SLOVENSKÁ TECHNICKÁ
UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY
A INFORMATIKY

Ing. Dominik Sopiak

Autoreferát dizertačnej práce

Biometrické rozpoznávanie osôb na základe biometrík tvár a ucho

na získanie: vedecko-akademickej hodnosti
philosophiae doctor, PhD.

v doktorandskom študijnom programe: Aplikovaná informatika
v študijnom odbore: 9.2.9 aplikovaná informatika

Miesto a dátum: Bratislava, 7. 9. 2018

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Ing. Dominik Sopiak

Autoreferát dizertačnej práce

Biometrické rozpoznávanie osôb na základe biometrík tvár a ucho

na získanie: vedecko-akademickej hodnosti
philosophiae doctor, PhD.

v doktorandskom študijnom programe: Aplikovaná informatika

Miesto a dátum: Bratislava, 7. 9. 2018

Dizertačná práca bola vypracovaná: v dennej forme doktorandského štúdia na Ústave informatiky a matematiky FEI STU v Bratislave.

Predkladateľ: Ing. Dominik Sopiak
ÚIM FEI STU
Ilkovičova 3
812 19 Bratislava

Školiteľ: prof. Dr. Ing. Miloš Oravec
FEI STU Bratislava

Oponenti: prof. Ing. Igor Farkaš, Dr.
Katedra aplikovanej informatiky
FMFI UK
Mlynská dolina
824 48 Bratislava

Ing. Matej Féder, PhD.
Hroznová 13
902 01 Pezinok

Autoreferát bol rozoslaný dňa: .....

Obhajoba dizertačnej práce sa koná: .....

Na: Fakulte elektrotechniky a informatiky STU,
Ilkovičova 3, 812 19 Bratislava, v miestnosti C502.

prof. Dr. Ing. Miloš Oravec
dekan FEI STU

Obsah

1 Úvod **1**

2 Ciele dizertačnej práce **2**

3 Dosiahnuté výsledky dizertačnej práce **3**

3.1 Využitie zhlukovacích aloritmov pre rozpoznávanie osôb podľa tváre 3

3.2 Generovanie obrazov tváre z 3D modelu 5

3.3 Biometrická verifikácia tvárí použitím konvolučných neurónových sietí so siamskou architektúrou 7

3.4 Kontrola kvality obrazu pri rozpoznávaní osôb podľa tváre 8

3.5 Mobilná aplikácia na rozpoznávanie osôb podľa ucha 12

4 Splnenie cieľov dizertačnej práce **13**

5 Literatúra **15**

6 Zoznam prác dizertanta **17**

7 Riešiteľ projektov **18**

8 Summary **19**

1 Úvod

Od samotného zrodu spoločnosti sa ľudia zaraďovali do klanov pomocou rôznych tetovaní, oblečenia či iných zvyklostí. S vývojom spoločnosti sa zvyšovala aj náročnosť kvalitnej identifikácie. Dnes, v dobe internetu a mobilných sietí, kedy sa zvýšila úroveň globalizácie spoločnosti a je jednoduché cestovať medzi kontinentmi, je táto úloha extrémne náročná. Firma potrebuje mať prehľad o svojich zamestnancoch a návštevníkoch. Letiská, námestia a rušné miesta potrebujú mať prehľad o podozrivých osobách a samotné štáty chcú mať informácie o prechodoch cez hranicu. Na túto úlohu sa vo veľkej miere aplikujú poznatky z oblasti biometrie, kde sú často použité metódy počítačového videnia a strojového učenia.

Jeden z faktorov, ktorý prispieva k nárastu globalizácie je dostupnosť a rozšírenosť mobilných technológií. Nie len používatelia, ale aj firmy denne prispievajú veľkým množstvom informácií na internet. Podľa [1] publikovanom v 2017 90% dát, ktoré sú na internete, vznikli v roku 2016. Na sociálne siete sa každú minútu prihlási ďalší používateľ. V rokoch 2014-2016 bolo každú minútu nahratých 400 hodín nových video záznamov. 46740 fotografií je nahratých na Instagram každú minútu. Takéto množstvo dát otvára nové možnosti, ktoré doteraz pre hardvérové obmedzenia neboli, zároveň vzniká zodpovednosť tieto dáta chrániť a tiež obmedzovať škodlivý obsah. Metódy strojového učenia a počítačového videnia nám ponúkajú nástroje, ako s týmito dátami pracovať.

Keďže samotný problém verifikácie a identifikácie je veľmi zložitý, existuje viacero spôsobov, ako pristúpiť k jeho riešeniu. Prvým spôsobom, akým je možne identifikovať osobu, je na základe toho "čo vlastní". Bežným príkladom použitia tohoto spôsobu je pri vstupe do budovy pomocou identifikačnej karty, vstup do bytu pomocou kľúča alebo vjazd do parkoviska s platným parkovacím lístkom. Druhý často používaný spôsob identifikácie je založený na tom "čo subjekt vie". Využitie môžeme nájsť pri prístupe na mail, výbere peňazí z bankomatu, kde musíme zadať PIN alebo bezpečnostných otázkach, ktorými si zabezpečujeme rôzne kontá. V prvom aj v druhom prípade je výhodné, že nie je problém osobu identifikovať. Pri predložení preukazu alebo po zadaní hesla je jasné priradiť identitu subjektu. Netreba však zabudnúť na nevýhody a problémy, na ktoré je potrebné myslieť. Všetko čo môžeme chytiť, môžeme aj jednoducho stratiť, poškodiť alebo nám to môže niekto ukradnúť. Heslo alebo PIN sa poškodiť síce nemôžu, no výberom jednoduchého hesla riskujeme, že ho niekto uhádne, avšak zložité heslo je jednoduché zabudnúť. Tiež sa môže stať, že heslo nám niekto ukradne.

Ďalší spôsob identifikácie je založený na tom, "čím subjekt je". Ide o identifikáciu na

základe fyziologických znakov ako odtlačok prsta, výška postavy, geometria ruky, výzor tváre ale aj hlas či pach. Pre človeka je jednoduchá úloha rozpoznať člena rodiny v dave, spoznať kolegov v kancelárií, profesora na prednáške alebo známeho herca v obľúbenej rádiovej inscenácii. Posledný spôsob identifikácie používa informáciu o tom "čo subjekt robí". Každý z nás počas života nadobudol jedinečné návyky, ktoré je takmer nemožné zreplikovať druhou osobou. Ide napríklad o rukopis, chôdzu, gestikuláciu alebo melódiu hlasu. Oba tieto spôsoby sa zaraďujú do vedného oboru biometrie.

2 Ciele dizertačnej práce

Hlavným cieľom, ktorý sme si na začiatku doktorandského štúdia stanovili bolo navrhnúť biometrický systém, v ktorom využijeme najnovšie metódy na rozpoznávanie osôb podľa obrazu tváre a ucha. Aby sme tento systém vytvorili potrebovali sme sa zorientovať v rozsiahlom množstve publikácií. Získaný prehľad v tejto širokej oblasti nám pomohol zamerať sa na nájdené problémy, hlbšie ich analyzovať a navrhnuté riešenia experimentálne vyhodnotiť. V tejto kapitole budú popísané stanovené ciele mojej dizertačnej práce.

1. Navrhnúť a vytvoriť biometrický systém na rozpoznávanie osôb podľa biometriky tváre. Navrhnutý systém bude spĺňať nasledujúce požiadavky.
   - Návrh metód zamerať na rozpoznávanie v reálnom čase,
   - filtrovať vstup do systému na základe kvality obrazu,
   - extrahovať príznaky jednoducho klasifikovateľné v priestore príznakov,
   - navrhnuté metódy otestovať na dostupných databázach tvárí,
   - navrhnuté metódy optimalizovať pre mobilné zariadenia,
   - navrhnúť spôsob riešenia problému jedinej vzorky.
2. Navrhnúť a vytvoriť biometrický systém na rozpoznávanie osôb (identifikácia/verifikácia) podľa biometriky ucha v neriadených podmienkach. Navrhnutý systém bude spĺňať nasledujúce požiadavky.
   - Návrh metód zamerať na rozpoznávanie v reálnom čase,
   - extrahovať príznaky jednoducho klasifikovateľné v priestore príznakov,
   - navrhnuté metódy otestovať na dostupných databázach uší,
   - navrhnuté metódy optimalizovať pre mobilné zariadenia.
3. Aplikovať hlboké konvolučné siete na problematiku rozpoznávania osôb podľa tváre. Otestovať viacero architektúr a aplikovať ich na viaceré časti biometrického systému.

Pri rozpoznávaní sa dáva dôraz na neriadené podmienky pri získavaní snímok, ktorá je bežná pri systémoch použitých na mobilných zariadeniach.

3 Dosiahnuté výsledky dizertačnej práce

V tejto sekcii sa budeme venovať výsledkom experimentov, ktoré sme počas môjho doktorandského štúdia dosiahli. Výskum bol smerovaný k tvorbe biometrických systémov pre zariadenia s obmedzenými výpočtovými alebo pamäťovými nárokmi, akými sú mobilné zariadenia. Porovnávali sme viacero spôsobov extrakcie a klasifikácie príznakov a pracovali sme s biometrikami ucha a tváre.

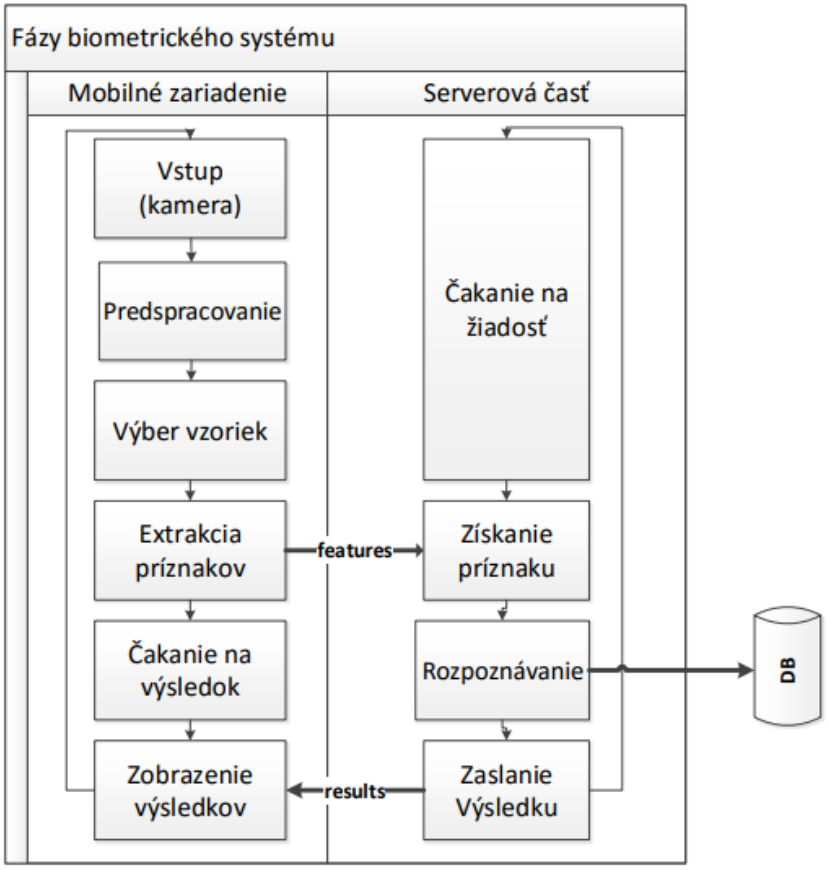
Prvá časť experimentov je venovaná biometrike tváre, kde sme navrhli biometrický systém založený na architektúre klient-server. Ďalej sme sa venovali riešeniu problému jedinej vzorky použitím 3D variabilného modelu, ktorý slúžil na generovanie nových vzoriek. V našom riešení sme použili aj konvolučné siete, ktoré sú aktuálne najúspešnejšími metódami v oblasti počítačového videnia. Prvé testy sme vykonali so siamskou architektúrou, kde sme kontrastnú stratovú funkciu nahradili viacvrstvovým perceptrónom. Ďalšie testy sme vykonali na trojičke konvolučných sietí, pričom vstup do systému bol filtrovaný na základe kvality obrazu.

Ďalšia časť je venovaná biometrike ucha, kde sme vytvorili systém, ktorý bol aplikovaný na mobilné zariadenia. Tento systém sme overili na viacerých dostupných databázach. V poslednej časti opisujeme použitie metód počítačového videnia na detekciu rýchlosti vozidla.

3.1 Využitie zhlukovacích aloritmov pre rozpoznávanie osôb podľa tváre

V tomto projekte sme sa venovali tvorbe riešenia, ktoré sme navrhli pre rozpoznávanie osôb podľa tváre na mobilnom zariadení. Napriek tomu, že v súčasnosti existuje mnoho riešení, ktoré pracujú priamo na mobilnom zariadení, tieto metódy musia čeliť problémom akými sú krátka životnosť batérie, obmedzené zdroje zariadenia a niekedy aj slabý výpočtový výkon. V tejto časti opíšeme systém, ktorý je vhodný pre akékoľvek zariadenie s limitovanými zdrojmi a prístupom ku sieti, pretože systém je založený na klient-server architektúre.

Ako rámec pre tvorbu nášho systému sme použili všeobecnú stavbu biometrických systémov, rozšírenú o výber vzoriek určených na trénovanie systému. Problém bol v tom, že kamera mobilných zariadení dokáže snímať 30fps, kde pri 10 sekundách snímania by



Obr. 1: Návrch architektúry pre klient-server systém.

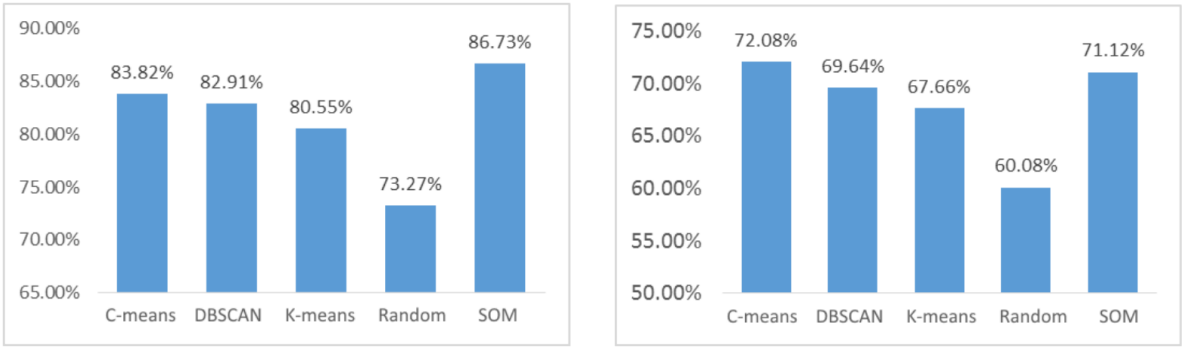
sme získali až 300 snímok, ktoré často obsahujú rovnaké alebo podobné snímky, ktoré sme pred trénovaním potrebovali odfiltrovať. Návrh systému je možné vidieť aj na Obr. 1. Vstup do systému bol zabezpečený pomocou kamery na smartfóne. Systém sme stavali na komerčnom zariadení Android s fotoaparátom s rozlíšením 13 Mpx.

Ako vstup do systému sme použili stredne veľké databázy EURECOM [2] a CMU PIE [3]. Tieto databázy nám slúžili na trénovanie ale aj testovanie našich systémov. Vo fáze predspracovania obrazu sme odstránili pozadie, ktoré neobsahuje tvár, následne zmenšili obraz na jednotnú veľkosť. Tvár sme hľadali v šedotónovom obraze pomocou Viola-Jones algoritmu. Namiesto Haarových príznakov sme však použili LBP príznaky [4]. Nájdené tváre sme vystrihli a všetky takto získané tváre sme zmenšili na 64×64 px. Na záver sme aplikovali histogramovú ekvalizáciu na vyrovnanie jasu a zvýraznenie textúry v obraze.

Výber vzoriek sme riešili pomocou zhlukovacích algoritmov. Konkrétne išlo o algoritmy K-means, C-means, DBSCAN a SOM. Keďže sme chceli získať presný počet snímok aplikácia algoritmov K-means, C-means a SOM bola bezproblémová, pretože pri týchto algoritmoch

je nutné udať parameter, ktorý určuje počet zhlukov. Pre DBSCAN sme museli prehľadávať parameter ϵ, kým sme nedostali potrebný počet zhlukov.

Všetky snímky sme zakódovali pomocou algoritmu LBP kde veľkosť okolia sme zvolili $r = 2$ a počet susedov, ktorými sme okolie kódovali bol $p = 8$. Na klasifikáciu vzoriek sme využili jednoduchú Manhattanskú vzdialenosť. Získané výsledky sú znázornené na Obr. 2. Aplikáciou zhlukovacích algoritmov sa nám podarilo zvýšiť úspešnosť o viac ako 10%. Aplikáciou sofistikovanejších algoritmov, ako SOM alebo C-means sme dosiahli lepšie výsledky. Avšak, tieto algoritmy je problematické využiť pri zariadeniach s nízkym výpočtovým výkonom pre ich výpočtovú náročnosť. V takom prípade je vhodnejšie aplikovať DBSCAN alebo K-means.



(a) Výsledky testov, ktoré sme vykonali na databáze PIE.

(b) Výsledky testov, ktoré sme vykonali na databáze Eurecom.

Obr. 2: Výsledky testov na oboch vstupných databázach.

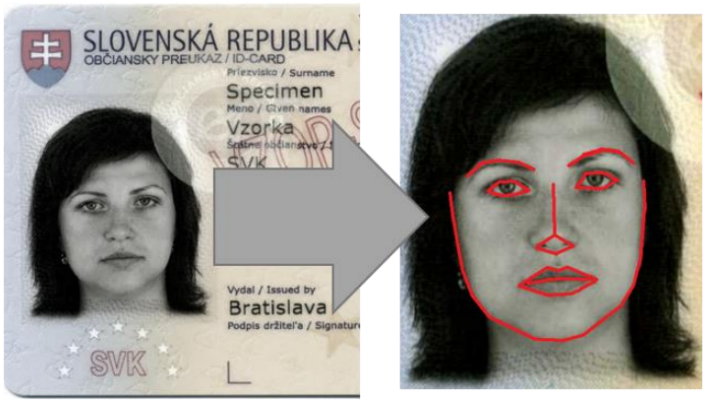
Výsledky experimentov sme prezentovali na domácej konferencii *Redžur v Smoleniciach* a *Medzinárodnej konferencii IWSSIP v Londýne* v 2015.

3.2 Generovanie obrazov tváre z 3D modelu

Táto publikácia predstavuje prístup tvorby a generovania nových syntetických obrazov tvárí. Prvým krokom je detekcia tváre. Po úspešnom nájdení tváre potrebujeme identifikovať kľúčové body, vďaka ktorým našu 2D fotografiu bude možné namapovať na 3D model. Posledným krokom je rekonštrukcia 2D modelu do 3D priestoru, kde už môžeme ďalej s týmto modelom pracovať.

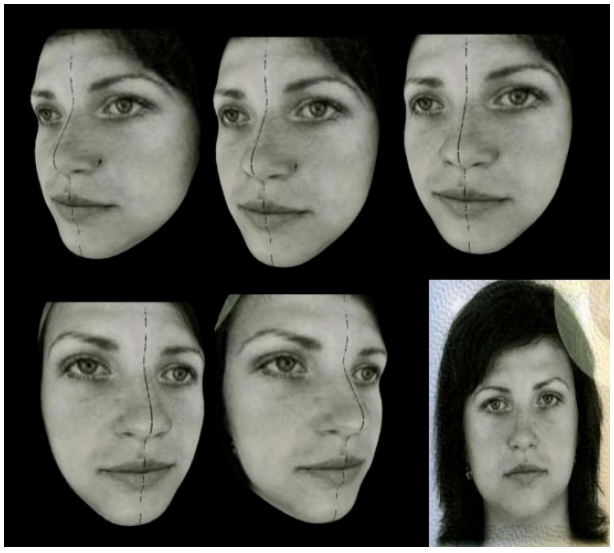
Úlohu detekcie a spracovanie tváre si vieme rozdeliť na dve nezávisle časti. Na lokálizáciu časti obrazu, kde sa tvár nachádza, a následne na získanie pozícií pixelov, kde sa jednotlivé črty tváre nachádzajú. Na detekciu tváre sme použili kombináciu HOG deskriptora a klasifikátora SVM. Na hľadanie tvárových čŕt sme použili algoritmus regresných stromov.

Pomocou tejto metódy sme rýchlo a presne našli kľúčové body.



Obr. 3: Príklad nájdenia tváre a zvýraznenie charakteristických čŕt.

Ďalším krokom ku generovaniu nových obrazov tváre je nájdenú 2D tvár namapovať na 3D model tváre. Museli sme vyriešiť ako namapovať nájdené 2D body a nájsť im ekvivalent v 3D priestore. Následne ako textúru z obrazu na základe zistenej transformácie namapovať na 3D model. Použili sme na to rekonštrukciu na 3D variabilný model v kombinácii so zvlnenou normalizovanou textúrou. Po namapovaní textúry je už možné generovať nové vierohodné vzorky. Natočené tváre sú premietnuté spätne na 2D obraz pomocou metód, ktoré sa bežne využívajú pri 3D projekciách.



Obr. 4: Vygenerované vzorky z ukážkového preukazu.

Takto navrhnutou metódou je však možné generovať vzorky nielen v prípade, že máme len jedinú vzorku, ale aj v prípade aktuálnych konvolučných sietí, ktorá na svoj tréning potrebuje veľké množstvo snímok.

Výsledky týchto experimentov boli prezentované na *medzinárodnej konferencii IWSSIP* v Londýne 2015.

3.3 Biometrická verifikácia tvárí použitím konvolučných neurónových sietí so siamskou architektúrou

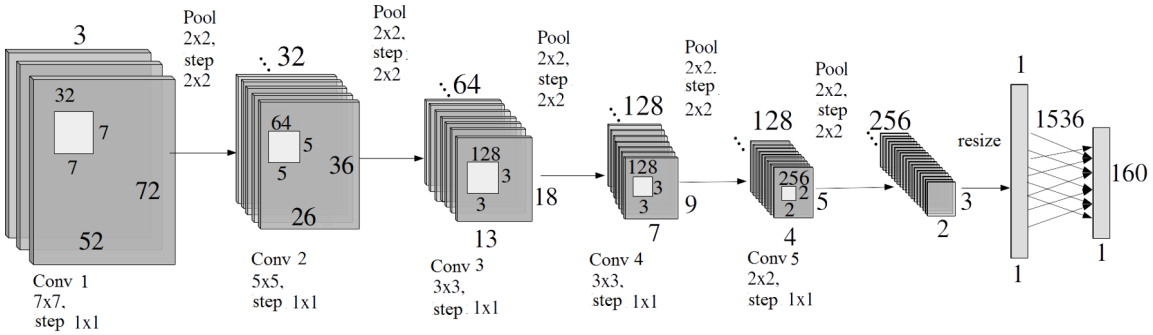
Vývoj v oblasti počítačového videnia ukázal, že pomocou metód konvolučných sietí sa zvýšila úspešnosť detekcie a identifikácie objektov v obraze. V publikácii sme predstavili návrh a implementáciu siamskej architektúry pozostávajúcej z dvoch konvolučných sietí, ktoré sme použili na verifikáciu tvárí na dátovej množine snímok. V obsahu tejto časti sa zameriavame na popis priebehu trénovania siete a ohodnotenie jednotlivých vstupných parametrov, pričom sledujeme ich vplyv na učenie siete.

Inšpirovali sme sa prácami [5] a [6], kde autori navrhli a využívali konvolučnú sieť so siamskou architektúrou. O fúziu dvoch paralelných sietí sa stará kriteriálna funkcia, ktorej hlavná úloha je klasifikovať príznaky získané zo sietí. Úlohou tejto siete je získať reprezentáciu tváre, vhodnú na klasifikáciu aj pomocou jednoduchších metód akou je aj KNN. Cieľom našej práce bolo nahradiť túto kriteriálnu funkciu viacvrstvovým perceptrónom.

Ako vstupnú dátovú množinu sme vybrali tvárovú databázu CelebA, ktorá bsahuje 10 177 subjektov, pričom celkový počet snímok je 202 599. Pred trénovaním sme všetky obrázky orezali podľa pozícií, ktoré sú súčasťou dátovej množiny. Takto získané obrazy sme upravili na rovnakú veľkosť 73×53 pixelov.

Nami navrhnutá konvolučná sieť je založená na architektúre siete AlexNet [7]. Naša sieť je zložená z piatich konvolučných vrstiev, za ktorými nasleduje jedna plne prepojená vrstva. Použitá architektúra je zobrazená aj na Obr. 5. Ako aktivačnú funkciu sme použili ReLU. Za každou konvolučnou vrstvou nasleduje poolingová vrstva s veľkosťou jadra 2×2 a krokom 2, čím sme znižovali dimenziu vstupných príznakov.Výstupné hodnoty z konvolučných jadier sme transformovali do vektora, ktoré slúžili ako vstup pre plne prepojené vrstvy. Pri plne prepojených vrstvách sme ako aktivačnú funkciu pre výstupnú vrstvu použili hyperbolický tangens.

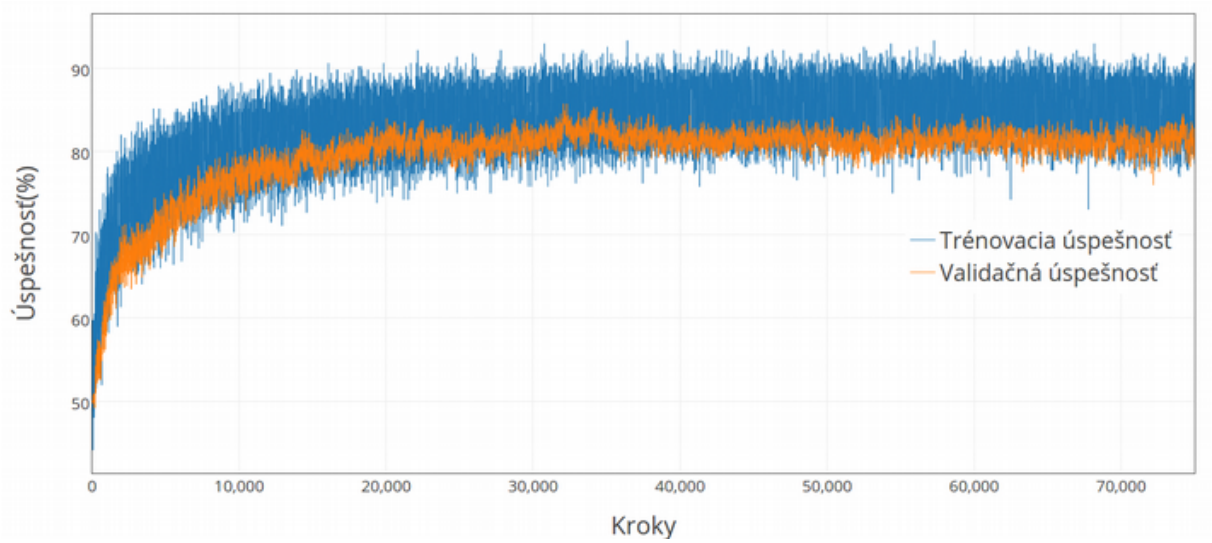
Vstupom pre náš systém boli dve snímky a úlohou systému je zistiť, či sa jedná o tú istú osobu alebo ide o dve rozdielne osoby. Obrazy rovnakých osôb nazveme pravé dvojice a dvojice, na ktorých sú rozdielne osoby nazveme podvodné dvojice. Namiesto kontrastnej stratovej funkcie použitej v článkoch [8, 9], ktorá slúžila na fúziu ľavej a pravej strany



Obr. 5: Architektúra konvolučnej siete.

siete, sme natrénovali ďalšiu plne prepojenú vrstvu. Takéto riešenie sme sa rozhodli použiť ako možnú alternatívu ku kontrastnej stratovej funkcií.

Na trénovanie sme si pripravili trénovaciu sadu s počtom snímok 198 656, z ktorých sme vytvorili 99 328 dvojíc s vyrovnaným pomerom medzi pravými a podvodnými dvojicami. Pomer medzi pravými a podvodnými dvojicami bol vybraný náhodne v intervale medzi 9/11 a 11/9. Testovacia a validačná množina pozostávala z 1024 snímok. Úspešnosť rozpoznávania behom trénovania je zobrazená na Obr. 6. Pomocou konvolučnej neurónovej siete so siamskou architektúrou sme dosiahli úspešnosť 85,74 %.



Obr. 6: Úspešnosť rozpoznávania počas trénovania siamskej konvolučnej neurónovej siete.

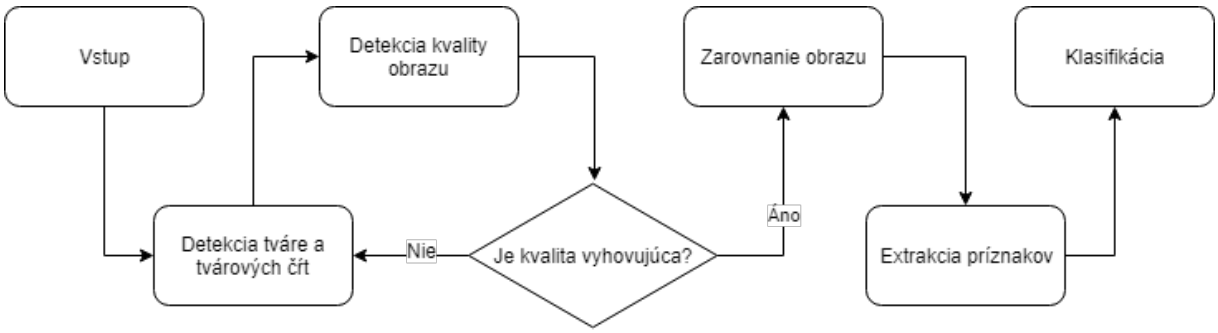
Výsledky boli publikované na *medzinárodnej konferencii ELMAR v Zadare* v 2017.

3.4 Kontrola kvality obrazu pri rozpoznávaní osôb podľa tváre

Cieľom tohto projektu bolo zachytiť snímku spĺňajúcu normy, ktoré sú odporúčané Medzinárodnou organizáciou pre civilné letectvo (angl. The International Civil Aviation Or-

ganization (ICAO)). Konkrétne sa venujú popisu biometrík, ktoré je možné uložiť na identifikačné preukazy (občiansky preukaz, vodičský preukaz, pas), a to tvár, odtlačok prsta a dúhovka. V prípade biometriky tváre, ktorej sa venujeme aj my, uložené biometrické príznaky sa musia pridŕžať normy ISO/IEC 19794-5.

Nami navrhnutý systém je popísaný na Obr. 7. Tento systém aplikuje poznatky získané z predchádzajúcich experimentov. Pri vstupe sme pracovali s viacerými možnosťami. Pred spustením systému je možné nastaviť, či akceptujeme video, vstup z kamery (vstavanej, USB alebo IP) alebo sekvenciu snímok pripravených v adresári.



Obr. 7: Bloková schéma systému an rozpoznávanie osôb podľa tváre.

Na detekciu tváre sme využili viacero algoritmov, ktoré je možné vzájomne zamieňať poprípade rozšíriť o ďalší spôsob detekcie tváre. Jeden z nich využíva príznaky získané algoritmom HOG v kombinácii s jednoduchým lineárnym klasifikátorom. Pomocou posuvného okna prechádzame pyramídu obrazov, kde hľadáme tvár. Jedným z najzaujímavejších detekčných algoritmov, ktoré sme použili je sieť MTCNN (angl. Multi-task CNN), ktorá je popísaná článku [10]. Navrhnutá štruktúra siete sa snaží riešiť tieto problémy pomocou série troch natrénovaných sietí. V prvej fáze sa používa konvolučná sieť, nazývaná P-Net (Proposal Network), na získanie kandidátov pre ohraničenie tváre a ich ohraničovacích vektorov. Získaní kandidáti vstupujú do druhej konvolučnej siete, volanej R-Net (Refine Network), ktorá ďalej odstraňuje veľký počet nesprávnych kandidátov, kalibruje ich ohraničenia. Tretia fáza sa zameriava na identifikáciu regiónov tváre s väčšou presnosťou pomocou siete O-net. Výstupom siete sú najmä pozície piatich hlavných charakteristických znakov tváre (stredy očí, okraje úst a špička nosa).

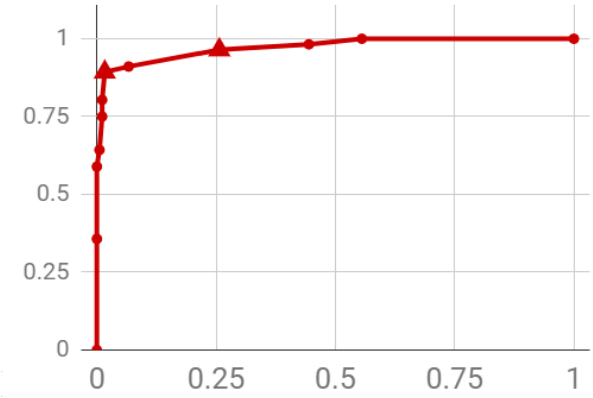
Na kontrolu obrazu sme použili sériu "ručne"vytvorených klasifikátorov a skalárnym súčinom sme získali celkové ohodnotenie obrazu. Kontrolovali sme kontrast, zašumenie a rozmazanie obrazu. Merali sme vzdialenosť očí, otvorenie očí a úst, detegovali sme emocie

a odhadovali pózu. Detekcia tváre na obraze je rýchla operácia na rozdiel od extrakcie príznakov a samotnej klasifikácie. Preto pred samotnou extrakciou skontrolujeme kvalitu obrazu a v prípade nízkej kvality zachytíme nový obraz tváre, až kým kvalita obrazu nebude vyhovujúca.

Ak obraz tváre prejde kontrolou kvality, získanú tvár zarovnáme a získaný obraz zmenšíme na jednotnú veľkosť 160×160 px. Obraz tváre je zarovnaný na základe získaných bodov z 68 bodového modelu. Z nájdených bodov tváre sa do zdrojovej sady vyberie päť bodov, ktoré predstavujú vonkajšiu časť očí, úst a stred nosa.

Pri extrakcii príznakov sme sa inšpirovali architektúrou siete, ktorá je popísaná v [11]. Na rozdiel od siamskej architektúry v tomto prípade použijeme trojicu sietí. Vstupom tým pádom bude trojica snímok, kde 2 snímky budú patriť rovnakej osobe a tretia snímka rozdielnej osobe. Použili sme konvolučnú sieť s architektúrou Inception ResNet v1 popísanú v článku [12]. Keďže predpokladáme, že príznaky z nášho datasetu sú jednoznačne separovateľné, môžeme aplikovať aj jednoduchšie metódy. My sme sa rozhodli aplikovať SVM.

Náš systém sme sa rozhodli aplikovať na dáta získané z video sekvencií televízneho vysielania. Išlo o 701 735 statických obrazov, na ktorých bolo nájdených 502 452 tvárí. Tváre boli následne porovnávane s 103 obrazmi známych osobností, aby sme vedeli popísať, ako dobre systém pracuje pri hľadí tvárí v televíznom vysielaní. Získané výsledky meraní sú vyjadrené ROC krivkou zobrazenou na Obr. 8.



Obr. 8: ROC krivka pre nami navrhnutý systém. Trojuholníkmi vyznačené body zobrazujú tresholdy používané vo zvyšku našich experimentov.

Aby sme zistili, či detekcia kvality má vplyv na rozpoznávanie a ako veľmi kvalita snímky

Obr. 9: Výsledky našich experimentov. Výsledky v prvom rade boli vykonané na pravých snímkach a druhý rad grafov popisuje výsledky získané na podvodných snímkach. Každý zo stĺpcov popisuje iný ukazovateľ. Stĺpec a) rozmazanie, stĺpec b) šum, stĺpec c) rozsah farieb, stĺpec d) štatistickú odchýlku kontrastu, stĺpec e) otvorenie úst a stĺpec f) otvorenie očí. Na všetkých grafoch sú na osi x vynesené vzdialenosti medzi snímkami a na osi y sú zobrazené namerané hodnoty ukazovateľov. Modrá krivka reprezentuje priemernú hodnotu nameraných hodnôt, oranžová hodnotu mediánu, modrá bodkovaná krivka reprezentuje maximálne hodnoty a oranžová bodkovaná krivka reprezentuje minimálne hodnoty.

na rozpoznávanie vplýva, vykonali sme test, kde sme zobrali 2100 podvodných snímok a 3000 pravých snímok. Pri týchto vzorkách sme hľadali závislosti medzi ukazovateľmi kvality a vzdialenosťou jednotlivých prvkov. Na Obr. 9 sú zobrazené merania pre rozmazanie, kontrast a otvorenie a zavretie očí a úst.

Výsledky tejto práce sú priaté na *medzinárodnú konferenciu ELMAR* 2018.

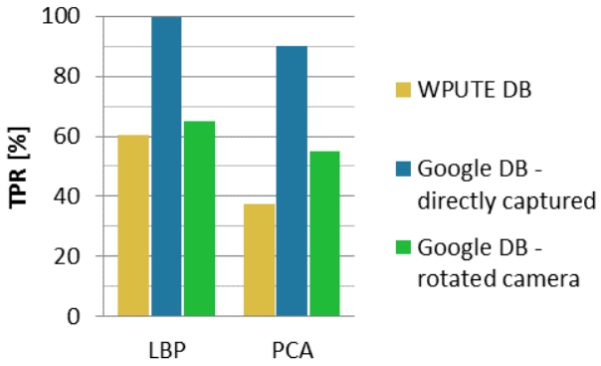
3.5 Mobilná aplikácia na rozpoznávanie osôb podľa ucha

Na základe predchádzajúcich experimentov sme sa rozhodli navrhnúť mobilnú aplikáciu aj pre biometriku ucha. Snahou bolo aplikovať jednoduchšie metódy, ktoré nám umožnia rozpoznávať ucho na mobilnom zariadení v reálnom čase.

Základ aplikácie však beží v 4 základných krokoch, ktorými sú získanie snímky, predspracovanie, extrakcia príznakov a klasifikácia. Ako vstup sme využívali kameru smartfónu. Vo fáze predspracovania sme obraz zmenšili a previedli do šedotónového obrazu, aby sme zrýchlili nájdenie ucha a jeho následné spracovanie. Ucho sme prehľadávali pomocou Viola-Jones algoritmu, kde sme využili LBP príznaky. Takto získané ucho sme z obrazu vystrihli a aplikovali sme histogramovú ekvivalizáciu, aby sme vyrovnali v obraze jas a tým získali nejasné zložky textúry ucha. Veľkosť výsledného obrazu, z ktorého sme získavali príznaky bola 80×160 px.

Obraz ucha obsahuje výrazné štruktúry, ktoré sú detegovateľné z jeho textúry, a preto sme sa rozhodli na extrakciu príznakov využiť LBP algoritmus. Na extrakciu príznakov sme využili aj metódu PCA, ktorá na rozdiel od LBP sa stará o nájdenie najväčšej variácie v obraze, a tak stratíme iba menej významné informácie. Pri klasifikácii sme sa zamerali na jednoduchšie metódy, aby sme efektívne zrýchlili výpočet na úkor úspešnosti rozpoznávania. Konkrétne sme použili euklidovskú metriku a keďže sme pracovali aj s LBP príznakmi, rozhodli sme vyskúšať aj chi-kvadrát vzdialenosť.

Systém sme sa rozhodli aj otestovať, aby sme zistili jeho úspešnosť. Na testovanie sme použili 2 databázy. Jedna databáza bola zozbieraná z internetu, kde sme získali 20 osôb. Ďalšiu databázu sme použili WPUTE, kde každá snímka mala rozlíšenie 380×500 px. Každý subjekt mal presne 4 snímky z čoho sme použili 3 na natrénovanie systému a 1 na testovanie systému. Výsledky týchto testov sú zobrazené na Obr. 10. Tieto výsledky sme aj prezentovali na *konferencii IWSSIP* v 2016, ktorá sa konala v Bratislave.
Získané výsledky bolo potrebné potvrdiť na väčších databázach, aby systém mohol byť natrénovaný a otestovaný na viacerých snímkach. V spolupráci s diplomantom Ing. Fač-



Obr. 10: Získané výsledky testov na rôznych databázach.

kovcom [13] sme jednoduché klasifikačné metriky nahradili sofistikovanejšími algoritmami ako SVM a KNN a príznaky sme extrahovali pomocou algoritmu SIFT a jeho optimalizovanou variantnou SURF. Systém sme testovali na databázach AMI, IIT Delhi a USTB databázach. Na natrénovanie systému sme použili vždy 2/3 dát a zvyšok bol použitý na testovanie. Dosiahnuté výsledky sú zaznačené v tabuľke.

Tabuľka : Výsledky testov navrhnutého systému na rôznych databázach [85].

		klasifikačné algoritmy	
databázy	extrakcia príznakov	SVM	KNN
AMI EAR	SURF	93,3%	40%
	SIFT	91.6%	46.6%
IIT Delhi ear	SURF	52%	19.5%
	SIFT	40.3%	33.6%
USBT ear DB1	SURF	63.2%	47.2%
	SIFT	60.8%	39.2%
USBT ear DB2	SURF	35.5%	43.5%
	SIFT	36.5%	28%

4 Splnenie cieľov dizertačnej práce

V nasledujúcich bodoch sa vyjadrím k splneniu cieľov dizertačnej práce, ktoré sú stanovené v kapitole 2.

- Podarilo sa nám navrhnúť metódy na detekciu tváre, detekciu pozícií tvárových čŕt, extrakciu príznakov a klasifikáciu tváre. Všetky časti boli implementované a experimentálne overené. Ako námet do ďalšej práce uvediem možnosť porovnať všetky kombinácie najčastejšie využívaných metód.

  - Využitie metódy založenej na Viola-Jones algoritme na detekciu tváre je stále vhodný spôsob, ako riešiť tento problém. Veľkou výhodou je najmä jeho rýchlosť a spoľahlivosť.
  - Výber vhodnej metódy na extrakciu príznakov je kľúčové v rozpoznávaní osôb podľa tváre. Konvolučná sieť je aktuálne najlepším nástrojom na získanie týchto príznakov. V kombinácií s klasifikáčnou metódou SVM vieme vytvoriť spoľahlivý systém na rozpoznávanie osôb podľa tváre. Článok popisujúci tieto výsledky je prijatý na konferencii ELMAR [**?**].

- Filtrácia vstupu bola riešená dvoma spôsobmi. Ako prvé sme aplikovali zhlukovacie algoritmy, ktorými sme vedeli vyselektovať reprezentatívne obrazy. Tento spôsob je vhodné použiť v prípade, že na vstup dostaneme veľký počet snímok ako v prípade, keď vstupom je kamera, ktorej rýchlosť snímania je aj 30 fps. Tento spôsob filtrácie bol súčasťou publikácie [**?**]. Druhý spôsob bolo aplikovanie jednoduchých filtrov, popisujúcich vlastnosti ako póza, osvetlenie, kvalita vstupu a emócie. Následne lineárnou kombináciou týchto ukazovateľov vieme vybrať obrazy vhodné na rozpoznávanie. Tento spôsob je náročný na parametrizáciu, keďže každý z ukazovateľov je potrebné vyhotoviť zvlášť. Na druhej strane vieme takto získať väčšiu kontrolu nad vstupným obrazom.

- Na extrakciu príznakov sme použili viaceré metódy. Experimentálne bolo overené, že na extrakciu je najlepšie aplikovať konvolučnú sieť, čo je popísane v kapitole 3.4. Nevýhodou takéhoto riešenia je, že na správne natrénovanie tejto siete je potrebné veľké množstvo snímok, čo v prípade tvárí nie je problém.

- Tiež sme vytvorili systém, ktorý bol implementovaný na mobilnom zariadení. Ten popisujeme v kapitole 3.1.

- Riešenie problému jednej vzorky je opísané v kapitole 3.2, kde sme nové obrazy získavali z 3D modelu tváre.

- Navrhli a implementovali sme všetky časti potrebné pre systém na rozpoznávanie osôb podľa biometriky ucha. Išlo o detekciu ucha v obraze, extrakciu charakteristických príznakov a následnú klasifikáciu týchto príznakov. Navrhnutý systém bol publikovaný v [**?**]. Doplňujúce testy sú súčasťou diplomovej práce [13]. Získané poznatky môžeme zosumarizovať nasledovne:

  - Rovnako ako pri tvári sme aplikovali metódu založenú na algoritme Viola-Jones. Tento algoritmus bol zvolený na základe experimentov, ktoré sú popísané v práci [**?**].

  - Na extrakciu príznakov sme aplikovali viacero metód, ktorých porovnanie nájdeme v kapitole 3.5. Ukázalo sa, že metódy LBP, SIFT a SURF sú vhodné na extrakciu príznakov. V kombinácií s klasifikačnou metódou SVM sme dosahovali najlepšie výsledky.

- Tiež sme vytvorili systém, ktorý bol aplikovaný na mobilné zariadenie. Ten popisujeme v kapitole 3.5.

- V kapitolách 3.3 a 3.4 sme opísali experimenty, kde sme poukázali na aplikácie hlbokých konvolučných sietí. Ukázali sme, že ich je možné použiť ako na extrakciu príznakov, tak aj na detekciu objektov v obraze. Taktiež aktuálny trend výskumu v oblasti počítačového videnia a biometrického rozpoznávania tváre tento fakt potvrdzuje.

Okrem spomenutých publikácií som počas PhD štúdia aktívne spolupracoval so študentmi, čoho výsledkom je odovzdaných a úspešne obhájených 10 bakalárskych projektov, 12 diplomových projektov a 7 tímových projektov. Niektoré zo získaných výsledkov boli súčasťou publikovaných prác.
Naštudované metódy strojového učenia sme aplikovali aj v iných oblastiach. Jednou z nich je rozpoznávanie športových aktivít z video záznamu [**?**], kde sme použili 3D konvolučnú sieť. Výsledky týchto experimentov sú prijaté na konferenciu ELMAR 2018. Tiež sme pomocou metód strojového učenia merali rýchlosť vozidiel [**?**]. Tieto výsledky boli prezentované na medzinárodnej konferencii ELMAR 2017.
Ďalším krokom by mala byť tvorba multimodálneho systému, kde aplikujeme nadobudnuté poznatky z oblasti rozpoznávania osôb podľa biometriky tváre a podľa ucha. Primárnou biometrikou by bola tvár, čo by sa zohľadnilo by fúzii systémov. Pri tvorbe tohto systému je potrebné postupne rozširovať databázu uší. Takto vytvorený systém je vhodné zabezpečiť, keďže pracujeme s citlivými informáciami. Inak by mohlo dôjsť ku zneužitiu týchto citlivých informácií.

5 Literatúra

[1] IBM research team, 10 Key Marketing Trends for 2017. In *10 Key Marketing Trends for 2017*, [Online] [Dátum prístupu: 20.6.2018].

[2] MIN, R., KOSE, N., AND DUGELAY, J.-L. Kinectfacedb: A kinect database for face recognition. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on* 44, 11 (Nov 2014), 1534–1548.

[3] SIM, T., BAKER, S., AND BSAT, M. The cmu pose, illumination, and expression (pie) database. *In Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition* (May 2002), pp. 46–51.

[4] OJALA, T., PIETIKÄINEN, M., AND HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition 29, 1* (1996), 51 – 59.

[5] MELEKHOV, I., KANNALA, J., AND RAHTU, E. Siamese network features for image matching. *In Pattern Recognition (ICPR)*, 2016 23rd International Conference on (2016), IEEE, pp. 378–383.

[6] KHALIL-HANI, M., AND SUNG, L. S. A convolutional neural network approach for face verification. In High Performance Computing & Simulation (HPCS), 2014 *International Conference on (2014)*, IEEE, pp. 707–714.

[7] KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. E. Imagenet classification with deep convolutional neural networks. *In Advances in neural information processing systems* (2012), pp. 1097–1105.

[8] BROMLEY, J., GUYON, I., LECUN, Y., SÄCKINGER, E., AND SHAH, R. Signature verification using a"siamese"time delay neural network. *In Advances in Neural Information Processing Systems* (1994), pp. 737–744.

[9] CHOPRA, S., HADSELL, R., AND LECUN, Y. Learning a similarity metric discriminatively, with application to face verification. *In Computer Vision and Pattern Recognition*, 2005. CVPR 2005. IEEE Computer Society Conference on (2005), vol. 1, IEEE, pp. 539–546.

[10] ZHANG, K., ZHANG, Z., LI, Z., AND QIAO, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters* 23, 10 (2016), 1499–1503.

[11] SCHROFF, F., KALENICHENKO, D., AND PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. *In Proceedings of the IEEE conference* on computer vision and pattern recognition (2015), pp. 815–823.

[12] SZEGEDY, C., IOFFE, S., VANHOUCKE, V., AND ALEMI, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. *In AAAI* (2017), vol. 4, p. 12.

[13] FACKOVEC, M. Tvorba biometrického systému na zariadení raspberry pi podľa biometriky ucha. Diplomová práca, 2017.

6 Zoznam prác dizertanta

1. Sopiak, D. Rozpoznávanie obrazov tvárí v neriadených podmienkach z hľadiska trénovacích množín, Diplomová práca, FEI-5384-64721

2. Oravec, M., Sopiak, D., and Jirka, V. Face recognition on mobile devices based on client-server architecture. In 2015 9th International workshop on multimedia and signal processing (Redžúr) (Mar 2015), pp. 121–124.

3. Oravec, M., Sopiak, D., Jirka, V., Pavlovičová, J., and Budiak, M. Clustering algorithms for face recognition based on client-server architecture. In 2015 International Conference on Systems, Signals and Image Processing (IWSSIP) (Sept 2015), pp. 241–244.

   Citácia:

   - Chowdhury, D.P., Bakshi, S., Guo, G. and Sa, P.K., On applicability of tunable filter bank based feature for ear biometrics: a study from constrained to unconstrained. Journal of medical systems, 42(1), 2018, p.11.

4. Oravec, M., Pavlovičová, J., Sopiak, D., Jirka, V., Loderer, M., Ľ. Lehota, Vodička, M., Fačkovec, M., Mihalik, M., Tomík, M., and Gerát, J. Mobile ear recognition application. In 2016 International Conference on Systems, Signals and Image Processing (IWSSIP) (May 2016), pp. 1–4.

   Citácia:

   - Sumana Achar, Kavana Anahga B. V., A Novel Study on Face Recognition Algorithms and its Challenges, In International Journal for Research in Applied Science & Engineering Technology (IJRASET),Volume 6, Issue 5, (May 2018)

5. Sopiak, D., Oravec, M., Pavlovičová, J., Bukovčíková, Z., Dittingerová, M., Biľanská, A., Novotná, M., and Gontkovič, J. Generating face images based on 3d morphable model. In 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (Aug 2016), pp. 58–62.

6. Gerát, J., Sopiak, D., Oravec, M., and Pavlovicová, J. Vehicle speed detection from camera stream using image processing methods. In 2017 International Symposium ELMAR (Sept 2017), pp. 201–204.

   Citácia:

- Naphade, M., Chang, M. C., Sharma, A., Anastasiu, D. C., Jagarlamudi, V., Chakraborty, P. & Hwang, J. N. The 2018 NVIDIA AI City Challenge. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2018), pp. 53-60.

7. Bukovčiková, Z., Sopiak, D., Oravec, M., and Pavlovičová, J. Face verification using convolutional neural networks with siamese architecture. In 2017 International Symposium ELMAR (Sept 2017), pp. 205–208.

8. Sopiak, D., Bukovčiková, Z., Oravec, M., and Pavlovičová, J. The analysis of quality indicators on face recognition in video frames. *podané a prijaté na* International Symposium ELMAR (Sept 2018)

9. Brezovský, M., Sopiak, D., Oravec, M. Action recognition by 3D convolutional network. *podané a prijaté na* International Symposium ELMAR (Sept 2018)

7 Riešiteľ projektov

1. Založenie výskumného centra pre analýzu a ochranu dát. Operačný program Výskum a vývoj, spolufinancovaný z Európskeho fondu regionálneho rozvoja a štátneho rozpočtu. 26240120037, 2013-2015

2. NaMuBiSy, Návrh multimodálneho biometrického systému, Projekt na podporu mladých výskumníkov, zodpovedný riešiteľ D. Sopiak 2015

3. Návrh pokročilých metód biometrického rozpoznávania na základe obrazov tváre a dúhovky (Proposal of Advanced Methods of Biometric Recognition Based on Face and Iris Images), VEGA 1/0529/13 (zodpovedný riešiteľ M.Oravec, FEI STU), 2013-2016

4. Inovácia vzdelávania v číslicovom spracovaní obrazu, v biometrii a v strojovom učení a neurónových sieťach (Education Innovation in Digital Image Processing, Biometrics, Machine Learning and Neural Networks), KEGA 010STU-4-2014 (zodpovedný riešiteľ J.Pavlovičová, FEI STU), 2014-2016

5. MLbiomedia - pokročilé metódy strojového učenia na návrh biometrických a medicínskych diagnostických systémov (MLbiomedia - Advanced Machine Learning Methods for Proposal of Biometrics and Medical Diagnostic Systems), VEGA 1/0867/17 (zodpovedný riešiteľ M.Oravec, FEI STU), 2017-2020

8 Summary

This work deals with human recognition based upon biometric characteristics, specifically about human face and ear. We review their basic function, development, algorithms that they use and a global overview of state of the art techniques. We also describe the databases of human faces and ears that are used for testing purposes of biometric systems. In the experimental part of this work we deal with face and ear of mentioned methods. Goal and future work based on previous experiments are mentioned in the end of this paper.

In the first paper we presented a client-server approach for face recognition in mobile devices. In addition we have evaluated the use of clustering algorithms for training sample selection. We tested the system with two different databases. In both cases the random selection achieved the worst success rate. Using the CMU-PIE database the random selection reached 73.27 %. The most successful algorithm in this case was the SOM reaching 86.73 %. Using the EURECOM database the random selection reached 60.08 % and the most successful algorithm was the C-means reaching 83.82 %. Globally the success rate increased significantly with the use of any of the clustering algorithms. Advantages and limitations of the proposed methodology related to the used algorithms are as follows: K-means is a very simple algorithm but its convergence to a local minimum may produce wrong results. The C-means algorithm minimizes intra-cluster variance as well, but shows the same problems as K-means; the problems with a local minimum and the results depend on the initial choice of weights. DBSCAN method can find arbitrarily shaped clusters, has a notion of noise, and is robust to outliers. On the other hand it cannot cluster data sets well with large differences in densities, since the $minPts - \epsilon$ combination cannot then be chosen appropriately for all clusters. SOM algorithm is less sensitive to the noise or initialization compared to K-means or C-means.

In our second paper, we consider relevant issues related to single sample problem in the area of face recognition. We proposed method for generating new poses from one frontal face image cropped from ID document using morphable 3D model. The first step is locating the face of a subject. This topic is explained in a first part of the paper, where we discuss locating face using histogram of oriented gradients. Next section explains face alignment and usage of the regression trees and the last part of paper shows process of reconstruction of 2D model to three dimensional (3D) space. We explain our approach for creating and generating new face models which can help in a recognition process. We researched technique for enlargement of the training set by new samples, in order to improve recognition accuracy.

Our next paper evaluates the ability of convolutional networks to solve the problems arising with face classification in unconstrained environment. It contains design and implementation of Siamese architecture consisting of two convolutional networks used for face verification on sets of photographs. Thanks to CNN with Siamese architecture, we reached 85.74% success rate. The results confirm our presumption, that Siamese convolutional network is effective for face recognition in uncontrolled conditions like various poses, illumination or overlapping of human faces with other objects. On the other hand, we need to partially prepare our data to processing by normalization or selection of region of interest. Next improvement of our system can be expected by increasing the size of training dataset or number of training epochs.

In the last paper we examine the effects of various environmental and camera constraints on the accuracy of deeplearning-based face recognition system. More specifically the paper takes a look on the decline of accuracy of the famous CNN architecture trained on "prettyïmages of celebrities, when it is used on stills from video sequences from real-life television broadcast. It introduces a number of different quality indicators (connected to the sharpness of image, the expression of subject and image contrast) and describes their effect on the results of the convolutional network. We have shown that even the face recognition systems trained on professionally-taken still images can be used to recognize people in video sequences. However not all faces in a sequence need to be examined and various indicators of quality can be derived from the frame to speed up the process. We also need to note that even though only results for positive pairs were reported in the paper, the results for impostor pairs follow a similar pattern. Therefore it can be noted that the indicators mentioned in the paper can be an universal indicator for the quality of detection and that the less sharp image is or stranger the pose of the subject, the bigger distance between the descriptors. Further research is required for other indicators of the image quality - e.g. various head tilts and occlusions and the better estimation of image coloring.