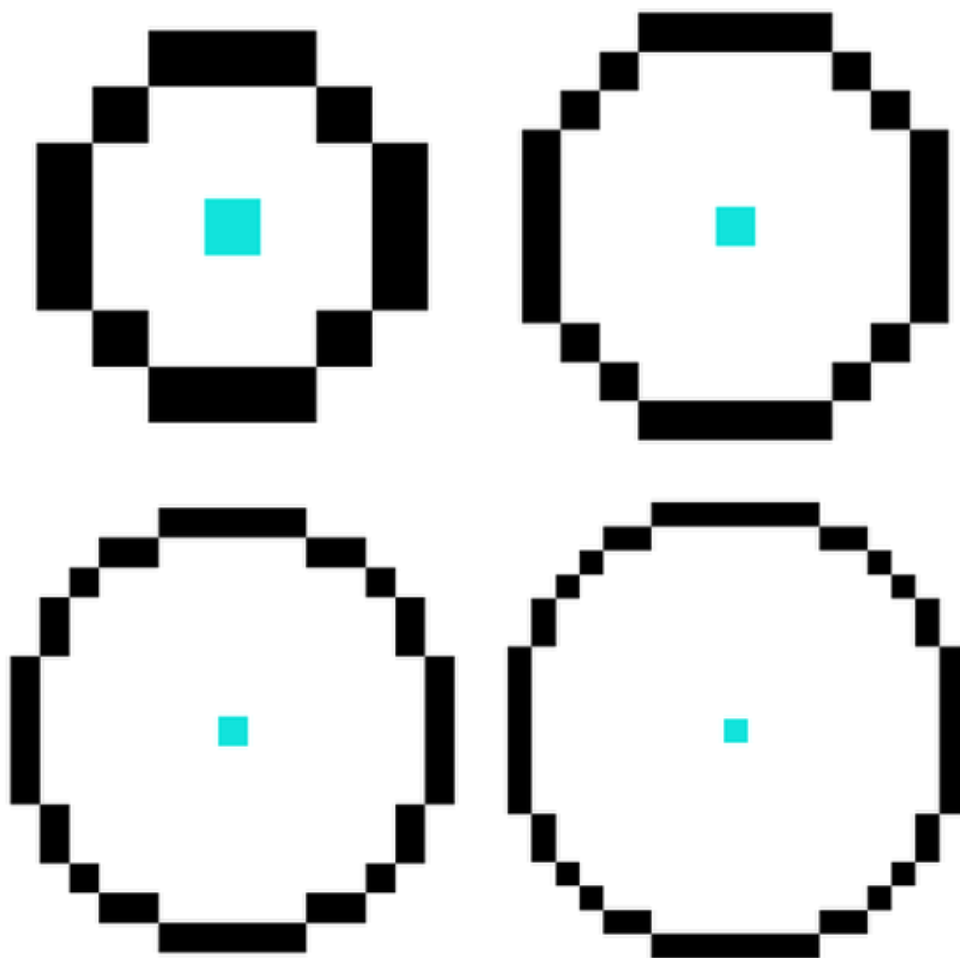
Na hľadanie lokálnych extrémov bol zvolený algoritmus podobný algoritmu FAST [65], ktorý porovnáva každý bod s jeho okolím (s niekoľkými úpravami). Pôvodný algoritmus FAST pre každý analyzovaný bod našiel jeho okolie v podobe Bresenhamovho kruhu veľkosti 3.



**Obr. 9.** Pôvodný FAST algoritmus

Body sa očísľujú od 1 po 16 ( $I_1, I_2, \dots, I_{16}$ ). Ak existuje  $N$  spojitých pixelov, pre ktoré platí, že hodnota stredového pixelu  $P$  je väčšia ako  $I_x + \beta$  alebo menšia ako  $I_x - \beta$ , tak môžeme prehlásiť, že daný stredový bod je hrana. Hodnoty  $\beta$  a  $N$  sa určujú experimentálne v závislosti od typu hrany, ktorú chceme detegovať. Pre detekciu rohov sa volí  $N \geq 12$ .  $\beta$  sa väčšinou určuje podľa dynamického rozsahu jasu v analyzovanom obraze.

Výhodou opísaného algoritmu je extrémna rýchlosť v porovnaní s inými algoritmami na extrakciu príznakov. Autori uvádzajú rýchlosť spracovania  $200\mu s$  na snímku s rozlíšením  $640 \times 480$  pixelov. To bolo realizované v roku 2006 na jednojadrovom počítači. Jeho značnou nevýhodou je zlá škálovateľnosť. Pri obrazoch s väčším rozlíšením algoritmus zlyháva z dôvodu dlhších prechodov medzi hranami. Aby bolo možné tento algoritmus použiť na obrazy s väčším rozlíšením, musí byť upravený. V našom finálnom algoritme bol na nájdenie lokálnych extrémov použitý pôvodný algoritmus FAST rozšírený o kruhy väčších veľkostí. Na spracovanie obrazov  $512 \times 512$  až  $1024 \times 1024$  bol použitý pôvodný FAST algoritmus s pôvodným Bresenhamovým kruhom s polomerom 3. Na obrazy  $1024 \times 1024$  až  $2048 \times 2048$  bol zvolený Bresenhamov kruh s polomerom 5, pre obrazy  $2048 \times 2048$  až  $4096 \times 4096$  bol zvolený kruh s polomerom 7. Najväčší použitý Bresenhamov kruh bol s polomerom 9 a bol použitý na obrazy s rozlíšením nad  $4096 \times 4096$ . Algoritmus je do budúca škálovateľný v prípade zvyšujúceho sa rozlíšenia snímok.



**Obr. 10.** Ukážka predgenerovaných Bresenhamových kruhov okolo stredového pixelu s polomeri 3, 5, 7 a 9



V prípade, ak je daný bod uznaný ako lokálny extrém, je jeho poloha uložená v databáze. Po aplikovaní algoritmu na celý obraz uložíme získané príznaky a ich relatívnu polohu vzťahnutú na veľkosť daného obrazu. Na tréovanie a rozpoznávanie sa dané vektory budú ešte spracovávať. Vo vektore sú totiž príznaky usporiadané v poradí, ako boli zozbierané, čo je od ľavého horného bodu po riadkoch. Použiť ich v tejto podobe na rozpoznávanie by nebolo ideálne. Vložilo by to do daného procesu náhodnosť, ktorú by bol problém odstrániť. Dĺžka vektorov by taktiež nebola rovnaká.

Na klasifikáciu príznakov bola zvolená technológia konvolučných neurónových sietí. Tá vyžaduje, aby na jej vstupe boli dáta rovnakej veľkosti. Vektory príznakov preto interpolujeme do prázdnej nulovej matice a všade tam, kde sa nachádza príznak, sa vloží jednotka. Poloha jednotiek sa z vektora relatívnych polôh určí pre násobením veľkosťou nulovej (prázdnej) matice a zaokrúhlením. Experiment prebiehal s viacerými veľkosťami. Na nasledujúcom obrázku je zobrazené, ako dané príznaky vyzerajú po interpolovaní do matíc veľkosti 1024x1024.



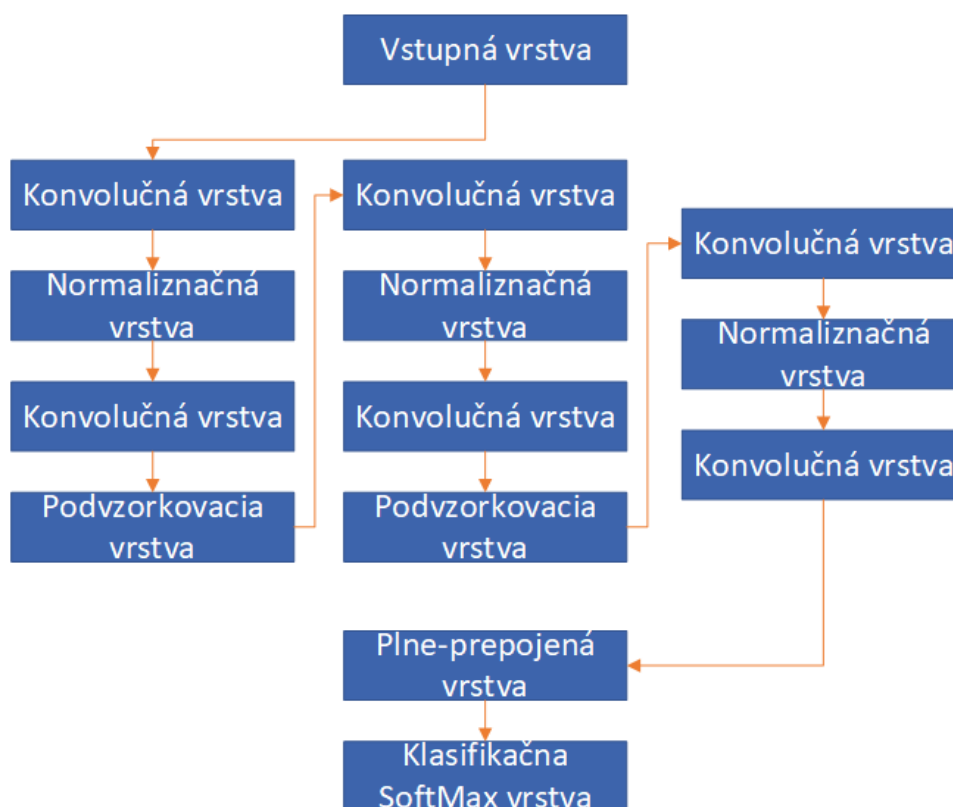
**Obr. 11.** Vektor príznakov znaku B interpolovaný do matice 1024x1024 (negatív)

## 4.4 Klasifikácia obrazov

Z už uvedených dôvodov je extrémne zložitá nájsť model, ktorý by dokázal úspešne klasifikovať znaky prstovej abecedy. Tvary prstov a rúk nie sú úplne presne determinované a medzi rôznymi ľuďmi dochádza k značným rozdielom. Aj pri ukázaní toho istého znaku dochádza k rozdielom, ako ich daný jednotlivec reprodukuje.

Je žiadúce, aby bol použitý algoritmus, ktorý bude konvergovať k nejakému suboptimálnemu modelu. Z tohto dôvodu bola použitá na klasifikáciu technológia CNN (konvolučných neurónových sietí). Tá je schopná pri zachovaní určitých predpokladov konvergovať k nejakému suboptimálnemu riešeniu daného problému. Štruktúra danej CNN bude opísaná v nasledujúcich podkapitolách.

### 4.4.1 Štruktúra neurónovej siete a jej tréning



Obr. 12. Výsledná štruktúra neurónovej siete

Neurónová sieť bola implementovaná v simulačnom programovacom prostredí Matlab. Prostredie Matlab bolo zvolené pre svoju modulárnosť, obrovské množstvo dostupných knižníc open-source programov a veľkej medzinárodnej komunite venujúcej sa vývoju na tejto platforme. Vzorový príklad, ako implementácia vyzerá, je uvedený v nasledujúcom pseudokóde:



```
#####
layers=[
    imageInputLayer([1024 1024 1])

    convolution2dLayer(7,8,'Padding',3)
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2,'Stride',2)

    convolution2dLayer(7,16,'Padding',3)
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2,'Stride',2)

    convolution2dLayer(7,32,'Padding',3)
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2,'Stride',2)

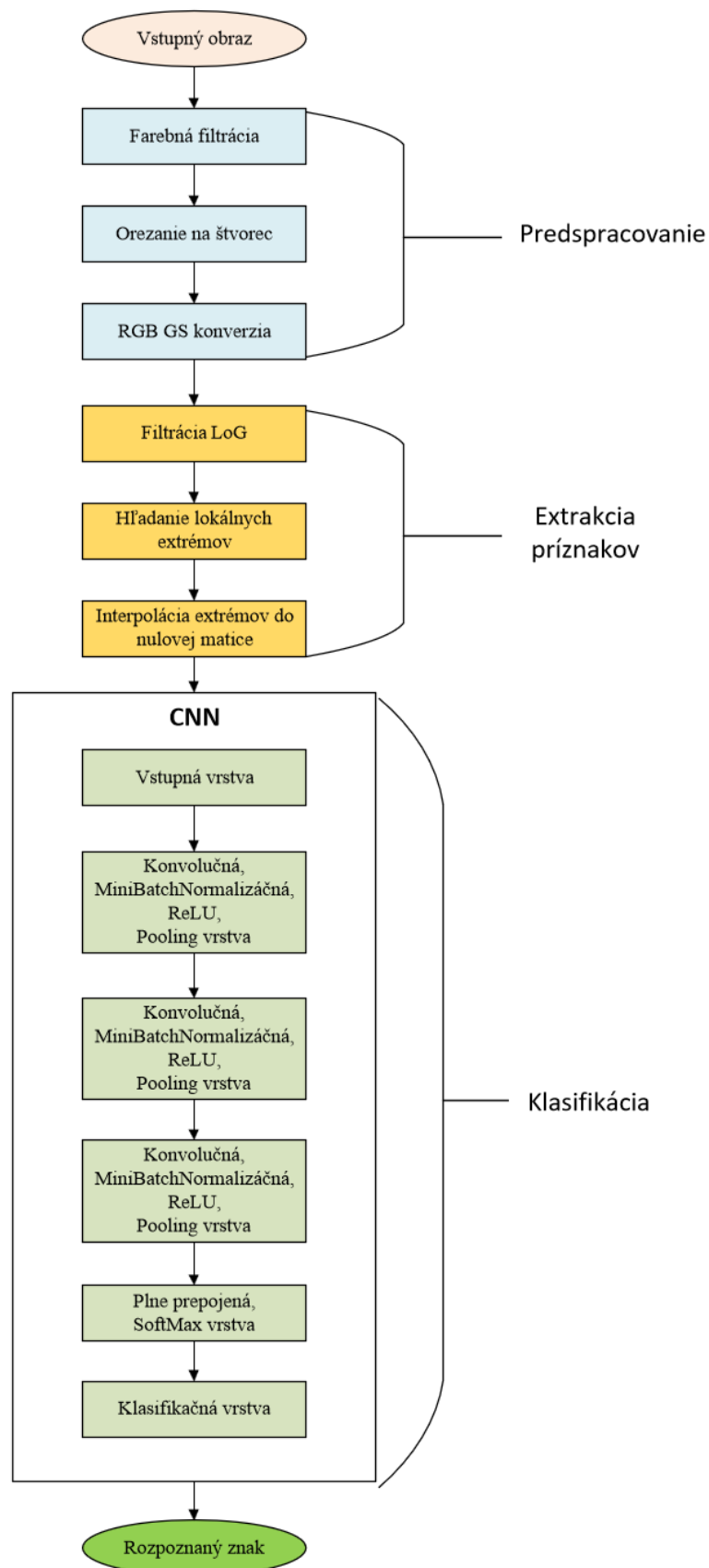
    convolution2dLayer(7,64,'Padding',3)
    batchNormalizationLayer
    reluLayer

    fullyConnectedLayer(26)
    softmaxLayer
    classificationLayer];
#####
```

Pričom jednotlivé riadky predstavujú implementáciu daných vrstiev v Matlabe. Následne musí byť celá sieť natrénovaná.

```
#####
options=trainingOptions('sgd',...
    'MaxEpochs',4,...
    'ValidationData',valDigitData,...
    'ValidationFrequency',1,...
    'ValidationPatience',Inf,...
    'Verbose',false,...
    'Plots','training-progress');
net =trainNetwork(trainDigitData,layers,options);
#####
```

Trénovanie neurónovej siete je založené na metóde klesajúceho gradientu a je detailne opísané v dizertačnej práci.

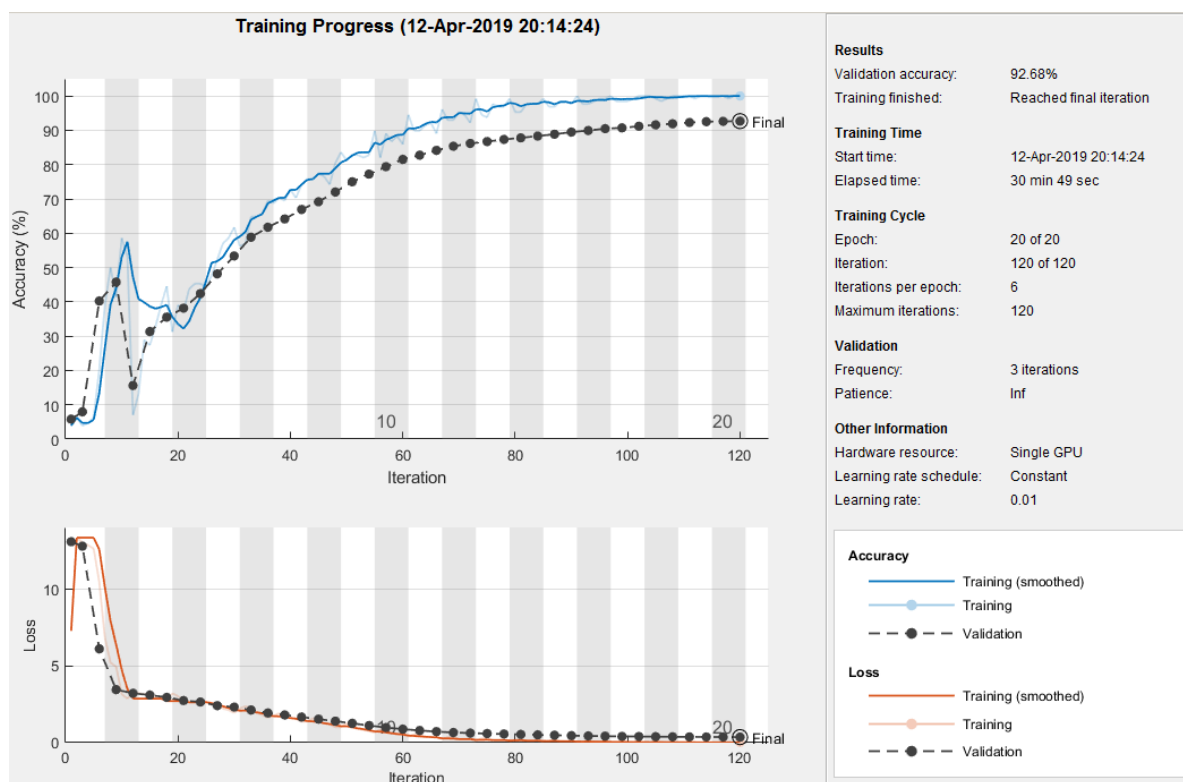


**Obr. 13.** Proces klasifikácie znakov

## 5 Výsledky a zhodnotenie

Na vyhodnotenie výsledkov musíme vedieť, s čím ich porovnávať, aby sme vedeli zhodnotiť ich objektivitu. Zobrali sa pôvodné snímky rúk ukazujúce znaky prstovej abecedy a boli použité na natréňovanie konvolučnej neurónovej siete. Z každého písmena pôvodnej prstovej abecedy sme mali 10 (od desiatich rôznych ľudí). Trénovanie prebiehalo do 26 klasifikačných tried (bez mäkkčov a dĺžňov). Každá z nich sa trénovala 9 znakmi a jeden sa použil na rozpoznávanie. Neurónová sieť dané znaky neklasifikovala do tried náhodne, ale na niečo sa natrénovala. Pri náhodnej klasifikácii by bola úspešnosť približne 3,64% (100/26). Vo väčšine prípadov sa na tréningové dáta natrénovala s úspešnosťou nad 90% a v niektorých prípadoch aj na 100%.

Nasledujúci obrázok zobrazuje záchyt tréningu a klasifikácie, kde bol aplikovaný celý náš algoritmus. Bolo vysegmentované pozadie, prebehla LoG filtrácia, našli sa charakteristické príznaky, boli interpolované do štvorcovej matice 1024x1024 a nimi bola daná sieť tréňovaná. Taktiež bolo klasifikované do 26 znakov prstovej abecedy a každý znak má 3969 tréningových vzoriek (9\*441 rotovaných a deformovaných verzií) a 441 testovacích snímkov (detailný popis mechanizmu ktorým boli z pôvodných snímkov vytvorené rotované a deformované snímky na tréningovanie neurónovej siete je uvedený v dizertačnej práci).



**Obr. 14.** Screenshot s Matlabu po dokončení tréningu s novým navrhovaným postupom

Ako možno z dosiahnutých výsledkov na obrázku vidieť, úspešnosť rozpoznávania znakov bola v tomto prípade 92,68%. Naznačuje to veľmi pozitívny vplyv aplikovaného postupu na schopnosť neurónovej siete rozpoznávať a klasifikovať dané znaky. V tomto prípade sme obmedzili množstvo dát, na ktoré sa môže neurónová sieť natréňovať, a tým sme veľmi výrazne zvýšili úspešnosť klasifikácie. Navrhnutá technika je založená na podobných princípoch ako

nové v súčasnosti skúmané techniky OSL (One Shot Learning). Priemerná úspešnosť bola 93,27% a najlepšia dosiahnutá úspešnosť rozpoznania testovacích dát bola 96,79%.

Každý krok použitého navrhovaného postupu zvýšil úspešnosť klasifikácie. Priemerná úspešnosť rozoznávania pôvodných nespracovaných obrazov bola 21,82% (pre 10 obrazov 6000x4000 a CNN s kernelom 7x7 a hĺbkou 3 vrstvy). Prvý krok, vyrezanie ruky (vo väčšine prípadov približne 1000x1000 pixelov) vymaže približne 95% informácie z obrazu a na natrénovanie sa bude používať len jeho relevantná časť. Tým sa pri klasifikácii zvýšila priemerná úspešnosť o 19,34% na 41,16% (tento obrovský skok je spôsobený zrušením vplyvu viac ako 90% obrazu na klasifikáciu). Ďalší krok, farebná filtrácia, zanechá v štvorcovom výreze len objekt farby kože, teda ruku. Tento krok vymaže približne 60% pixelov a zvyšuje úspešnosť klasifikácie o priemerne 11,43% na 52,59%. Najväčší skok v úspešnosti klasifikácie ale spôsobilo vygenerovanie nových tréningových dát. Vytvorením z jednej tréningovej snímky 441 rôznych snímok (deformáciami a rotáciami) sa zvýšila úspešnosť rozpoznávania o 29,32% na 81,91%. Tieto kroky by sa dali nahradiť extrémnym zvýšením tréningovej databázy na tisíce rôznych snímok (v takom prípade by sa úspešnosť mohla blížiti 100 percentám). Posledným krokom bola extrakcia príznakov pomocou LoG filtra a klasifikácie len charakteristických príznakov. Táto operácia čiastočne dekorelovala redundantné dáta v obraze a zvýšila úspešnosť približne o 10% na približne 92%.

Samostatným problémom bola aplikácia na prstové znaky špecifické pre slovenskú znakovú reč. Medzinárodne používaná prstová abeceda ASL používa statické ukážania znakov (z výnimkou znakov 'j' a 'z', ktoré sa tiež dajú klasifikovať pomocou ich finálnych statických polôh). Znaky 'ä' a 'ch' majú svoje nové jedinečné znaky, ktoré mali pri natrénovaní klasifikačnú úspešnosť podobnú priemernej úspešnosti danej sady. Znaky s mäčkami a dĺžkami sa klasifikovali ako dva krát po sebe ukázaný ten istý znak. Ich priemerná úspešnosť znaku bez mäčka/dĺžky na druhú. Napríklad znak 'c' mal pri jednom natrénovaní úspešnosť 94.36% (0.9436) a znak 'č' približne 89.97% (približne  $0.9436 \cdot 0.9436$  teda pravdepodobnosť úspešnej detekcie 2 krát po sebe).

Množstvo korektných modelov, ku ktorým daná sieť môže skonvergovať pri tréningu (ktoré začne v náhodnom modeli), je nepredvídateľne veľké. Z týchto modelov budú mať niektoré lepšiu úspešnosť a iné výrazne horšiu. Každý bude ale aproximáciou minimalizovania stratovej funkcie a každý z nich môže aproximovať iné riešenie tejto minimalizácie. Taktiež môže každé z nich skončiť v inom lokálnom minime gradientu stratovej funkcie (a pravdepodobne aj skonči).

Bolo vykonané veľké množstvo meraní pre rôzne konfigurácie CNN a rôzne varianty navrhovaného algoritmu rozpoznávania prstovej abecedy. Tie sú všetky detailne opísané a vysvetlené v dizertačnej práci.

## 6 Záver

Témou tejto dizertačnej práce bolo rozpoznávanie znakov prstovej abecedy. Ide o špecifický problém v problematike strojového videnia, ktorá sa pokúša klasifikovať obraz tak, ako ľudský neuro-vizuálny systém.

Prvá časť dizertačnej práce sa zaoberá prehľadom aktuálneho stavu problematiky počítačového videnia a rozpoznávania prstovej abecedy. Analyzuje existujúce postupy a porovnáva praktickosť ich aplikácie. V tejto časti je aj čiastočne vysvetlená problematika klasifikácie obrazov tak, ako by to klasifikoval človek. Sú tu porovnané najčastejšie používané algoritmy strojového videnia s vysvetlením, prečo poskytujú také výsledky, aké majú.

Nasledujúce kapitoly boli praktické. Boli v nich opísané nami navrhované postupy, ktorými sme sa danú problematiku pokúšali riešiť, a ktoré aspoň čiastočne fungovali. Prvé dva postupy poskytovali čiastočné riešenie tejto problematiky, ale ani jedno samo o sebe neposkytovalo dostatočnú presnosť výsledkov. Prvý bol založený na porovnávaní šablón transformovaných vzdialenostnou transformáciou a druhý na porovnávaní obrysov ruky zakódovaných novým, nami navrhnutým 8 smerovým kódom. Výsledky oboch postupov boli prezentované a publikované na zahraničných medzinárodných vedeckých konferenciách.

Oba predošlé postupy mali vážne nedostatky. Aj keď mali znaky s veľmi veľkou úspešnosťou rozpoznávania, tak mali aj znaky s veľmi nízkou úspešnosťou rozpoznávania. Aj keby boli teoreticky tieto postupy použiteľné na inú aplikáciu (napríklad ovládanie gestami, len s použitím gest s dobrou úspešnosťou rozpoznávania), ich použitie na rozpoznávanie všetkých znakov prstovej abecedy nie je prakticky možné.

Nasledujúce opísané riešenie je finálnym riešením dizertačnej práce. Je založené na čiastočnej aproximácii ľudského vizuálneho systému a hlbokých konvolučných neurónových sieťach. Toto navrhované riešenie sa snaží napodobiť spracovanie obrazových vnemov v mozgu tak, aby jeho výsledok bol podobný tomu, čo nakoniec človek vidí (ľudský mozog podvedome filtruje veľkú časť informácií, ktoré považuje za málo významné). Teória, ktorá sa zaoberá rozdielom medzi obrazom, ktorý dopadne na sietnicu, a tým čo človek vníma ako „to, čo vidí“, nie je úplná. Ľudské poznatky v tejto problematike nie sú úplné a tie poznatky, ktoré máme, nie sú dostatočné na navrhnutie modelu spracovania vizuálnych podnetov ľudským mozgom.

To, čo ale vieme, je, že ľudský mozog sa dá aproximovať ako systém hodnôt. Objekty, ktoré ľudský mozog vyhodnotí ako v danej chvíli bezvýznamné, človek „nevidí“, a tie, ktoré vyhodnotí ako dostatočne významné, „vidí“.

Tento princíp je v našom algoritme aproximovaný farebnými filtrami. V skúmanom regióne obrazu sú zachované len objekty farby ľudskej kože alebo podobnej farby. Z nich sa vymažú všetky až na ten najväčší. Za predpokladu, že ukazujúci vložil ruku do skúmaného regiónu obrazu alebo že na jeho ruku niekto zameral kameru, bude tento objekt práve ruka. Táto časť algoritmu aproximuje to, že počítač bude „vidieť“ iba ruku. Následne bude obraz orezaný na najmenší možný opísaný štvorec na zachovanie veľkosti potrebnej pre neurónové siete.

V ďalšom kroku sa vo vysegmentovanom obraze ruky vyhľadávajú charakteristické príznaky. Tie boli nájdené ako lokálne extrémny (minimá a maximá) po vyfiltrovaní obrazu ruky filtrom Laplacian Gaussianu (LoG). Výsledné príznaky predstavujú polohu vnútorných a vonkajších charakteristických črt ruky. Človek je schopný len na základe týchto príznakov obraz ruky rozoznať, a teda v nich zostala charakteristická informácia potrebná na rozoznávanie jednotlivých obrazov. Tento krok je nutný na zníženie množstva informácie v obraze. Neurónová sieť je teoreticky schopná natréňovať sa na aproximovanie akéhokoľvek modelu, ale nevieme dopredu určiť, na aký model sa natrénuje (najmä ak množstvo dát vo vstupnom obraze je značne redundantné a korelované). Elimináciou redundantnej informácie zvyšujeme pravdepodobnosť skonvergovania ku korektnému modelu neurónovej siete pri tréningu (obzvlášť v prípadoch, keď nemáme dostupné obrovské tréningové databázy desiatok tisícov vzoriek každej klasifikačnej triedy, ako to v neurónových sieťach býva bežné).

Posledným krokom bola klasifikácia získaných príznakov pomocou konvolučnej neurónovej siete. Neurónová sieť bola tréňovaná maticami, kde príznaky boli interpolované ako jednotky do nulových matic, pričom poloha týchto jednotiek je určená relatívnou polohou príznakov v pôvodnom obraze. Na zväčšenie množstva tréningových a testovacích dát boli obrazy rúk rotované a deformované. Detailný opis celého postupu a postupu ako prebiehalo vytváranie týchto databáz, je uvedený v kapitole 6. Ide o postup čiastočne podobný najmodernejším skúmaným technológiám OSL (One Shot Learning), pri ktorých sa neurónová sieť dá úspešne natréňovať relatívne malým množstvom dát.

Výsledky klasifikácie sú komplexne zhodnotené v kapitole 7. Najlepší dosiahnutý výsledok bol 96.79% pre jedno konkrétne natréňovanie a priemerný výsledok 93.27% pre všetky natréňovania v danej konfigurácii.

Komplexné zhodnotenie úspešností pre rôzne natréňovania je uvedené v kapitole 7. V nej sú uvedené aj rôzne konfigurácie neurónovej siete, ktoré boli pri tréňovaní použité.

Náš finálny navrhovaný algoritmus sa ukázal byť použiteľný na rozpoznávanie prstovej abecedy. Je založený na princípoch podobných ľudskému videniu a na rozdiel od bežných neurónových sietí je schopný skonvergovať ku správne klasifikačnému modelu už pri relatívne malej tréningovej databáze. Aj pri jeho najzložitejšej a najpomalšej implementácii je náš algoritmus schopný klasifikovať viac ako 3 snímky za sekundu. To viac ako postačuje na klasifikovanie znakov vo videu v reálnom čase. Maximálna priemerná doba klasifikácie na znak bola 297.451ms pri vstupnom obraze veľkosti 6000x4000 a z neho príznaky interpolované do matice 1024x1024. Tento výsledok bol dosiahnutý v programe Matlab. V prípade optimalizovanej implementácie v natívnom programovacom jazyku by sa rýchlosť klasifikácie výrazne zvýšila.

Nami navrhovaný algoritmus by mohol byť použitý aj na klasifikáciu iných obrazov, ktoré sa dajú jednoznačne charakterizovať svojimi vonkajšími a vnútornými črtami.

## 7 **Resumé**

This dissertation thesis focuses on the problem of finger spelling sign language's visual recognition. It is a specific problem in the field of computer vision, where images need to be classified by a computer with the same result as by a real human.

The first part of this dissertation thesis analyzes the methodology used in computer vision and describes state-of-the-art techniques and algorithms for sign language recognition. Most commonly used algorithms are compared from the practical point of view, as well as the basic principles of human vision from the neuropsychological point of view. This chapter also describes why some algorithms of computer vision work better than others.

Following chapters of dissertation thesis are practical. They describe the proposed and implemented methods, which yielded at least partially acceptable results. First two proposed methods offered partial solution of the problem at hand, but neither of them provided satisfactory results on its own. The first one is based on comparing the distance transformed images and the second one on comparing the images encoded by the novel proposed complex 8 directional code. Both methods were published at respected international peer-reviewed conferences.

Both previous methods had serious limitations - the character recognition success rate varied widely for different characters – some were correctly classified with a high success rates, but other ended up classified incorrectly more than others. Both could be theoretically used for different applications (such as device remote control using gestures with good recognition rate). However, they can't be used for recognition of the entire alphabet.

The method described in the next chapter is the final proposed solution of this dissertation thesis. It is based on partial approximation of human neurovisual system and deep convolutional neural networks. The proposed method attempts to approximate the human brain processing of visual stimuli in such a way, that the result of this process shall be similar to what a human finally sees (the human brain automatically filters information considered unimportant). Human knowledge about the difference between the light hitting retina and the final representation of what a "human sees" is incomplete. The theory of visual stimuli processing in a human brain is not exact enough for us to make a deterministic model of how the human neurovisual system works.

What we do know is that human brain can be approximated as a value system. Objects evaluated in a certain situation as worthless are unperceived even when the eyes pass over them. Objects considered valuable are perceived ("seen").

This principle is implemented in our algorithm using color filters. All objects in an analyzed image are preserved only if the color present in the image resembles that of a skin color. All other pixels are zeroed out. The remaining objects of skin color are morphologically closed (deleted) from smallest to largest, until only one remains. If there is a hand present in the analyzed area, the only remaining object will be the hand. The result of these operations is that the hand will be the only object analyzed in the image, which translates as the only object "seen". The image is then cropped to the smallest possible square (it is necessary to have the size ratio of images unified for the CNN).

Local characteristic features are found in the next step. They are identified as local extrema (local minima and maxima) of the analyzed image convolved with the Laplacian of Gaussian filter (LoG). The resulting features represent location of inner and outer characteristic hand contours. A human is still able to recognize images of different letters after this operation, which means that the information necessary to successfully classify them is still present. This step is necessary to remove the redundant information in the image and to remove unnecessary information. A neural network is theoretically capable of being trained to classify any information and approximate any model, but we cannot predict what model it will converge to during training, especially if the training data is redundant and highly correlated. By removing the redundant data, we increase the likelihood that the network will converge to a desired model based on the training data (particularly when no large dataset is available).

The last step of the proposed method is classification of obtained features using convolutional neural network. The CNN was trained using square matrices, where features were interpolated as ones (into zero matrices), where the position of ones was chosen as relative position of the feature in the original image, regardless of the image resolution. To increase the size of the training data set, training character images were rotated and deformed. A detailed description of this process can be found in chapter 6 of the dissertation thesis. This process is in principal, partly similar to the currently researched OSL (One Shoot Learning) technology, which can be used to train neural networks with relatively small training databases.

The results of classification are analyzed in detail in chapter 7. The best reached classification rate was 96.79% for a single training and the best average result was 93.27% for all trainings with a specific configuration.

The final proposed algorithm proved itself to be acceptable for sign language alphabet classification. It is based on principles akin to human vision, and unlike only using conventional neural networks, it is capable of converging into correct classification model with relatively small training data set. Even in its slowest and most complex implementation it was able to classify more than 3 signs per second – more than sufficient rate to classify signs in real time in video sequences. The largest average time of alphabet classification was 297.451ms for images with resolution 6000x4000 whose features were interpolated as 1024x1024 matrices. This result was obtained using MATLAB. The classification rate could be drastically increased by an implementation optimized in a low-level programming language.

The proposed novel algorithm could also be used to classify any objects that can be described by their inner and outer contours.



## 8 Prínosy dizertačnej práce

Hlavným cieľom tejto práce bol návrh algoritmu na rozpoznávanie sekvencií znakov prstovej abecedy.

Na základe uvedených faktov pokladám za pôvodné prínosy tejto dizertačnej práce a svojej vedeckej činnosti nasledovné:

1. Návrh algoritmu na rozpoznávanie znakov prstovej abecedy založený na Gáborových waveletoch a chamferovej vzdialenosti.  
(Publikované v [AFD04])
2. Algoritmus na zakódovanie tvaru obrazu pomocou nového, komplexného 8 smerového kódu, a následné rozpoznávanie pomocou Pearsonovho korelačného koeficientu.  
(Publikované v [AFC02])
3. Nový postup na aproximáciu ľudského neurovizuálneho systému, založený na adaptívnej extrakcii príznakov a hlbokých konvolučných neurónových sieťach.  
(Publikované v [ADM01])
4. Úprava algoritmu na hľadanie kľúčových snímok v kanáloch s maskovanými chybami.  
(Publikované v [AFC01] a [AFD07])

## 9 Vlastná publikačná činnosť

### 9.1 *Publikácie týkajúce sa dizertačnej práce*

**ADM Vedecké práce v zahraničných časopisoch registrovaných v databázach Web of Science alebo SCOPUS**

ADM01 CSÓKA, Filip [70%] - POLEC, Jaroslav [10%] - CSÓKA, Tibor [10%] - KAČUR, Juraj [10%]. Recognition of Sign Language from High Resolution Images Using Adaptive Feature Extraction and Classification. In INTL Journal Of Electronics And Telecommunications, 2019, VOL. 65, NO.2 , PP. 309-314 Manuscript received January 14, 2019; revised April, 2019. (Článok prijatý, očakávaný dátum vydania: Jún 2019) In: Master Journal List, **Kateg. B**

**AFC Publikované príspevky na zahraničných vedeckých konferenciách**

AFC01 CSÓKA, Filip [45 %] - POLEC, Jaroslav [45 %] - VARGIC, Radoslav [10 %]. Key frame extraction from video sequences containing ASL signs with concealed transmission errors. In CoDIT 2017 : 4th International conference on control, decision and information technologies. Barcelona, Spain. April 5-7, 2017. Danvers : IEEE, 2017, S. 208-213. ISBN 978-1-5090-6466-3. V databáze: IEEE ; WOS: 000450826500036. **Kateg. B**

AFC02 CSÓKA, Filip [60 %] - POLEC, Jaroslav [40 %]. Algorithm for real-time finger spelling alphabet recognition in video sequences. In SPSympo 2017 : Signal processing symposium. Jachranka, Poland. September 12-14, 2017. Piscataway : IEEE, 2017, S. 271-274. ISBN 978-1-5090-6755-8. V databáze: IEEE ; WOS: 000427086800052. **Kateg. B**

**AFD Publikované príspevky na domácich vedeckých konferenciách**

AFD03 CSÓKA, Filip [50 %] - PUTERKA, Boris [30 %] - POLEC, Jaroslav [20 %]. Real-time recognition of Slovak sign language. In ELITECH'17 [elektronický zdroj] : 19th Conference of doctoral students. Bratislava, Slovakia. May 24, 2017. 1. ed. Bratislava : Spektrum STU, 2017, CD-ROM, [5] p. ISBN 978-80-227-4686-1.

AFD04 CSÓKA, Filip [50 %] - PUTERKA, Boris [20 %] - POLEC, Jaroslav [30 %]. SSL finger spelling recognition using Gabor wavelet and chamfer distance. In Redžúr 2017 : 11th International workshop on multimedia information and communication technologies. Bratislava, Slovakia. May 19, 2017. 1. vyd. Bratislava : Vydavateľstvo Spektrum STU, 2017, S. 11-14. ISBN 978-80-227-4691-5. **Kateg. C**

AFD05 CSÓKA, Filip [50 %] - POLEC, Jaroslav [50 %]. Algorithm for Real-time Feature Extraction and Recognition of ASL Using Laplacian of Gaussian and Convolutional Neural Networks. In ELITECH'18 [elektronický zdroj] : 20th Conference of doctoral students. Bratislava, Slovakia. May 23, 2018. 1. ed. Bratislava : Vydavateľstvo Spektrum STU, 2018, CD-ROM, [5] p. ISBN 978-80-227-4794-3.

AFD07 HERIBANOVÁ, Petra [20 %] - CSÓKA, Filip [40 %] - POLEC, Jaroslav [40 %]. Key frame extraction for fingerspelling alphabets recognition using DCT. In Redžúr 2016

[elektronický zdroj] : 10th International workshop on multimedia and signal processing. Bratislava, Slovakia. May 24, 2016. 1. ed. Bratislava : Nakladateľstvo STU, 2016, CD-ROM, S. 39-43. ISBN 978-80-227-4560-4. **Kateg. C**

## 9.2 *Ostatné publikácie*

### **AAB Vedecké monografie vydané v domácich vydavateľstvách**

AAB01 BAROŇÁK, Ivan [30 %] - CHAMRAZ, Filip [10 %] - CSÓKA, Filip [30 %] - ZAFČÍK, Ján [10 %] - ŠÍP, Tomáš [10 %] - HARTMANN, Matej [10 %]. IMS technológie. 1. vyd. Bratislava : Spektrum STU, 2017. 307 s. Edícia monografií. ISBN 978-80-227-4738-7.

### **ADN Vedecké práce v domácich časopisoch registrovaných v databázach Web of Science alebo SCOPUS**

ADN01 CSÓKA, Tibor [30 %] - POLEC, Jaroslav [30 %] - CSÓKA, Filip [30 %] - KOTULIAKOVÁ, Kvetoslava [10 %]. VQ-based model for binary error process. In Journal of Electrical Engineering. Vol. 68, No. 3 (2017), s. 167-179. ISSN 1335-3632 (2017: 0.508 - IF, 4 - JCR Best Q, 0.205 - SJR, Q3 - SJR Best Q). V databáze: WOS: 000404615900001 ; SCOPUS: 2-s2.0-85028028933. **Kateg. A**

### **AED Vedecké práce v domácich recenzovaných vedeckých zborníkoch, monografiách**

AED01 CSÓKA, Filip [30 %] - BAROŇÁK, Ivan [70 %]. Smerovanie SIP správ pomocou SIP servera Kamailio. In Ročenka 2014. Fórum pre komunikačné technológie. Banská Bystrica : DALI-BB, 2015, S. 65-74. ISBN 978-80-8141-084-0.

### **AFC Publikované príspevky na zahraničných vedeckých konferenciách**

AFC03 CSÓKA, Filip [40 %] - POLEC, Jaroslav [60 %]. Superpixels merging in corrupted color images. In SPSympo 2017 : Signal processing symposium. Jachranka, Poland. September 12-14, 2017. Piscataway : IEEE, 2017, S. 280-283. ISBN 978-1-5090-6755-8. V databáze: IEEE ; WOS: 000427086800054. **Kateg. B**

AFC04 CSÓKA, Tibor [30 %] - POLEC, Jaroslav [30 %] - CSÓKA, Filip [30 %] - KOTULIAKOVÁ, Kvetoslava [10 %]. Classification-based VQ model for simulation of binary error process on the wireless channel. In MCSI 2016 : Third international conference on mathematics and computers in science and industry. Chania, Greece. 27-29 August 2016. Piscataway : IEEE, 2016, S. 77-84. ISBN 978-1-5090-0973-2. V databáze: WOS: 000404434600014. **Kateg. A**

AFC05 KOTULIAKOVÁ, Kvetoslava [50 %] - POLEC, Jaroslav [30 %] - CSÓKA, Filip [20 %]. An adaptive ARQ-HARQ method with BCH codes. In AICT 2017 : 11th IEEE International conference on application of information and communication technologies. Moscow, Russia. September 20-22, 2017. Piscataway : IEEE, 2017, S. 409-413. ISBN 978-1-5386-0500-4. **Kateg. B**

AFC06 POLEC, Jaroslav [15 %] - VARGIC, Radoslav [15 %] - CSÓKA, Filip [10 %] - BIELIKOVÁ, Mária [10 %] - SVRČEK, Martin [10 %] - MÓRO, Róbert [10 %] -

SMOLEJOVÁ, Eva [20 %] - HERETIK, Anton [10 %]. Detection of schizophrenia spectrum disorders using saliency maps. In AICT 2017 : 11th IEEE International conference on application of information and communication technologies. Moscow, Russia. September 20-22, 2017. Piscataway : IEEE, 2017, S. 398-402. ISBN 978-1-5386-0500-4. **Kateg. B**

AFC07 VARGIC, Radoslav [45 %] - KAČUR, Juraj [45 %] - CSÓKA, Filip [10 %]. Enhanced hierarchical mask creation for image coding using saliency maps. In CoDIT 2017 : 4th International conference on control, decision and information technologies. Barcelona, Spain. April 5-7, 2017. Danvers : IEEE, 2017, S. 288-293. ISBN 978-1-5090-6466-3. V databáze: IEEE ; WOS: 000450826500050. **Kateg. B**

AFC08 KAČUR, Juraj [35 %] - POLEC, Jaroslav [35 %] - CSÓKA, Filip [10 %] - SMOLEJOVÁ, Eva [20 %]. GMM Based Detection of Schizophrenia Using Eye Tracking. In IEEE Conference on Computational Intelligence and its application to problems in biology. CIBCB2019. (Článok prijatý, Dátum konferencie: Júl 2019).

#### **AFD Publikované príspevky na domácich vedeckých konferenciách**

AFD01 CSÓKA, Filip [50 %] - BAROŇÁK, Ivan [50 %]. Konfigurácia smerovača Kamailio ako SIP servera v sieti LAN. In ŠVOČ 2014 [elektronický zdroj] : Zborník vybraných prác 2014, Bratislava, 29. apríl 2014. 1.vyd. Bratislava : FEI STU, 2014, CD-ROM, s. 290-294. ISBN 978-80-227-5154-5.

AFD02 CSÓKA, Filip [50 %] - BAROŇÁK, Ivan [50 %]. Spracovanie prichádzajúcich SIP správ a ich následné smerovanie pomocou servera Kamailio. In EE časopis pre elektrotechniku, elektroenergetiku, informačné a komunikačné technológie [elektronický zdroj] : Zborník ku konferencii s medzinárodnou účasťou Elektrotechnika, Informatika a Telekomunikácie ELOSYS 2014. Roč. 20, mimoriadne číslo (2014), CD-ROM, s. 9-13. ISSN 1335-2547.

AFD06 CSÓKA, Tibor [70 %] - POLEC, Jaroslav [10 %] - CSÓKA, Filip [10 %] - ILČÍKOVÁ, Ivana [10 %]. SDR based capture of physical and data-link layer error process characteristics in low-rate WSN. In ELITECH'16 [elektronický zdroj] : 18th Conference of doctoral students. Bratislava, Slovakia. June 8, 2016. 1. vyd. Bratislava : STU, 2016, CD-ROM, [5] s. ISBN 978-80-227-4561-1.

AFD08 SZABÓ, Vladimír [10 %] - POLEC, Jaroslav [50 %] - CSÓKA, Filip [20 %] - KOTULIAKOVÁ, Kvetoslava [20 %]. Binary VQ-based model for error process. In Redžúr 2017 : 11th International workshop on multimedia information and communication technologies. Bratislava, Slovakia. May 19, 2017. 1. vyd. Bratislava : Vydavateľstvo Spektrum STU, 2017, S. 15-18. ISBN 978-80-227-4691-5.

## 10 Zoznam použitej literatúry

- [1] J. Klincková, "Neverbálna komunikácia alebo komunikujeme nielen slovami," Univerzita Mateja Bela, Banská Bystrica, 2011, p. 93, ISBN 978-80-557-0168-4.
- [2] S. S. Fels, G. E. Hinton, "Glove-talk: a neural network interface between a data-glove and a speech synthesizer," IEEE Transactions on Neural Networks, vol. 4, no. 1, p. 2–8, 1993.
- [3] R. H. Liang, M. Ouhyoung, "A real-time continuous gesture recognition system for sign language," Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, 1998, p. 558-567.
- [4] C. Savur, F. Sahin, "American Sign Language Recognition system by using surface EMG signal," 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, 2016, p. 002872-002877.
- [5] C. S. Weerasekera, M. H. Jaward, N. Kamrani, "Robust ASL Fingerspelling Recognition Using Local Binary Patterns and Geometric Features," 2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Hobart, TAS, 2013, p. 1-8.
- [6] K. Sadeddine, F. Z. Chelali, R. Djeradi, "Sign language recognition using PCA, wavelet and neural network," 2015 3rd International Conference on Control, Engineering & Information Technology (CEIT), Tlemcen, 2015, p. 1-6.
- [7] N. Pugeault, R. Bowden, "Spelling it out: Real-time ASL fingerspelling recognition," 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, 2011, p. 1114-1119.
- [8] A. Thalange, S. Dixit, "Effect of thinning extent on ASL number recognition using open-finger distance feature measurement technique," 2015 International Conference on Signal Processing and Communication Engineering Systems, Guntur, 2015, p. 39-43.
- [9] H. Kaur and J. Rani, "A review: Study of various techniques of Hand gesture recognition," 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, 2016, p. 1-5.
- [10] P. V. V. Kishore, M. V. D. Prasad, C. R. Prasad, R. Rahul, "4-Camera model for sign language recognition using elliptical fourier descriptors and ANN," 2015 International Conference on Signal Processing and Communication Engineering Systems, Guntur, 2015, p. 34-38.
- [11] P. Shukla, A. Garg, K. Sharma, A. Mittal, "A DTW and Fourier Descriptor based approach for Indian Sign Language recognition," 2015 Third International Conference on Image Information Processing (ICIIP), Wanknaghat, 2015, p. 113-118.
- [12] W. El-Tarhouni, L. Boubchir, N. Al-Maadeed, M. Elbendak, A. Bouridane, "Multispectral palmprint recognition based on local binary pattern histogram fourier features and gabor filter," 2016 6th European Workshop on Visual Information Processing (EUVIP), Marseille, 2016, p. 1-6.
- [13] S. T. Hassan, J. A. Abolarinwa, C. O. Alenoghena, S. A. Bala, M. David, A. Farzaminia, "Intelligent sign language recognition using enhanced fourier descriptor: A case of Hausa sign language," 2017 IEEE 2nd International Conference on Automatic Control and Intelligent Systems (I2CACIS), Kota Kinabalu, Malaysia, 2017, p. 104-109.
- [14] X. Liu, C. Li, L. Tian, "Hand Gesture Recognition Based on Wavelet Invariant Moments," 2017 IEEE International Symposium on Multimedia (ISM), Taichung, Taiwan, 2017, p. 459-464.

- [15] A. Ghasemzadeh, H. Demirel, "3D discrete wavelet transform-based feature extraction for hyperspectral face recognition," in IET Biometrics, vol. 7, no. 1, p. 49-55, 1 2018.
- [16] D. Huang, S. Zhang, Y. Zhang, "EEG-based emotion recognition using empirical wavelet transform," 2017 4th International Conference on Systems and Informatics (ICSAI), Hangzhou, China, 2017, p. 1444-1449.
- [17] A. G. Funde, S. D. Thepade, "Improved sign language recognition using Cosine-Haar hybrid wavelet transform and assorted similarity measures with energy compaction," International Conference & Workshop on Electronics & Telecommunication Engineering (ICWET 2016), Mumbai, 2016, p. 174-177.
- [18] P. Samant, R. Agarwal, A. Bansal, "Enhanced discrete cosine transformation feature based iris recognition using various scanning techniques," 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), Mathura, India, 2017, p. 660-665.
- [19] M. Ashourian, "Embedding sign language interpretation of a video in the wavelet transform domain," 2004 12th European Signal Processing Conference, Vienna, 2004, p. 669-672.
- [20] Y. C. See, N. M. Noor, J. L. Low, E. Liew, "Investigation of face recognition using Gabor filter with random forest as learning framework," TENCON 2017 - 2017 IEEE Region 10 Conference, Penang, Malaysia, 2017, p. 1153-1158.
- [21] T. Pumlumchiak, S. Vittayakorn, "Facial expression recognition using local Gabor filters and PCA plus LDA," 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), Phuket, Thailand, 2017, p. 1-6.
- [22] Zhishen Huang, Dongyang Jiang, Wentao Zhao, "Study of sign language recognition based on Gabor wavelet transforms," 2010 International Conference On Computer Design and Applications, Qinhuangdao, 2010, p. V1-151-V1-154.
- [23] D. A. Kumar, P. V. V. Kishore, A. S. C. S. Sastry, P. R. G. Swamy, "Selfie continuous sign language recognition using neural network," 2016 IEEE Annual India Conference (INDICON), Bangalore, 2016, p. 1-6.
- [24] E. Cardarelli, P. Medici, P. P. Porta, G. Ghisio, "Road signs shapes detection based on Sobel phase analysis," 2009 IEEE Intelligent Vehicles Symposium, Xi'an, 2009, p. 376-381.
- [25] S. Saha, S. Bhattacharya, A. Konar, "Comparison between type-1 fuzzy membership functions for sign language applications," 2016 International Conference on Microelectronics, Computing and Communications (MicroCom), Durgapur, 2016, p. 1-6.
- [26] R. Moradi, R. Yousefzadeh, "Recognizing objectionable images using convolutional neural nets," 2015 Signal Processing and Intelligent Systems Conference (SPIS), Tehran, 2015, p. 133-137.
- [27] Y. H. Chen, H. C. Huang, "Collusion-Resilient Fingerprinting Using Cropping and Subsampling," 2009 Fifth International Conference on Information Assurance and Security, Xian, 2009, p. 98-101.
- [28] R. Amrutha, V. M. Ladwani, "Bharatanatyam hand gesture recognition using normalized chain codes and oriented distances," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, 2016, p. 1-6.
- [29] E. Rachmawati, M. L. Khodra, I. Supriana, "Shape based recognition using freeman chain code and modified Needleman-Wunsch," 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, 2016, p. 1-6.
- [30] H. Althobaiti, Chao Lu, "A survey on Arabic Optical Character Recognition and an isolated handwritten Arabic Character Recognition algorithm using encoded freeman chain code," 2017 51st Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, 2017, p. 1-6.

- [31] V. Hebballi, V. Rojit, "Latent Semantic Analysis (LSA) based object recognition and clustering," 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Noida, 2015, p. 416-421.
- [32] T. K. Senthilkumar, S. Rajalingam, S. Manimegalai, V. G. Srinivasan, "Human facial emotion recognition through automatic clustering based morphological segmentation and shape/orientation feature analysis," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, p. 1-5.
- [33] T. Song, J. Li, X. Han, L. Shao, "A quick Otsu-Kmeans algorithm for the internal pipeline detection," 2017 IEEE International Conference on Mechatronics and Automation (ICMA), Takamatsu, 2017, p. 193-197.
- [34] P. Kaur, "Outlier Detection Using Kmeans, Fuzzy Min Max Neural Network in Network Data," 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN), Tehri, 2016, p. 693-696.
- [35] D. de Santos-Sierra, M. F. Arriaga-Gómez, G. Bailador, C. Sánchez-Ávila, "Low computational cost multilayer graph-based segmentation algorithms for hand recognition on mobile phones," 2014 International Carnahan Conference on Security Technology (ICCST), Rome, 2014, p. 1-5.
- [36] The University Of British Columbia, Vancouver, " MA: Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image", Inventors: D. G. Lowe, United States Patent, US6711293 B1, 08.03. 1999.
- [37] R. Wang, J. W. Jeon, "Design of real-time SIFT feature extraction," IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, China, 2017, p. 3545-3549.
- [38] H. B. Joo, J. W. Jeon, "Feature-point extraction based on an improved SIFT algorithm," 2017 17th International Conference on Control, Automation and Systems (ICCAS), Jeju, Korea (South), 2017, p. 345-350.
- [39] E. Šikudová, Z. Černeková, W. Benešová, Z. Haladová, J. Kučerová, "Počítačové videnie Detekcia a rozpoznávanie objektov," Wikina Praha, Praha, 2013, p. 397, ISBN 978-80-879-2506-5.
- [40] M. Oravec, J. Polec, S. Marchevský, "Neurónové siete pre číslicové spracovanie signálov," FABER, Bratislava, 1998, p. 196, ISBN 978-80-967-5039-9.
- [41] V. Kumar, D. Kalitin, P. Tiwari, "Unsupervised learning dimensionality reduction algorithm PCA for face recognition," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 2017, p. 32-37.
- [42] F. Csóka, B. Puterka, J. Polec, "Real-time recognition of Slovak sign language," ELITECH'17, Bratislava, 2017, p. 5, ISBN 978-80-227-4686-1.
- [43] F. Csóka, B. Puterka, J. Polec, "SSL finger spelling recognition using Gabor wavelet and chamfer distance," Redžúr 2017, Bratislava, p. 11-14. ISBN 978-80-227-4691-5.
- [44] F. Csóka, J. Polec, "Algorithm for real-time finger spelling alphabet recognition in video sequences," SPSympo 2017, Jachranka, 2017, p. 271-274, ISBN 978-1-5090-6755-8.
- [45] RBCS - Ray Bernard Consulting Services, "Technology Growth Curves" [online], 2017, <<http://www.go-rbcs.com/articles/tech-growth-curves>>.
- [46] B. Choi et al., "In-Pixel Aperture CMOS Image Sensor for 2-D and 3-D Imaging," in IEEE Sensors Journal, vol. 18, no. 22, p. 9163-9168, 15 Nov.15, 2018.
- [47] A. Srivastava, A. Lee, B. Simoncelli, S. Zhu, "On Advances in Statistical Modeling of Natural Images", Journal of Mathematical Imaging and Vision, 2003, vol.18, p.17-33, ISSN 1573-7683.

- [48] R. Cui, H. Liu, C. Zhang, "A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training," in *IEEE Transactions on Multimedia*.
- [49] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti, C. Massaroni, "Exploiting Recurrent Neural Networks and Leap Motion Controller for the Recognition of Sign Language and Semaphoric Hand Gestures," in *IEEE Transactions on Multimedia*, vol. 21, no. 1, p. 234-245, Jan. 2019.
- [50] N. Soodtoetong, E. Gedkhaw, "The Efficiency of Sign Language Recognition using 3D Convolutional Neural Networks," 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Chiang Rai, Thailand, 2018, p. 70-73.
- [51] S. S Kumar, T. Wangyal, V. Saboo, R. Srinath, "Time Series Neural Networks for Real Time Sign Language Translation", 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 2018, p. 243-248.
- [52] James J. Gibson, "The Ecological Approach to Visual Perception", Routledge, 1986, ISBN 0898599598.
- [53] Lynne A. Isbell, "The Fruit, the Tree, and the Serpent: Why We See So Well", Harvard University Press, 2011, ISBN 0674061969.
- [54] Martha J. Farah, "The Cognitive Neuroscience of Vision", Blackwell Publishers, 2000, ISBN 0631214038.
- [55] Christopher Chabris, Daniel Simons "The Invisible Gorilla", Harmony, 2010, ISBN 0307459659.
- [56] D. Erhan, C. Szegedy, A. Toshev, D. Anguelov, "Scalable Object Detection Using Deep Neural Networks," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, p. 2155-2162.
- [57] Google LLC, TensorFlow devision, Neural Network Guide, "TensorFlow Guide", [online], 2019, <[https://www.tensorflow.org/tutorials/images/deep\\_cnn](https://www.tensorflow.org/tutorials/images/deep_cnn)>.
- [58] K. Pulli, A. Baksheev, K. Korniyakov, V. Eruhimov, "Realtime Computer Vision with OpenCV", Communications of the ACM, 2012, Doi:10.1145/2184319.2184337.
- [59] ASL-recognition, N. Wilkins, Q. Haqq, B. Njinju, "ASL-recognition", 2016, [online], <<https://github.com/npw3202/ASL-recognition>>.
- [60] A. Yedida, "Sign Language Recognition", 2017, [online], <[https://github.com/AnupamYedida/Sign\\_Language\\_Recognition](https://github.com/AnupamYedida/Sign_Language_Recognition)>.
- [61] Y. Ye, Y. Tian, M. Huenerfauth, J. Liu, "Recognizing American Sign Language Gestures from Within Continuous Videos", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, p. 2145-2149.
- [62] S. R. Kalbhor, A. M. Deshpande, "Digit Recognition Using Machine Learning and Convolutional Neural Network," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, 2018, p. 604-609.
- [63] F. Ertam, G. Aydın, "Data classification with deep learning using Tensorflow," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, 2017, p. 755-758.
- [64] D. Demirović, E. Skejić, A. Šerifović-Trbalić, "Performance of Some Image Processing Algorithms in Tensorflow," 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP), Maribor, 2018, p. 1-4.
- [65] S. Taylor, E. Rosten, T. Drummond, "Robust feature matching in 2.3μs," 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Miami, FL, 2009, p. 15-22.



- [66] Z. Zhang, "Derivation of Backpropagation in Convolutional Neural Network (CNN) ", 2016, Zhang2016DerivationOB.
- [67] L. Schelkes, "Evaluation of Neural Networks for Image Recognition Applications: Designing a 0-1 MILP Model of a CNN to create adversarial", ArXiv, 2018. vol. 1809.00216.
- [68] A. Wills, T. B. Schön, "Stochastic quasi-Newton with adaptive step lengths for large-scale problems.", ArXiv, CoRR abs/1802.04310 (2018).
- [69] B. C. Csáji (2001) Approximation with Artificial Neural Networks; Faculty of Sciences; Eötvös Loránd University, Hungary.
- [70] G. Cybenko "Approximations by superpositions of sigmoidal functions", Mathematics of Control, Signals and Systems, 1989, p. 303–314. doi:10.1007/BF02551274.
- [71] Z. Lu, H. Pu, F. Wang, Z. Hu, L. Wang, "The Expressive Power of Neural Networks: A View from the Width. " Neural Information Processing Systems, 2017, p. 6231-6239.
- [72] B. Hanin, "Universal function approximation by deep neural nets with bounded width and ReLU activations. " arXiv, 2017, preprint arXiv:1708.02691.
- [73] A. Anand, S. S. Tripathy and R. S. Kumar, "An improved edge detection using morphological Laplacian of Gaussian operator," 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, 2015, pp. 532-536.
- [74] A. Kessentini, N. Bahri, N. Masmoudi, A. Samet and M. A. Ben Ayed, "DSP-based down-sampling process using lanczos filter bank," 2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP),