SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Ing. Vojtěch Jirka

Autoreferát dizertačnej práce

Nové metódy biometrického rozpoznávania tvárí a dúhoviek v neriadených podmienkach využitím metód strojového učenia

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

v doktorandskom študijnom programe: aplikovaná informatika

v študijnom odbore: informatika

Forma štúdia: externá

Miesto a dátum: Bratislava 31.5.2022

STU

Dizertačná práca bola vypracovaná na Ústave informatiky a matematiky FEI STU

Predkladateľ: Ing. Vojtěch Jirka
ÚIM FEI STU
Ilkovičova 3
812 19 Bratislava

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

Oponenti: prof. Ing. Róbert Hudec, PhD.
Žilinská univerzita v Žiline, Fakulta elektrotechniky a informačných technológií, Katedra multimédií a informačno-komunikačných technológií, Univerzitná 8215/1, 010 26 Žilina (člen Odborovej komisie Informatika na STU)

doc. Ing. Matúš Pleva, PhD.
Technická univerzita Košice, Fakulta elektrotechniky a informatiky, Katedra elektroniky a multimediálnych telekomunikácií, Boženy Němcovej 32, 040 01 Košice

Autoreferát bol rozoslaný: .....

Obhajoba dizertačnej práce sa bude konať dňa 20.12.2022 o 11:00 h.

na Fakulte elektrotechniky a informatiky STU v Bratislave v zasadačke dekana

.....

prof. Dr. Ing. Miloš Oravec
dekan FEI STU

STU

Nové metódy biometrického rozpoznávania tvárí a dúhoviek v neriadených podmienkach využitím metód strojového učenia

Ing. Vojtěch Jirka

Slovenská technická univerzita v Bratislave, Bratislava, 31.5.2022

Školiteľ: prof. Dr. Ing. Miloš Oravec

Práca sa zaoberá problematikou rozpoznávania osôb na základe ľudskej tváre a očnej dúhovky v neriadených podmienkach. Zamerali sme sa hlavne na rozpoznávanie podľa ľudskej tváre so zreteľom na časť predspracovania snímok, kde analyzujeme aktuálny stav riešenia danej problematiky a venujeme sa návrhu spôsobu riešenia problému automatického výberu trénovacích vzoriek. Vypracovali sme návrh, ako vyriešiť tento problém použitím zhlukovacích algoritmov a hlbokého učenia a súčasne sme navrhli použitie ďalších operácií a algoritmov na zlepšenie kvality obrazov. Navrhnuté riešenia sme overili na konvenčných systémoch rozpoznávania ľudských tvárí, použitím overených algoritmov, ako sú analýza hlavných komponentov, lokálne binárne vzory na extrakciu príznakov, stroje s podpornými vektormi na klasifikáciu a rôzne metriky na porovnávanie ako aj aktuálne metódy založené na hlbokom učení. Neriadené podmienky sme simulovali použitím vhodných testovacích databáz. Ich popisu venujeme v práci samostatnú kapitolu. Postup sme rovnako overili aj s použitím mobilných zariadení, kde sa predpokladá vysoká miera neriadených podmienok pri snímaní. Zamerali sme sa na všeobecnú použiteľnosť tohto prístupu v zariadeniach, ktoré nemajú dostatočnú výpočtovú kapacitu, preto sme na vyhodnotenie opäť použili testovacie databázy. Pre použitie v praxi ale navrhujeme overiť riešenie aj s použitím reálnych snímok získaných z konkrétnych zariadení. V problematike rozpoznávaniu podľa dúhovky sme sa zamerali na zostrojenie systému vhodného na kvalitné snímanie obrazov očných dúhoviek. V práci uvádzame aj výskum v oblasti rozpoznávania na základe ucha. Tento výskum, nad rámec práce, bol vykonaný spoločne naším celým tímom pri vykonávaní vedeckej práce, preto ich v práci uvádzame.

Obsah

1. Úvod

Biometria, ako vedná disciplína, sa začala rozvíjať už začiatkom 19. storočia [Orav13], kedy Louis Alphonse Bertillon vyvinul a praktizoval metódu zistenia totožnosti zločincov založenú na meraní ľudského tela. Jeho metóda vychádzala z predpokladu, že sa dĺžky niektorých ľudských kostí a obvod lebky počas života nemenia.

Druhým dôležitým míľnikom Biometrie bol koniec 19. storočia. Juan Vucetich, policajný úradník z Buenos Aires, podľa práce Sira Francisa Galtona položil základy daktyloskopie a ako prvý dokázal identifikovať vraha vďaka krvavému odtlačku, ktorý zanechal.

Biometrické rozpoznávacie systémy využívajú rôzne metódy (biometriky), na základe ktorých sú osoby rozpoznávané. Uviesť môžeme napríklad už spomenutý odtlačok prsta alebo tvár a okrem iných aj dúhovka, sietnica a pod. V komerčne používaných systémoch sú to biometriky, ktoré sú klasifikačne silné, ľahko získateľné a prijateľné ako pre užívateľa, tak aj pre prevádzkovateľa.

V našej práci sa zaoberáme najmä témou rozpoznávania podľa tváre a to konkrétne problematikou predspracovania. Pod predspracovaním rozumieme prípravu zosnímaných snímok tak, aby boli vhodné pre spracovanie biometrickým systémom jednak z pohľadu kvality, ale aj z pohľadu výberu správnych snímok. Pri 30fps (z angl. fps – frames per second - snímok za sekundu), ktoré je možné dosiahnuť takmer akoukoľvek kamerou, získame už len počas 10 sekúnd až 300 snímok tváre. Hlavným problémom je preto výber tých správnych snímok, ktoré zabezpečia najlepšiu úspešnosť biometrického systému. V práci popisujeme nami navrhnuté riešenie, kde zozbierané snímky rozdelíme do skupín a následne z každej vyberieme snímku najlepšie reprezentujúcu danú skupinu. Inak povedané navrhujeme použiť zhlukovacie algoritmy.

Prvá kapitola dizertačnej práce popisuje Biometriu ako vednú disciplínu. Definujeme tu všeobecný model biometrického systému a popisujeme jeho jednotlivé časti. V druhej kapitole dizertačnej práce uvádzame teoretický základ biometrického rozpoznávania tvárí vychádzajúc zo všeobecného modelu biometrického systému definovaného v prvej kapitole dizertačnej práce. Overenie úspešnosti biometrických systémov sa vykonáva porovnávaním výsledkov na referenčných databázach, preto sme vyčlenili samostatnú kapitolu, v ktorej sa venujeme databázam tvárí. V štvrtej a piatej kapitole dizertačnej práce sú uvedené jej ciele a samotný výskum vykonaný v danej problematike.

Okrem hlavného zamerania práce, rozpoznávania tvárí, sme sa počas nášho výskumu zaoberali aj inými metódami rozpoznávania osôb a to pomocou dúhovky a ucha. Na konci piatej kapitoly dizertačnej práce uvádzame aj výskum v týchto oblastiach.

2. Ciele dizertačnej práce

Všeobecným cieľom tejto práce je zlepšiť a umožniť úspešné rozpoznávanie podľa ľudskej tváre a dúhovky v neriadených podmienkach. Je to veľmi komplexný problém, pri ktorom aj všeobecne úspešné metódy rozpoznávania nedosahujú vždy uspokojivé výsledky. Súčasne sme si určili za cieľ, aby navrhnuté metódy a riešenia boli použiteľné v reálnom čase.

V tejto kapitole uvedieme základné problémy, ktorým sa v práci venujeme. Následne popíšeme stanovené ciele dizertačnej práce.

2.1 Problém neriadených podmienok

Rôzne vonkajšie faktory vplývajúce na systémy rozpoznávania spôsobujú nižšiu až neuspokojivú úspešnosť. V reálnom využití rozpoznávacích aplikácií sú snímky často nekvalitné (kvalita snímok záleží od kvality samotného snímacieho zariadenia), s variabilným osvetlením (snímanie ráno verzus večer prípadne v noci alebo cez deň, vo vnútri alebo vonku), natočenia tváre (poloha pri snímaní) a pod [Sing13].

Niektoré vplyvy, ako napríklad variabilita osvetlenia, je možné adresovať použitím vhodných algoritmov. Iné vplyvy, ako prekrytie tváre alebo natočenia, už nie je možné relatívne jednoducho odstrániť. Ako najlepším riešením sa v tomto prípade ponúka natrénovanie rozpoznávacieho systému tak, aby mal na trénovanie dostupné všetky možné variácie, ktoré môžu nastať. Toto riešenie nie je jednoduché uskutočniť z hľadiska použiteľnosti v reálnom čase. Zamerali sme sa preto na riešenie, kde snahou nie je pokryť všetky možné variácie, ale zahrnúť ich čo najväčšie dostupné množstvo. Našou snahou je teda maximalizovať rôznorodosť snímok použitých pre trénovanie, inak povedané, maximalizovať vnútro-triednu variabilitu trénovacej množiny. Za týmto účelom sme navrhli použitie zhlukovacích algoritmov, ktorých centrá zhlukov budú využité na získanie trénovacej množiny.

V oblasti rozpoznávania podľa očnej dúhovky sú dôsledky neriadených podmienok ešte výraznejšie. Rôzne odlesky a nedostatočné osvetlenie spôsobia, že dúhovka nemusí byť vôbec detegovaná alebo jej štruktúra málo rozlíšiteľná. Najdôležitejším faktorom pri snímaní snímok je práve samotné snímacie zariadenie alebo systém, ktorý už vďaka svojej konštrukcii dokáže minimalizovať spomenuté neriadené podmienky.

2.2 Rozpoznávanie v mobilných zariadeniach

Vhodným kandidátom na pozorovanie a testovanie neriadených podmienok je rozpoznávanie v mobilných zariadeniach. Rôznorodosť ich využitia a ich samotná vlastnosť, že sú mobilné, ich

predurčujú na to, aby v nich boli implementované riešenia pre zmiernenie vplyvov neriadených podmienok. Príkladom môže byť databáza MobiFace [Yimi20], ktorá pozostáva zo zozbieraných videí zosnímaných počas živého vysielania z mobilných zariadení.

Ďalšou, nie natoľko výhodnou vlastnosťou mobilných zariadení, je ich hardware-ové obmedzenie. Či už z pohľadu samotného výpočtového výkonu, kapacity batérie, prípadne obmedzeného úložného priestoru. Zamerali sme sa preto na výskum metód použiteľných v zariadeniach, kde nie je priaznivé, či dokonca možné spracovávať komplexné výpočtové operácie.

2.3 Rozpoznávanie v reálnom čase

Prvotnou motiváciou podmienky rozpoznávania v reálnom čase bola aj spoluúčasť na projekte „Hybrid Broadcast Broadband Next Generation“ [P4]. V tomto projekte sme sa venovali rozpoznávaniu používateľov služieb pomocou tváre práve v reálnom čase. Jednalo sa o komplexný biometrický systém s automatickým natrénovaním subjektu a následným rozpoznávaním v reálnom čase. S touto podmienkou sme sa zaoberali aj naďalej pri skúmaní vo využití v mobilných zariadeniach.

2.4 Ciele

Na základe uvedených problémov sme si stanovili nasledovné ciele:

1. Návrh metód s využitím strojového učenia na definíciu vhodnej trénovacej množiny pre rozpoznávanie podľa tváre tak, aby bola maximalizovaná vnútro-triedna variabilita.
2. Zameranie sa na rozpoznávanie podľa tváre v mobilných zariadeniach.
3. Zameranie sa na rozpoznávanie podľa tváre v reálnom čase.
4. Vytvorenie vhodného systému snímania obrazov očných dúhoviek na minimalizáciu neriadených podmienok osvetlenia.

3. Dosiahnuté výsledky dizertačnej práce

Popíšeme dosiahnuté výsledky experimentov, ktoré sme počas doktorandského štúdia vykonali vzhľadom na zadané ciele dizertačnej práce. Zameriavali sme sa hlavne na rozpoznávanie podľa ľudskej tváre, na oblasť predspracovania a na definíciu vhodnej trénovacej množiny.

Následne popíšeme náš systém na snímanie obrazov očných dúhoviek. Zostrojený systém je schopný minimalizovať vplyvy neriadeného osvetlenia a to potlačiť väčšinu odleskov a zvýrazniť samotnú štruktúru dúhovky. V poslednej časti uvedieme výskum vykonaný v oblasti rozpoznávania podľa ľudského ucha, ktorý bol nad rámec tejto práce.

3.1 Výber trénovacích vzoriek zhlukovaním a vplyv predspracovania

Prvý experiment spočíval vo výbere trénovacích snímok pomocou rôznych zhlukovacích algoritmov a zistenia vplyvu rôzneho predspracovania na snímky a následný výber vzorky. Vybrali sme tri zhlukovacie algoritmy. S cieľom využiť strojové učenie bola prvou metódou vybraná neurónová sieť samo-organizujúca sa mapa (SOM) [Orav12]. Ako ďalšiu metódu sme vybrali klasický zhlukovací algoritmus K-means [Spath85], ktorý patrí medzi najznámejšie zhlukovacie algoritmy a posledným vybraným algoritmom bol DBSCAN [Ester96]. Hlavnou motiváciou pre DBSCAN bola skutočnosť, že tento algoritmus nemá pevne daný počet zhlukov.

Na otestovanie výberu vzoriek sme použili databázu CMU-PIE [CMU02]. Táto databáza obsahuje fotky osôb v rôznych natočeniach a pózach, vďaka ktorým môžeme simulovať kontinuálne získanie sekvencie snímok napodobňujúce snímanie videa a neriadené podmienky. Pre každý subjekt sme použili snímky označené c05 (24 snímok na subjekt) (pohľad vľavo), c27 (49 snímok na subjekt) (frontálny pohľad) a c29 (24 snímok na subjekt) (pohľad vpravo) (Obrázok 1). Výsledkom je skupina 97 snímok pre každého zo 68 subjektov.

Obr. 1 Príklad použitých póz C05, C27, C29 jedného subjektu z databázy CMU-PIE

Vybrané snímky v rámci jednej pózy sú variabilné aj v zmysle rôzneho osvetlenia, preto sme výber vzoriek vykonávali na snímkach upravených predspracovaním. Testovali sme dva typy predspracovaní, histogramovú ekvalizáciu a adaptívnu histogramovú ekvalizáciu.

Pri testovaní úspešnosti vybraných zhlukovacích algoritmov sme vychádzali z predpokladu, že najlepšie výsledky rozpoznávania dosiahne biometrický systém, ak bude mať k dispozícii rovnaký počet reprezentatívnych snímok z každej z vybraných póz, resp. ak bude každá póza reprezentovaná aspoň jednou snímkou. Keďže sme si zadefinovali použitie troch rôznych natočení, naším cieľom bolo vybratie dvoch snímok na pózu, čiže výber 6 snímok, ktoré by reprezentovali dané pózy.

Výsledkom zhlukovania sú centrá zhlukov, ktoré nie sú ešte priamo použiteľné na natrénovanie rozpoznávacieho systému. Výsledné snímky sme vybrali porovnaním daných centier so všetkými snímkami subjektu pomocou euklidovskej vzdialenosti (príklad získaných snímok na obr. 2). Snímky, ktoré boli najbližšie k centrám sme vyhlásili za vybrané snímky.

Obr. 2 Ilustrácia centier zhlukov a vybraných snímok. Prvý riadok predstavuje centrá zhlukov, druhý riadok vybrané snímky pomocou euklidovskej vzdialenosti

Obr. 3 Vybraté snímky jedného subjektu pomocou SOM. Prvý riadok s použitím adaptívnej histogramovej ekvalizácie, druhý riadok s použitím histogramovej ekvalizácie, tretí riadok bez predspracovania

Obr. 4 Vybraté snímky pomocou algoritmu DBSCAN. Prvý riadok s použitím adaptívnej histogramovej ekvalizácie, druhý riadok s použitím histogramovej ekvalizácie, tretí riadok bez predspracovania

Obr. 5 Vybraté snímky pomocou algoritmu K-means. Prvý riadok s použitím adaptívnej histogramovej ekvalizácie, druhý riadok s použitím histogramovej ekvalizácie, tretí riadok bez predspracovania

Obrázky 3, 4 a 5 predstavujú príklad výsledku výberu snímok pomocou vybraných zhlukovacích algoritmov na jednej osobe s použitím troch rôznych predspracovaní.

V tabuľke 1 sú uvedené všetky dosiahnuté výsledky výberov. Uvedené hodnoty predstavujú priemerný počet vybraných póz. S prihliadnutím na vyššie spomenutý predpoklad sme najlepšie výsledky dosiahli pri použití adaptívnej histogramovej ekvalizácie a zhlukovania pomocou SOM. Môžeme teda tvrdiť, že z testovaných zhlukovacích algoritmov sa na automatizovaný výber vzoriek hodia najviac algoritmy SOM a K-Means. Dosahovali v priemere podobný počet vybraných vzoriek na pózu, natočenie, či osvetlenie pri použití adaptívnej histogramovej ekvalizácii. Pri ostatných predspracovaniach sme podobné výsledky nedosiahli. Algoritmus DBSCAN nepovažujeme za veľmi vhodný. Pri každej iterácii vyberá rôzny počet snímok, pričom v niektorých prípadoch bol počet snímok príliš nízky. Dosiahnuté výsledky sme uviedli v publikácii na konferencii ELMAR 2013 [V2].

Predspra-covanie	**Použité algoritmy**									
	K-MEANS			**SOM**			**DBSCAN**			
	Póza 05	**Póza 27**	**Póza 29**	**Póza 05**	**Póza 27**	**Póza 29**	**Póza 05**	**Póza 27**	**Póza 29**	**Stredný počet**
Bez HEq	0,588	4,941	0,471	2.603	3.074	0,323	1,397	3,427	1,029	5,853
HEq	1.382	3,528	1,089	2,265	2,632	1,103	1,412	3,00	1,735	6,147
CLAHE	**1,500**	**3.206**	**1,294**	**2,147**	**2,000**	**1,853**	**1,250**	**3,015**	**1,617**	**5,882**

Tab. 1 Priemerný počet vybraných snímok pomocou zhlukovania na pózu

3.2 Rozšírený výber trénovacích vzoriek pomocou SOM I.

Na základe výsledkov s predošlých experimentov sme ďalej rozvíjali použitie SOM ako metódy na výber trénovacích vzoriek. Sledovali sme výsledok výberu na rozšírenej vzorke piatich póz: pózy označené C05 (49 snímok na subjekt) (pohľad vľavo), C07 (24 snímok na subjekt) (pohľad nahor), C27 (49 snímok na subjekt) (frontálny pohľad), C09 (24 snímok na subjekt) (pohľad nadol) a C29 (24 snímok na subjekt) (pohľad vpravo) (obrázok 6).

Obr. 6 Príklad použitých póz jedného subjektu z databázy CMU-PIE (zľava doprava C05, C07, C27, C09, C29)

Obr. 7 Príklad vybraných snímok pomocou SOM bez predspracovania. Prvý riadok predstavuje reprezentáciu centier zhlukov a druhý riadok k nim nájdené snímky.

Obr. 8 Príklad vybraných snímok pomocou SOM s histogramovou ekvalizáciou. Prvý riadok predstavuje reprezentáciu centier zhlukov a druhý riadok k nim nájdené snímky.

Obr. 9 Príklad vybraných snímok pomocou SOM s adaptívnou histogramovou ekvalizáciou. Prvý riadok predstavuje reprezentáciu centier zhlukov a druhý riadok k nim nájdené snímky.

Postupovali sme obdobne ako v prvom experimente. Na obrázkoch 7, 8 a 9 uvádzame príklad výberu snímok jedného subjektu pre rôzne predspracovania. Výsledky výberu snímok na celej množine subjektov uvádzame v tabuľke 2. Je zrejmé, že ani pri použití adaptívnej histogramovej ekvalizácie sme nedosiahli zastúpenie každej pózy pre každý subjekt. Pózy natočenia nahor (C07) a natočenia nadol (C09) boli vo všeobecnosti málo vyberané. Môže to byť spôsobené jednak nízkym počtom snímok v týchto pózach oproti ostatným pózam (29 snímok pre pózu C07 verzus 49 snímok pre pózu C27), prípadne podobnosťou póz C07, C09 (natočenia nahor a nadol) a C27 (frontálny záber) pri niektorých subjektoch (obrázok 10). Výsledok z týchto experimentov sme publikovali na konferencii REDŽUR 2014 [V4].

Obr. 10 Príklad podobnosti póz C07, C09 a C27 u jedného zo subjektov.

Predspracovanie	**Pózy**				
	C05	**C07**	**C27**	**C09**	**C29**
Bez HEq	1.83	0.84	4.56	0.94	0.83
HEq	2.32	0.44	5.34	0.53	0.37
CLAHE	**2.59**	**0.66**	**4.15**	**0.69**	**0.91**

Tab. 2 Priemerný počet vybraných snímok pomocou SOM I.

3.3 Rozšírený výber trénovacích vzoriek pomocou SOM II.

V tomto experimente sme pokračovali v testovaní výberu vzoriek pomocou siete SOM. Vstupom do SOM bola aj tentoraz vybraná skupina snímok z databázy CMU-PIE. Použili sme rozšírenú vzorku s pózami pohľad vľavo (C05), pohľad nahor (C07), frontálny pohľad (C27), pohľad nadol (C09) a pohľad vpravo (C29), teda rovnako ako v pri predošlom experimente. Aby sme minimalizovali rozdiel počtu vybraných snímok (tabuľka 2), z každej pózy sme vybrali rovnaký počet 24 snímok, čo predstavuje 120 snímok na osobu.

Opäť sme výber vzoriek porovnávali aj s použitím vyššie spomenutých metód predspracovania – histogramovou ekvalizáciou a adaptívnou histogramovou ekvalizáciou.

Obr. 11 Príklad vybraných snímok pomocou SOM s predspracovaním pomocou adaptívnej histogramovej ekvalizácie. Prvý riadok predstavuje vizualizáciu centier zhlukov a druhý riadok snímky ktoré boli vybrané ako najbližšie pomocou euklidovskej vzdialenosti

Príklad výberu vzoriek pomocou SOM s predspracovaním pomocou adaptívnej histogramovej ekvalizácie je znázornený na obrázku 11. V tomto prípade sme dostali tri snímky s natočením vľavo (C05), jednu snímku s natočením nahor (C07), tri frontálne snímky (C27), jednu snímku s natočením nadol (C09) a jednu snímku s natočením vpravo (C29).

Z výsledkov výberu vzoriek (tabuľka 3) je zrejmé viditeľné zlepšenie rozdelenia počtu vybraných snímok na pózu. Aj keď rozdelenie výberu stále nie je úplne rovnomerné, je viditeľné zlepšenie oproti výsledku z tabuľky 2 vďaka použitiu rovnakého počtu snímok na pózu. Môžeme preto predpokladať, že nízky počet vybraných snímok z pózy C07 a C09 z predošlého experimentu bol spôsobený aj práve nerovnomerným rozdelením počtu snímok v jednotlivých pózach. Zostávajúce rozdiely v počte výberov môžeme pripísať veľkej podobnosti póz C07 s pózami C27 a C09 a hlavne podobnosti póz C27 a C09 osobitne, ako bolo spomenuté vyššie (obrázok 10). Najlepšie výsledky sme opäť dosiahli s použitím predspracovania pomocou adaptívnej histogramovej ekvalizácie. Výsledky týchto experimentov sme publikovali na konferencii ELMAR 2014 [V5].

Predspracovanie	**Póza**				
	C 05	**C 07**	**C 27**	**C 09**	**C 29**
-	0,412	2,353	3,440	0,927	1,868
HE	1,279	1,382	4,279	1,147	0,913
CLAHE	**1,720**	**1,573**	**2,882**	**1,119**	**1,706**

Tab. 3 Priemerný počet vybraných snímok pomocou SOM II.

3.4 Úspešnosť výberu trénovacích vzoriek pomocou SOM

Po teoretickom overení prístupu výberu vzoriek pomocou siete SOM v predošlých experimentoch sme pristúpili k overeniu reálnej úspešnosti rozpoznávania na rozpoznávacích systémoch. Testovanie úspešnosti prebiehalo s použitím rozpoznávacích systémov založených na PCA [Sing13], [Orav13], [Mull01], [Turk91] a SVM [Orav12], [Mull01]. Prvý rozpoznávací systém

bol jednoúrovňový, s jednoduchým SVM klasifikátorom. Druhý systém bol dvojúrovňový, kde prvá úroveň predstavovala extrakciu príznakov pomocou algoritmu PCA a druhá úroveň bola opäť SVM klasifikátor.

V tomto experimente sme sa vrátili k pôvodnému výberu šiestich snímok, kde bolo rozdelenie počtu vybraných snímok na pózu najrovnomernejšie (tabuľka 1). Výber trénovacej množiny pomocou siete SOM sme porovnávali s náhodnými výbermi. Celkovo sme uskutočnili päť výberov trénovacích vzoriek, ktoré sme testovali na vyššie spomenutých rozpoznávacích systémoch.

- ***Výber vzoriek pomocou SOM***
  Sieť SOM mala výpočtovú vrstvu o veľkosti 2x3 neuróny. Výsledkom bolo šesť centier ktoré sme pomocou euklidovskej vzdialenosti porovnávali so snímkami subjektu. Snímky, ktoré boli najbližšie k centrám, sme vybrali do trénovacej množiny.
- ***Výber vzoriek pomocou SOM s histogramovou ekvalizáciou***
  Výber prebiehal podobne ako v prvom prípade, ale na snímky bola najprv aplikovaná histogramová ekvalizácia.
- ***Výber vzoriek pomocou SOM s adaptívnou histogramovou ekvalizáciou***
  Výber prebiehal podobne ako v prvom prípade, ale na snímky bola najprv aplikovaná adaptívna histogramová ekvalizácia.
- ***Náhodný výber – RAND***
  Výber vzoriek prebiehal náhodným výberom. Do trénovacej množiny sa náhodne vybrala vzorka šiestich snímok z celého počtu dostupných snímok na subjekt.
- ***Kontrolovaný náhodný výber – CRAND***
  Výber vzoriek prebiehal kontrolovaným spôsobom. Do trénovacej množiny sa z každej z troch póz subjektu náhodne vybrali dve snímky.

Vybrané trénovacie vzorky boli použité na natrénovanie vyššie spomenutých rozpoznávacích systémov a následné vyhodnotenie úspešnosti. Samotné rozpoznávanie bolo vykonané s použitím originálnych snímok (bez predspracovania), aby neboli výsledky rozpoznávania touto operáciou ovplyvnené.

Výsledky úspešnosti rozpoznávania uvádzame v grafoch na obrázkoch 12 a 13. Najlepšie výsledky sme dosiahli pri výbere pomocou siete SOM, s použitím adaptívnej histogramovej ekvalizácie na predspracovanie snímok. Dosiahli sme 82,13 % úspešnosť pre jednoúrovňový systém a 81,33 % úspešnosť pre dvojúrovňový systém. Najnižšiu úspešnosť sme dosiahli pri

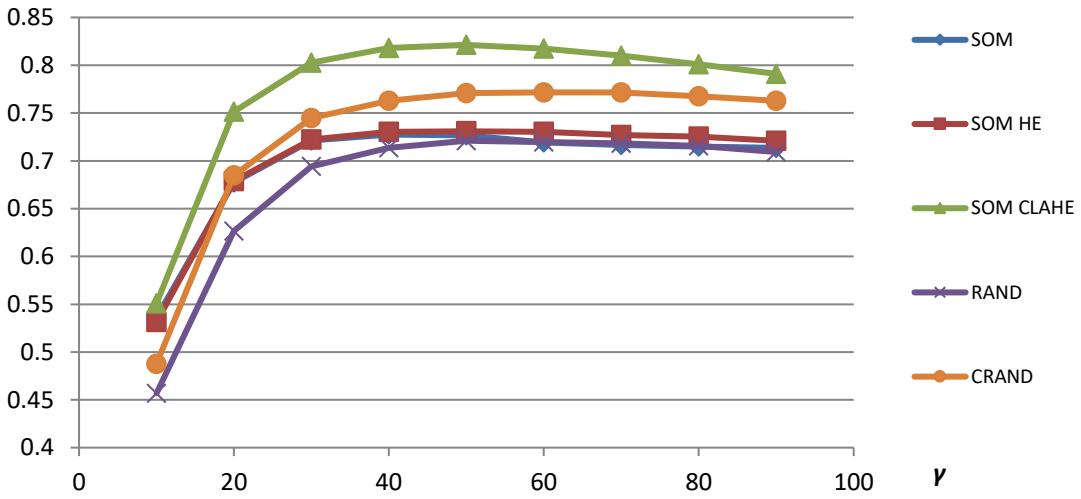
výbere RAND - nekontrolovanom náhodnom výbere, dosahujúc 72,10 % pre jednoúrovňový systém a 66,47 % pre dvojúrovňový systém.

Z výsledkov vyplýva, že samo-organizujúca sa mapa sa javí ako vhodný nástroj na automatizovaný výber vzoriek pre trénovaciu množinu, pričom najlepšie výsledky je možné dosiahnuť pri aplikácii predspracovania pomocou adaptívnej histogramovej ekvalizácie. Súčasne môžeme prehlásiť predpoklad, ktorý sme si zaviedli v prvom experimente, že najlepšie výsledky bude rozpoznávací systém dosahovať práve vtedy, ak bude každá dostupná póza rovnomerne zastúpená v trénovacej množine za nie úplne správny. Vyplýva to z výsledkov, ktoré sme dosiahli pri použití výberu CRAND - kontrolovaného náhodného výberu, kde sme jednotlivo z každej pózy vyberali presný počet dvoch snímok, čím sme zabezpečili splnenie rovnomerného rozdelenia zastúpených póz. Úspešnosť tohto prístupu bola v priemere o 5 až 6 % nižšia ako v prípade použitia samo-organizujúcej sa mapy a adaptívnej histogramovej ekvalizácii.

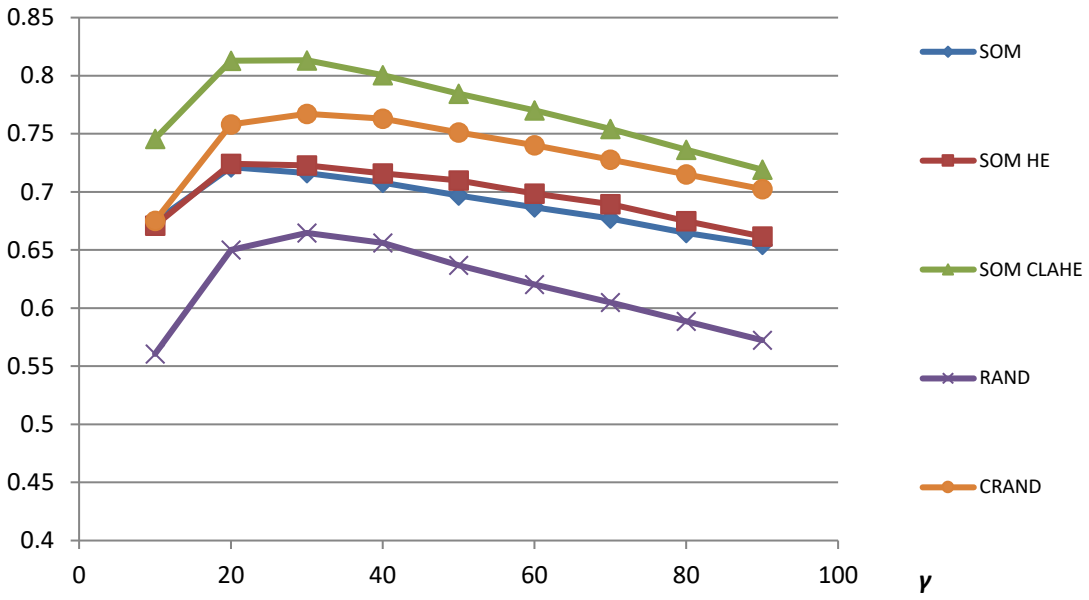
V tabuľke 4 uvádzame najlepší dosiahnutý výsledok pre každý spôsob výberu vzoriek. Výsledky týchto experimentov sme uviedli v publikácii na konferencii ELMAR2014 [V5].

Rozpoznávací systém	**Metóda výberu vzoriek**				
	SOM bez predspraco-vania	**SOM s HE**	**SOM s CLAHE**	**Kontrolovaný náhodný výber**	**Nekontrolovaný náhodný výber**
SVM	72,75 %	73,09 %	**82,13 %**	77,16 %	72,10 %
PCA + SVM	72,13 %	72,42 %	**81,33 %**	76,72 %	66,47 %

Tab. 4 Najvyššia úspešnosť rozpoznávania s automatickým výberom trénovacích vzoriek



Obr. 12 Úspešnosť rozpoznávania v závislosti od γ (parameter SVM) pre jedno-úrovňový rozpoznávací systém založený na SVM. Výber vzoriek bol uskutočnený pomocou SOM s rôznym predspracovaním a pomocou náhodného výberu.



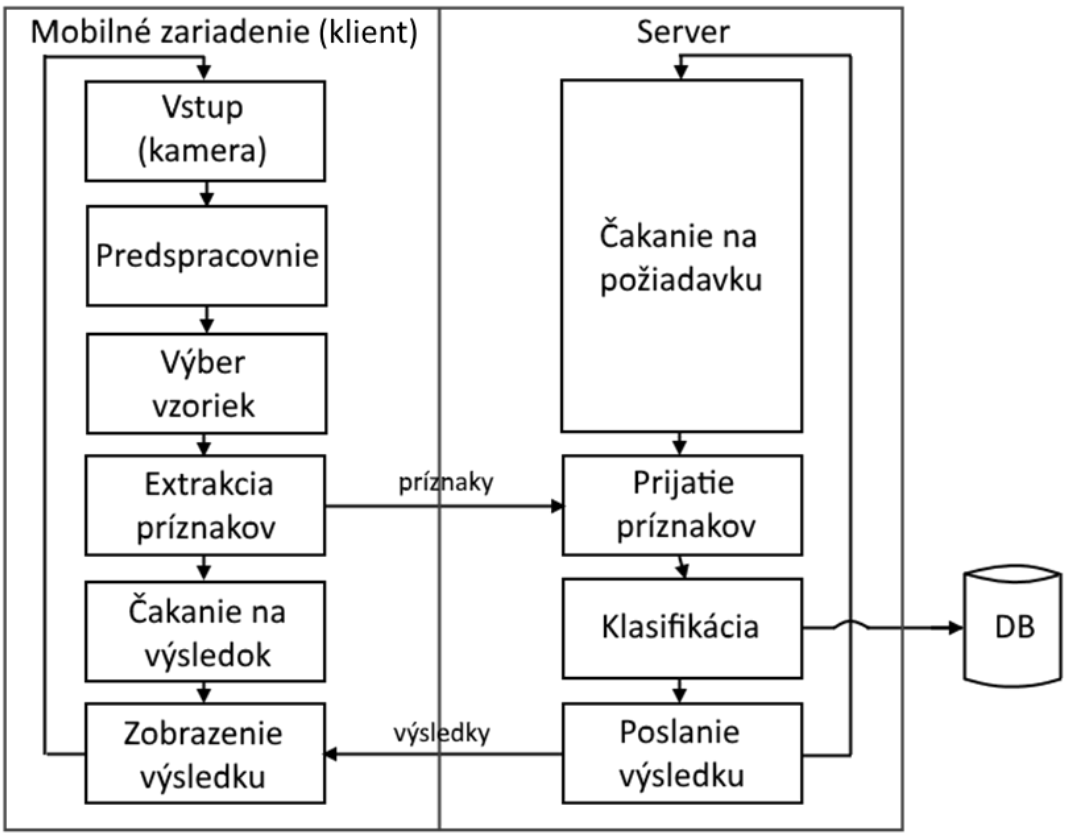
Obr. 13 Úspešnosť rozpoznávania v závislosti od γ (parameter SVM) pre dvoj- úrovňový rozpoznávací systém založený na PCA a SVM. Výber vzoriek bol uskutočnený pomocou SOM s rôznym predspracovaním a pomocou náhodného výberu

3.5 Systém rozpoznávania tvárí na mobilnom zariadení

V našej práci sa zaoberáme predspracovaním pri rozpoznávaní tváre v neriadených podmienkach. Najmenej riadené podmienky je možné očakávať pri rozpoznávaní na mobilných zariadeniach. Jedným z cieľov tejto práce preto bolo zamerať sa aj na návrh metód použiteľných v mobilných zariadeniach. Pri rozpoznávaní na týchto zariadeniach sa predpokladajú výrazné vplyvy neriadených podmienok, ako osvetlenie, natočenie, výrazy a pod.

Väčšina mobilných zariadení nedisponuje dostatočnou výpočtovou kapacitou (pozn. autora: v čase vykonávania daných experimentov bola táto skutočnosť výraznejšia, dnes je už situácia značne lepšia) na zvládnutie náročných algoritmov alebo dostatočným úložným priestorom tak, aby bolo umožnené rozpoznávanie subjektov v reálnom čase. Bolo preto potrebné navrhnúť architektúru, ktorá by zložité úkony a úložný priestor delegovala vzdialenému zariadeniu. Samotné mobilné zariadenie by sa používalo na snímanie subjektov a ako používateľské rozhranie. Systém je preto založený na klient-server architektúre, kde klient predstavuje mobilné zariadenie.

Na obrázku 14 je znázornený návrh našej architektúry. Pre čo najdlhšiu životnosť batérie mobilného zariadenia, ale zároveň aj pre minimalizáciu množstva odosielaných dát, sme extrakciu príznakov navrhli vykonávať na strane klienta, pričom klasifikácia a databáza je na strane servera. Na implementáciu bola použitá OpenCV [Brad00] knižnica, ktorá obsahuje veľké množstvo už implementovaných algoritmov použiteľných pri rozpoznávaní.



Obr. 14 Navrhnutá klient-server architektúra systému rozpoznávania tvárí

Úlohami mobilného zariadenia sú:

- Zaznamenávanie snímok
- Predspracovanie
  - Lokalizácia oblasti tváre
  - Orezanie fotografie
  - Prevedenie do odtieňov sivej
  - Histogramová ekvalizácia
- Výber vzoriek pomocou zhlukovacích algoritmov
- Extrakcia príznakov
- Komunikácia so serverom
- Zobrazenie výsledkov

Úlohami servera sú:

- Klasifikácia
- Úložisko príznakov
- Komunikácia s klientom

V tomto experimente sme sa zamerali na samotný návrh architektúry a zostrojenie proof-of-concept (z angl. overenie koncepcie) aplikácie. Výsledkom je funkčná aplikácia rozpoznávania tvárí na mobilných zariadeniach v reálnom čase. Overeniu úspešnosti rozpoznávania aj v zmysle použitých zhlukovacích algoritmov pre výber trénovacej množiny sa venujeme v nasledujúcom experimente. Návrh architektúry a prvotnú implementáciu nášho systému sme publikovali na konferencii REDŽUR 2015 [V6].

3.6 Úspešnosť výberu trénovacích vzoriek na klient-server architektúre

Podľa návrhu architektúry z predošlého experimentu bolo úlohou overiť prínos výberu trénovacích vzoriek na úspešnosť rozpoznávania na mobilných zariadeniach. Na testovanie úspešnosti a simuláciu neriadených podmienok sme použili referenčné databázy. Okrem databázy CMU-PIE použitej v prvom experimente, sme zahrnuli aj použitie databázy EURECOM Kinect Face Database [Rui14]. Z dôvodu veľkosti použitých databáz sme v týchto experimentoch mobilné zariadenie simulovali použitím osobného počítača. V konečnom dôsledku to na vyhodnotenie úspešnosti výberu vzoriek pomocou zhlukovacích algoritmov nemá žiaden vplyv.

Na výber vzoriek sme použili všetky algoritmy použité v predošlých experimentoch. Na extrakciu príznakov sme použili LBP deskriptory a na klasifikáciu sme použili metriku Manhattan.

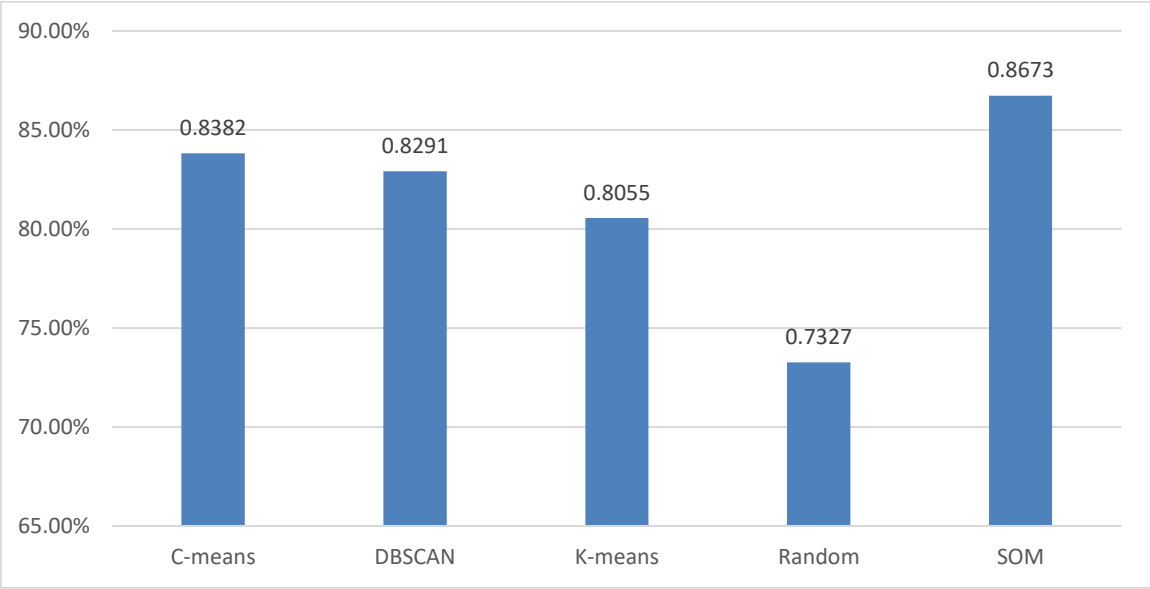
Postupovali sme odlišnou metodikou ako v predošlých experimentoch. Prvým krokom bolo vykonanie výberu trénovacích vzoriek pomocou algoritmu DBSCAN. Tento algoritmus, ako sme už vyššie spomenuli, nevracia pevne daný počet zhlukov, preto bolo najprv potrebné zistiť minimálny počet snímok, ktoré dokážeme získať pre každý subjekt. Na základe výberu sme dospeli k počtu piatich snímok na subjekt. Tento počet sme teda definovali aj pre ostatné algoritmy K-means, C-means a náhodný výber RANDOM. Neurónová sieť SOM má ale výpočtovú vrstvu definovanú ako maticu X*Y neurónov, preto najbližší vyšší počet ako päť zhlukov bol v tomto prípade počet 2x3 neurónov, čiže šesť centier, pričom sme do testovania vybrali vždy prvých päť centier, aby bol počet snímok rovnaký pre každý algoritmus. K výsledným centrám zhlukov boli pomocou euklidovskej vzdialenosti nájdené reálne snímky ktoré sa následne použili pri testovaní úspešnosti. Z vybraných snímok sme vytvorili LBP deskriptory, ktoré sme uložili do databázy na serveri. Klasifikácia prebiehala použitím metriky vzdialenosti Manhattan.

Výsledky meraní uvádzame v grafoch na obrázkoch 15 a 16. Graf na obrázku 15 predstavuje dosiahnuté úspešnosti pre databázu CMU-PIE. Najvyššiu úspešnosť 86,73 % sme dosiahli pri použití výberu pomocou siete SOM. Najnižšia úspešnosť 73,27 % bola dosiahnutá s náhodným výberom RANDOM. Na obrázku 16 je graf úspešnosti nad databázou EURECOM. Najvyššiu úspešnosť 72,08 % sme v tomto prípade dosiahli s použitím algoritmu C-means. Sieť SOM bola

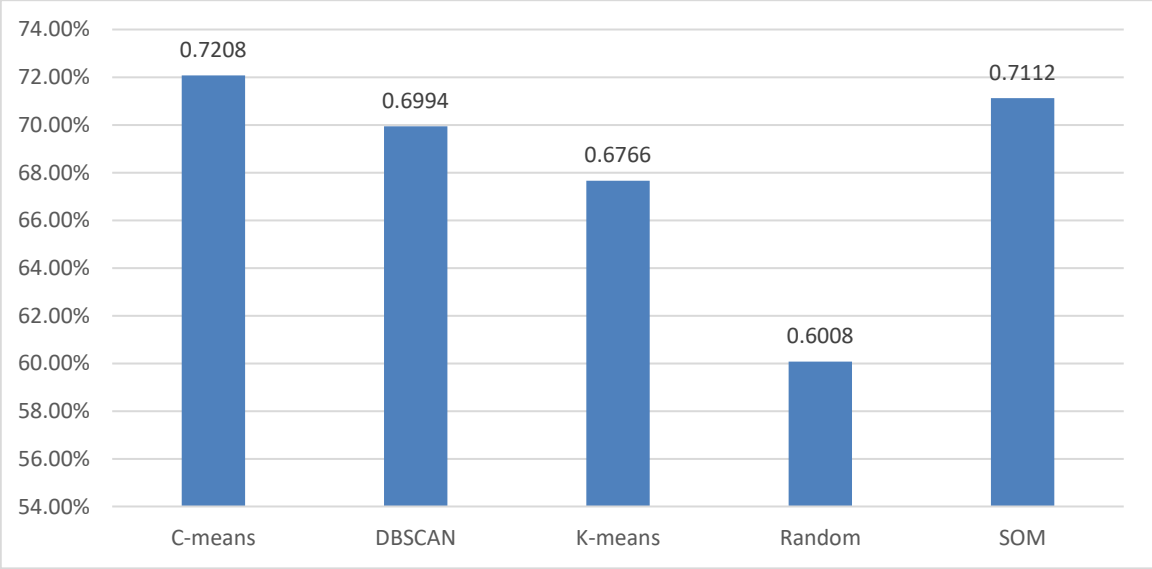
na druhom mieste s úspešnosťou 71,12 %. Najnižšiu úspešnosť 60,08 % sme opäť dosiahli použitím náhodného výberu RANDOM.

Na rozdiel od experimentu v kapitole 3.4 sme nepoužili adaptívnu histogramovú ekvalizáciu na predspracovanie snímok. V tomto prípade sme snímky upravili iba klasickou histogramovou ekvalizáciou. Z výsledkov experimentu z kapitoly 3.4 môžeme predpokladať, že pri použití adaptívnej histogramovej ekvalizácii by mohla byť úspešnosť vyššia. Ďalším faktorom, ktorý mohol ovplyvniť úspešnosť rozpoznávania pomocou SOM, bola skutočnosť, že sme na trénovanie použili vždy prvých päť snímok zo šiestich vybraných.

Výsledky dosiahnutých úspešností sme publikovali na konferencii IWSSIP 2015 [V7].



Obr. 15 Úspešnosť rozpoznávania na databáze CMU-PIE.

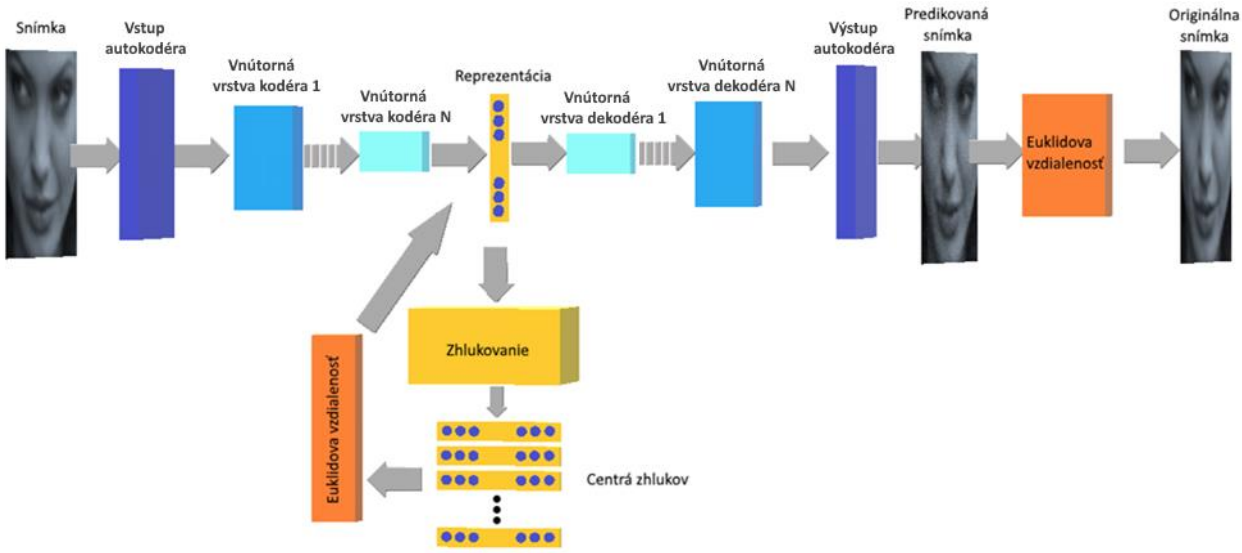


Obr. 16 Úspešnosť rozpoznávania na databáze EURECOM.

3.7 Výber trénovacej množiny pomocou hlbokého zhlukovania

V tomto experimente sme na výber trénovacej vzorky využili hlboké učenie. Vďaka rozsiahlemu vývoju v tejto oblasti a aj vďaka pokročilej technológii je reálne využitie algoritmov hlbokého učenia v biometrii čoraz dostupnejšie. Hlboké učenie je možné využiť aj v iných oblastiach, ako len v rozpoznávaní. Príkladom môže byť prehľadová práca [Min18], kde autori uvádzajú úspešné použitie hlbokých autokodérov pre zhlukovanie v rôznych oblastiach. Výhodou použitia autokodérov je, že sú schopné zo vstupných dát extrahovať príznaky, ktoré dokážu najlepšie popísať dané dáta.

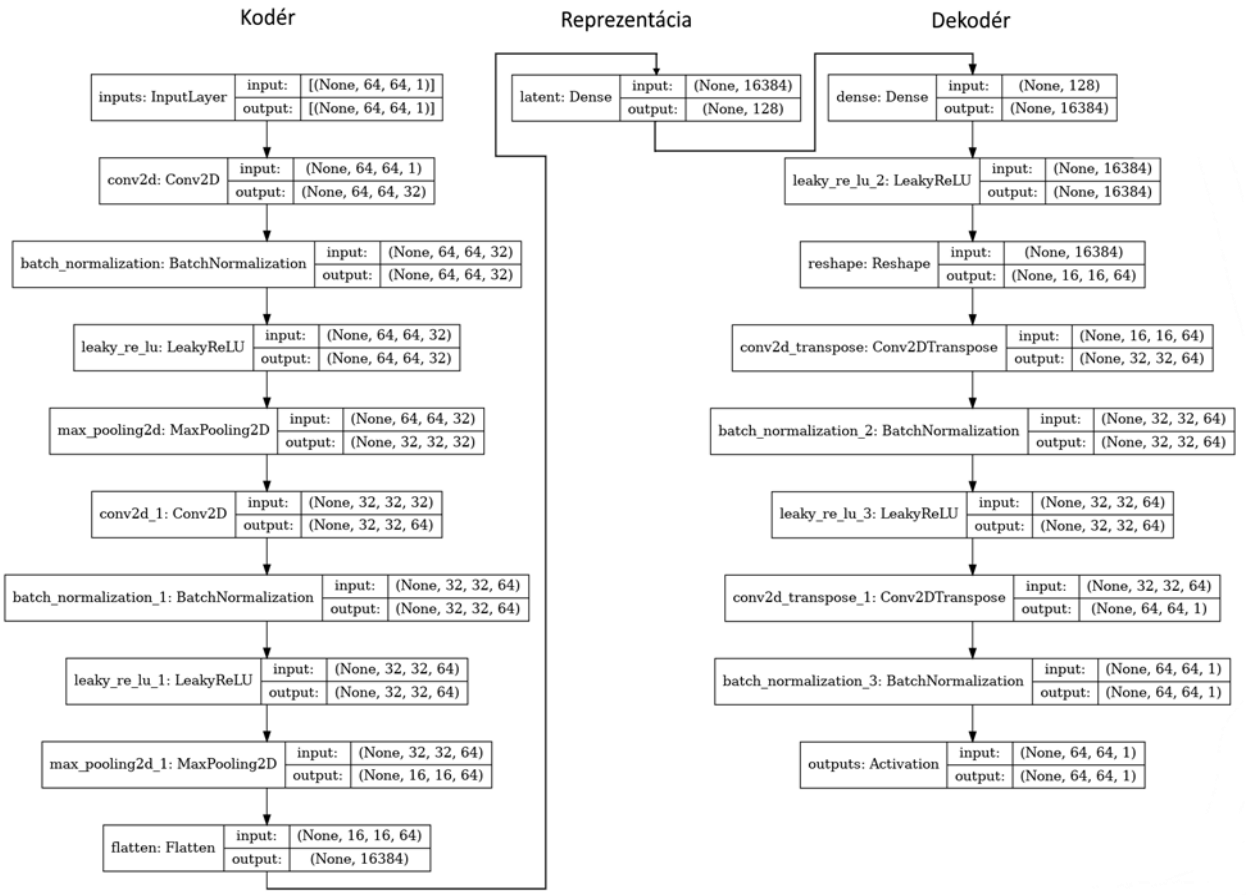
Ideou experimentu je pomocou autokodéra získať zakódovanú reprezentáciu snímok zo strednej vrstvy siete, nad nimi vykonať zhlukovanie pomocou zhlukovacích algoritmov a následne z vybraných zhlukov reprezentácií získať pomocou autokodéra späť pôvodné snímky. Tie budú použité ako trénovacia množina. Na obrázku 17 je znázornené schematické zobrazenie navrhovaného riešenia.



Obr. 17 Schematické zobrazenie výberu trénovacej množiny pomocou autokodérov

Autokodéry patria medzi najvýznamnejšie algoritmy učenia bez učiteľa [Min18]. Ich úlohou je natrénovať mapovaciu funkciu, pre ktorú bude chybovosť rekonštrukcie medzi kodérom a originálnymi dátami čo najmenšia.

Zvolili sme architektúru konvolučných autokodérov, kde každý neurón v konvolučnej vrstve je spojený iba s niekoľkými susediacimi neurónmi z predošlej vrstvy a každý neurón má rovnakú sadu váh [Min18]. Na obrázku 18 je znázornená architektúra, ktorú sme použili. Stredná vrstva siete – reprezentácia, ktorú používame v zhlukovaní predstavuje vektor dĺžky 128.



Obr. 18 Použitá architektúra autokodéra

Na testovanie sme použili databázu LFWcrop. Obsahuje orezané snímky z databázy LFW [Sand09]. Snímky boli prevedené do odtieňov šedi. Pre zhlukovanie sme vybrali iba subjekty, ktoré mali aspoň 20 snímok, aby bol možný samotný výber trénovacej množiny. Výsledkom je množina 62 subjektov po 20 snímkach. Autokodér bol natrénovaný pomocou trénovacej a validačnej množiny v pomere 2:3, pričom vyhodnocovanie prebiehalo s použitím celej množiny dostupných snímok. Na obrázku 19 je znázornená ukážka dvoch vybraných subjektov.

Obr. 19 Ukážka dvoch subjektov z databázy LFWcrop v odtieňoch šedi

Na zhlukovanie sme použili algoritmy SOM a K-means. Algoritmus SOM dosahoval v priemere najlepšie výsledky v našich doterajších experimentoch, preto sme chceli pokračovať s jeho overením aj v tomto prístupe. V práci [Kari21] sa zase uvádza K-means ako jeden z najlepších zhlukovacích algoritmov v spojení s autokodérmi.

Na subjekt sme mali k dispozícii 20 snímok. Vyberali sme v troch rôznych počtoch výberov. Z dôvodu využitia algoritmu SOM, ktorý má počet zhlukov daný vždy ako maticu X*Y neurónov (2*3, 3*3, 3*4 a pod.) sme zvolili nasledovné rozdelenia:

- 9 snímok na trénovanie a zbytok na testovanie,
- 12 snímok na trénovanie a zbytok na testovanie,
- 16 snímok na trénovanie a zbytok na testovanie.

Opäť sme overovali aj prínos vhodného predspracovania na snímky za účelom lepšieho výberu trénovacích vzoriek. Použili sme iba algoritmus CLAHE a nie histogramovú ekvalizáciu ako takú, keďže sa v predošlých experimentoch potvrdila jeho výhoda oproti klasickej histogramovej ekvalizácii.

Úspešnosť sme overili použitím knižnice DeepFace (https://github.com/serengil/deepface). Je to aplikačný rámec využívajúci najmodernejšie algoritmy pre rozpoznávanie podľa tváre a na analýzu tvárových vlastností. Na rozpoznávanie sme vybrali model VGG-Face, s ktorým sme dosahovali najlepšie výsledky. Vstupné snímky pri trénovaní a testovaní neboli žiadnym spôsobom predspracované.

Navrhnuté postupy výberu:

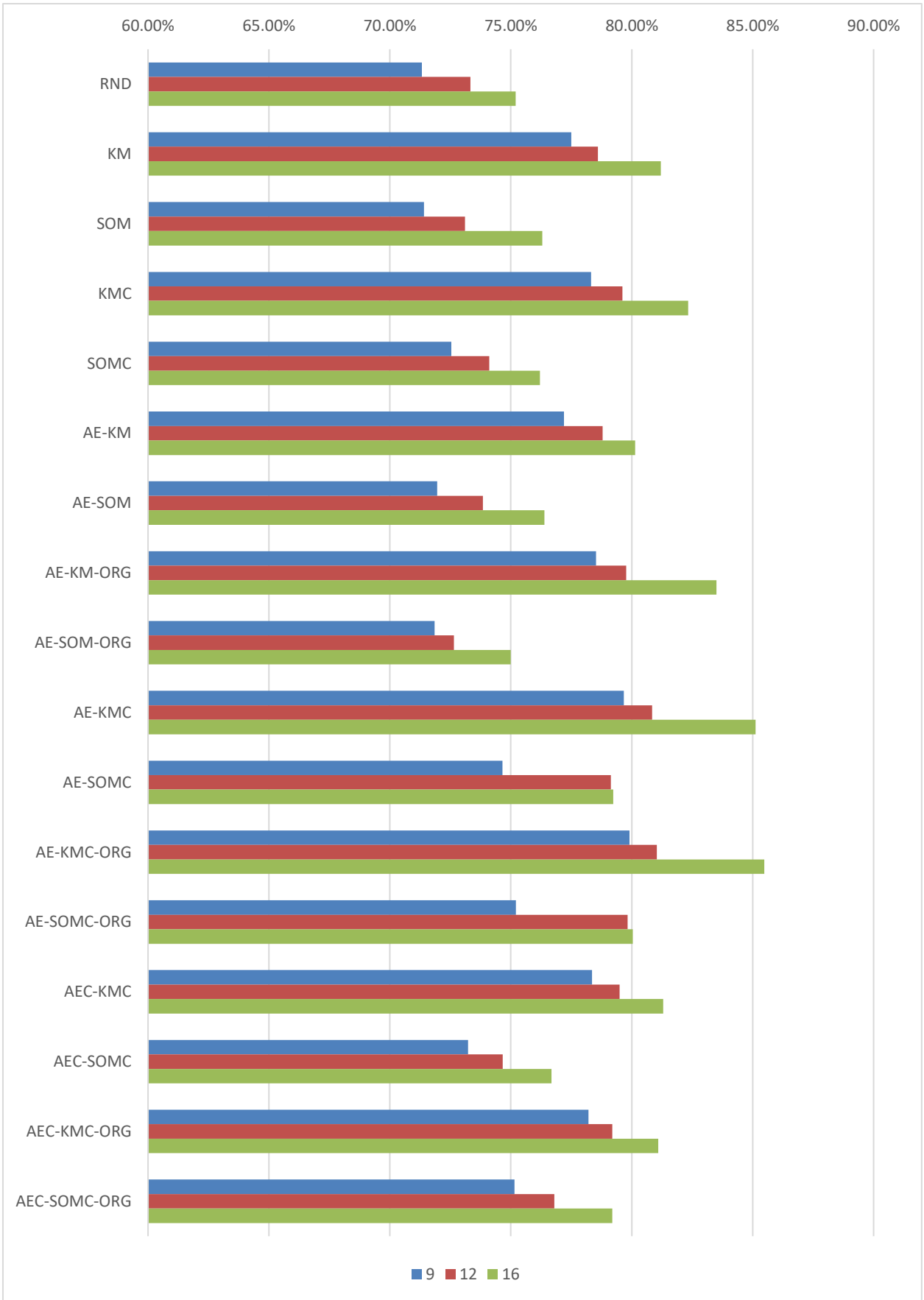
- ***Náhodný výber – RND***
    - Päť náhodných výberov
    - Vyhodnotenie na VGG-Face rozpoznávaní
    - Priemer všetkých výsledkov predstavuje celkový výsledok (RND)
- ***Výber vzoriek pomocou zhlukovania – KM, SOM***
    - Výber trénovacej množiny pomocou zhlukovacích algoritmov (K-means, SOM)
    - Vyhodnotenie na VGG-Face rozpoznávaní
- ***Výber pomocou zhlukovania s predspracovaním CLAHE – KMC, SOMC***
    - Predspracovanie snímok subjektu pomocou CLAHE
    - Výber trénovacej množiny pomocou zhlukovacích algoritmov (K-means, SOM)
    - Vyhodnotenie na VGG-Face rozpoznávaní
- ***Výber pomocou autokodéra s použitím centier – AE-KM, AE-SOM***
    - Natrénovanie autokodéra na snímkach LFWcrop databázy

- Zakódovanie snímok subjektu
- Aplikácia zhlukovania (K-means, SOM) na zakódovaných snímkach a získanie centier zhlukov
- Dekódovanie centier zhlukov na dekódované snímky
- Nájdenie najbližších originálnych snímok k dekódovaným snímkam s použitím euklidovskej vzdialenosti
- Vyhodnotenie na VGG-Face rozpoznávaní

- ***Výber pomocou autokodéra s použitím originálnych snímok – AE-KM-ORG, AE-SOM-ORG***
  - Natrénovanie autokodéra na snímkach LFWcrop databázy
  - Zakódovanie snímok
  - Aplikácia zhlukovania (K-means, SOM) na zakódovaných snímkach a získanie centier zhlukov
  - Nájdenie najbližších zakódovaných snímok k centrám zhlukov
  - Dekódovanie vybraných zakódovaných snímok na dekódované snímky
  - Nájdenie najbližších originálnych snímok k dekódovaným snímkam s použitím euklidovskej vzdialenosti
  - Vyhodnotenie na VGG-Face rozpoznávaní
- ***Výber pomocou autokodéra s použitím centier a predspracovaním CLAHE pri zakódovaní – AE-KMC, AE-SOMC***
  - Natrénovanie autokodéra na snímkach LFWcrop databázy
  - Predspracovanie snímok subjektu pomocou CLAHE
  - Zakódovanie snímok subjektu
  - Aplikácia zhlukovania (K-means, SOM) na zakódovaných snímkach a získanie centier zhlukov
  - Dekódovanie centier zhlukov na dekódované snímky
  - Nájdenie najbližších originálnych snímok k dekódovaným snímkam s použitím euklidovskej vzdialenosti
  - Vyhodnotenie na VGG-Face rozpoznávaní
- ***Výber pomocou autokodéra s použitím originálnych snímok a predspracovaním CLAHE pri zakódovaní – AE-KMC-ORG, AE-SOMC-ORG***
  - Natrénovanie autokodéra na snímkach LFWcrop databázy
  - Predspracovanie snímok subjektu pomocou CLAHE
  - Zakódovanie snímok subjektu

    - Aplikácia zhlukovania (K-means, SOM) na zakódovaných snímkach a získanie centier zhlukov
    - Nájdenie najbližších zakódovaných snímok k centrám zhlukov
    - Dekódovanie vybraných zakódovaných snímok na dekódované snímky
    - Nájdenie najbližších originálnych snímok k dekódovaným snímkam s použitím euklidovskej vzdialenosti
    - Vyhodnotenie na VGG-Face rozpoznávaní
- ***Výber pomocou autokodéra s použitím centier a predspracovaním CLAHE pri trénovaní aj zakódovaní – AEC-KMC, AEC-SOMC***
    - Predspracovanie všetkých snímok pomocou CLAHE
    - Natrénovanie autokodéra na snímkach LFWcrop databázy
    - Zakódovanie snímok subjektu
    - Aplikácia zhlukovania (K-means, SOM) na zakódovaných snímkach a získanie centier zhlukov
    - Dekódovanie centier zhlukov na dekódované snímky
    - Nájdenie najbližších originálnych snímok k dekódovaným snímkam s použitím euklidovskej vzdialenosti
    - Vyhodnotenie na VGG-Face rozpoznávaní
- ***Výber pomocou autokodéra s použitím originálnych snímok a predspracovaním CLAHE pri trénovaní aj zakódovaní – AEC-KM-ORG, AEC-SOM-ORG***
    - Predspracovanie všetkých snímok pomocou CLAHE
    - Natrénovanie autokodéra na snímkach LFWcrop databázy
    - Zakódovanie snímok subjektu
    - Aplikácia zhlukovania (K-means, SOM) na zakódovaných snímkach a získanie centier zhlukov
    - Nájdenie najbližších zakódovaných snímok k centrám zhlukov
    - Dekódovanie vybraných zakódovaných snímok na dekódované snímky
    - Nájdenie najbližších originálnych snímok k dekódovaným snímkam s použitím euklidovskej vzdialenosti
    - Vyhodnotenie na VGG-Face rozpoznávaní

Všetky navrhnuté postupy sme zrealizovali a výsledky uviedli v tabuľke 5 a v grafe na obrázku 20. Uvedená tabuľka a graf predstavujú dosiahnutú úspešnosť pre výber 9, 12 a 16 snímok. Zvýraznený riadok zodpovedá najlepšiemu dosiahnutému výsledku.



Obr. 20 Úspešnosť rozpoznávania pre všetky konfigurácie a pre výber 9, 12 a 16 snímok.

Metóda	9	12	16
RND	71,32 %	73,33 %	75,20 %
KM	77,50 %	78,60 %	81,20 %
SOM	71,41 %	73,11 %	76,30 %
KMC	78,32 %	79,61 %	82,33 %
SOMC	72,54 %	74,11 %	76,20 %
AE-KM	77,20 %	78,80 %	80,14 %
AE-SOM	71,96 %	73,85 %	76,39 %
AE-KM-ORG	78,52 %	79,77 %	83,50 %
AE-SOM-ORG	71,85 %	72,65 %	75,00 %
AE-KMC	79,67 %	80,84 %	85,12 %
AE-SOMC	74,65 %	79,14 %	79,23 %
AE-KMC-ORG	79,91 %	81,04 %	85,48 %
AE-SOMC-ORG	75,21 %	79,83 %	80,04 %
AEC-KMC	78,21 %	79,20 %	81,10 %
AEC-SOMC	73,23 %	74,66 %	76,68 %
AEC-KMC-ORG	78,36 %	79,50 %	81,30 %
AEC-SOMC-ORG	75,15 %	76,80 %	79,20 %

Tab. 5 Úspešnosť rozpoznávania pre všetky konfigurácie a pre výber 9, 12 a 16 snímok.

Najlepší výsledok sme dosiahli pre postup AE-KMC-ORG, pri všetkých počtoch výberu snímok. Autokodér bol natrénovaný na neupravených snímkach (bez CLAHE), kódovanie a následné zhlukovanie bolo vykonané na predspracovaných pomocou CLAHE algoritmom K-means a do dekodéra boli vložené originálne reprezentácie. Pre výber 9 snímok sme dosiahli úspešnosť rozpoznávania 79,91 % čo je oproti náhodnému výberu, pri ktorom sme dosiahli úspešnosť 71,32 %, až o 8,59 % viac. Pre výber 12 snímok bola úspešnosť 81,04 % a pre výber 16 snímok 85,48 %, kde v oboch prípadoch bola úspešnosť oproti náhodnému výberu vyššia v priemere o 8 %.

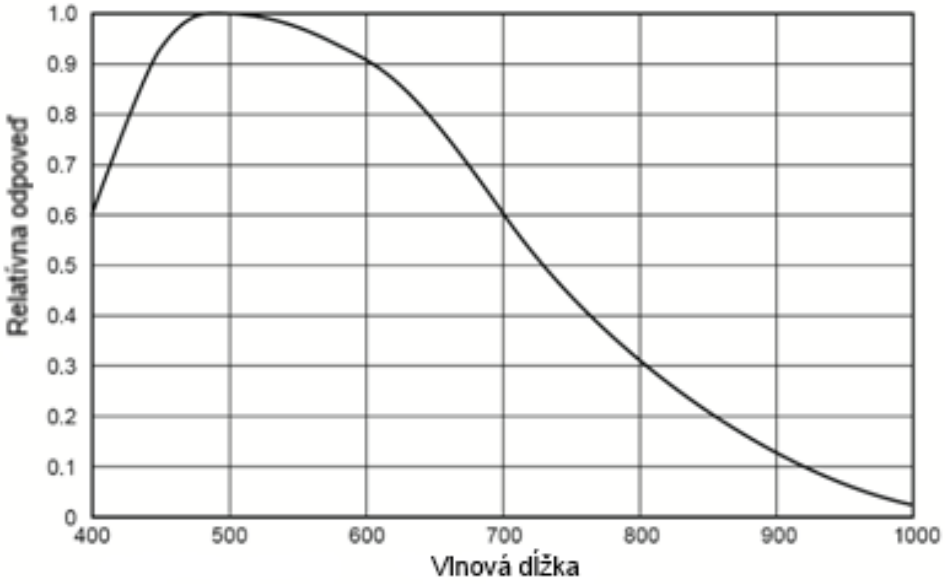
Zaujímavosťou je nižšia úspešnosť pre postupy, pri ktorých bol autokodér natrénovaný snímkami predspracovanými pomocou CLAHE. Rozdiel medzi najlepšou dosiahnutou úspešnosťou v týchto postupoch (AEC-KMC-ORG – 81,3 %) oproti najúspešnejšiemu postupu (AE-KMC-ORG) predstavuje 4,18 %. Predspracovanie CLAHE napriek tomu zlepšuje samotnú úspešnosť zhlukovania, ako je možné vidieť v porovnaní jednotlivých postupov, kde napr. AE-KMC-ORG dosahuje lepšie výsledky ako AE-KM-ORG (rozdiel v priemere 2 %), rovnako ako napr. pre AE-KMC verzus AE-KM (rozdiel v priemere 2 % až 5 %).

Z celkových výsledkov tohto experimentu môžeme prehlásiť, že použitie autokodérov v spojení so zhlukovaním zvyšuje úspešnosť rozpoznávania, preto je vhodné na toto využitie. Pritom najlepšie výsledky sa dosiahnu zhlukovacím algoritmom K-means s predspracovaním pomocou

CLAHE. SOM algoritmus v spojení s autokodérom a pre túto databázu sa javí ako nie veľmi vhodný. Výsledky tohto experimentu neboli ešte v čase odovzdávania tejto práce publikované.

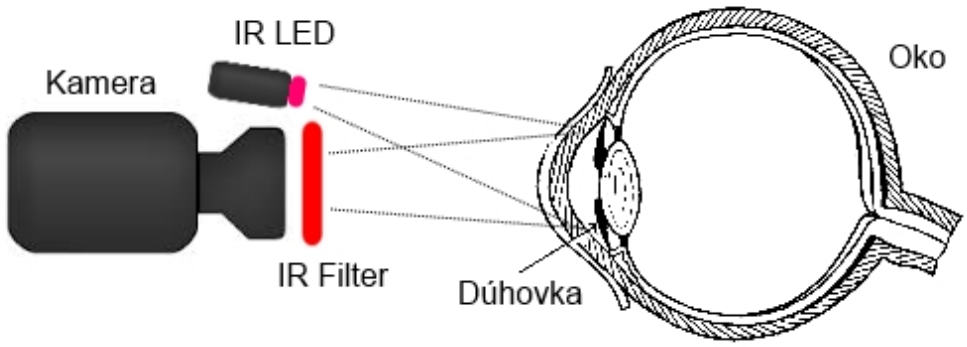
3.8 Systém snímania obrazov dúhovky

Nami zostavený systém na snímanie obrazov dúhoviek je založený na kamere Guppy F-038B NIR od spoločnosti Allied Vision Technologies. Táto kamera je citlivá v infračervenom spektre, pri ktorom je štruktúra dúhovky najlepšie pozorovateľná.



Obr. 21 Spektrálna citlivosť kamery Guppy F038B NIR [Gupp13]

Kamera je v spojení s objektívom značky Tamron M118FM25. Jeho zaostrenie je od 0,1 m až do nekonečna. To je postačujúce rozmedzie na získanie detailného záberu. Má manuálne zaostrovanie a clonu. Keďže pracujeme v spektre NIR, na dosiahnutie najlepších záberov je na objektíve nainštalovaný infračervený filter (šírka pásma 260 nm, kde centrum je 830 nm) a na kamere infračervené osvetlenie (najvyššia emisia pri 880 nm). Na obrázku 22 je zobrazený schematický návrh systému snímania dúhoviek, na obrázku 23 je zobrazená použitá kamera a objektív a na obrázku 24 celý zostavený systém.

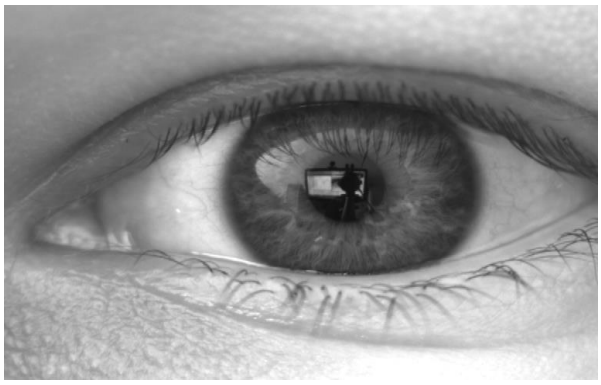


Obr. 22 Schematické zobrazenie systému snímania dúhovky

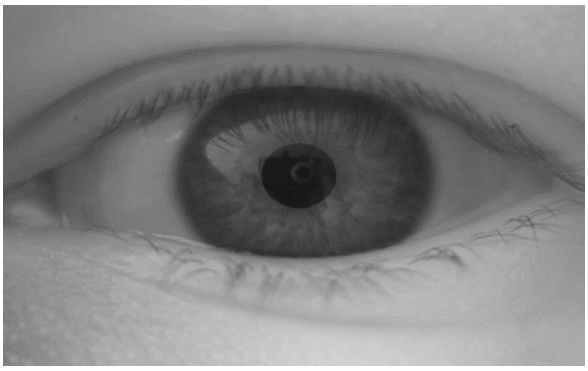
Obr. 23 Kamera Guppy F-038B NIR [Gupp13] a objektív Tamron M118FM25 [Tam13]

Obr. 24 Systém na zachytenie snímok dúhoviek

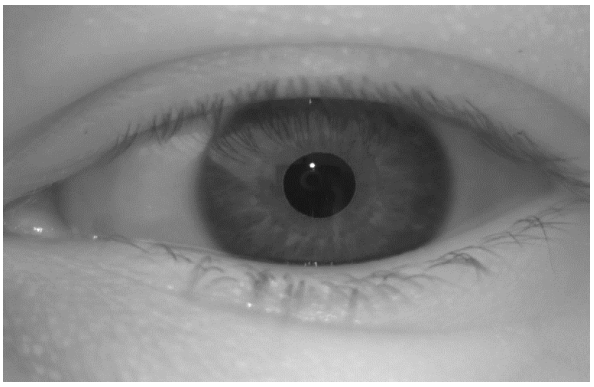
Pre porovnanie možností snímania systému sú na obrázkoch 25 až 27 zobrazené tri snímky dúhoviek. Prvá snímka je zachytená iba kamerou. Na druhej snímke je použitý infračervený filter a na tretej snímke je zapnuté aj infračervené osvetlenie.



Obr. 25 Dúhovka nasnímaná kamerou

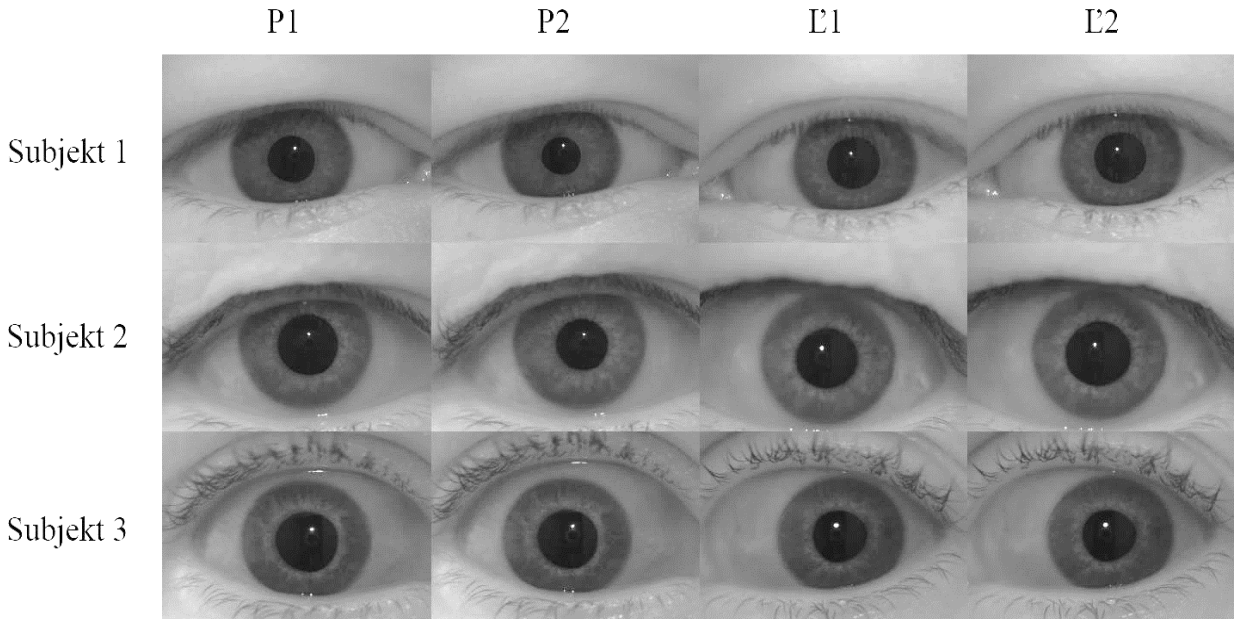


Obr. 26 Dúhovka nasnímaná kamerou s infračerveným filtrom



Obr. 27 Dúhovka nasnímaná kamerou s infračerveným filtrom a osvetlením

Zo snímok je zrejmé, že štruktúra dúhovky je najzreteľnejšia pri použití IR filtra a súčasne IR osvetlenia, vďaka ktorým bola odstránená aj väčšina nežiaducich odleskov. Na obrázku 28 je znázornený výber niekoľkých snímok, ktoré sme naším systémom zachytili. Nami zostrojený systém sme publikovali na konferencii REDŽUR 2014 [V4].



Obr. 28 Ukážka snímok dúhoviek

3.9 Rozpoznávanie ľudského ucha na mobilnom zariadení

Ako plynulým pokračovaním našej práce na mobilných zariadeniach a so značným nárastom ich výpočtovej kapacity sme sa zamerali na možnosti využitia multimodálneho rozpoznávania podľa tváre a ucha. Ucho patrí medzi biometriky, ktoré je možné úspešne použiť pri rozpoznávaní [Rah07]. V tomto výskume sme zostavili prvotnú verziu mobilnej aplikácie, ktorá pracuje s touto biometrikou.

Obr. 29 Ukážka snímok uší z databázy stiahnutých uší z vyhľadávača Google

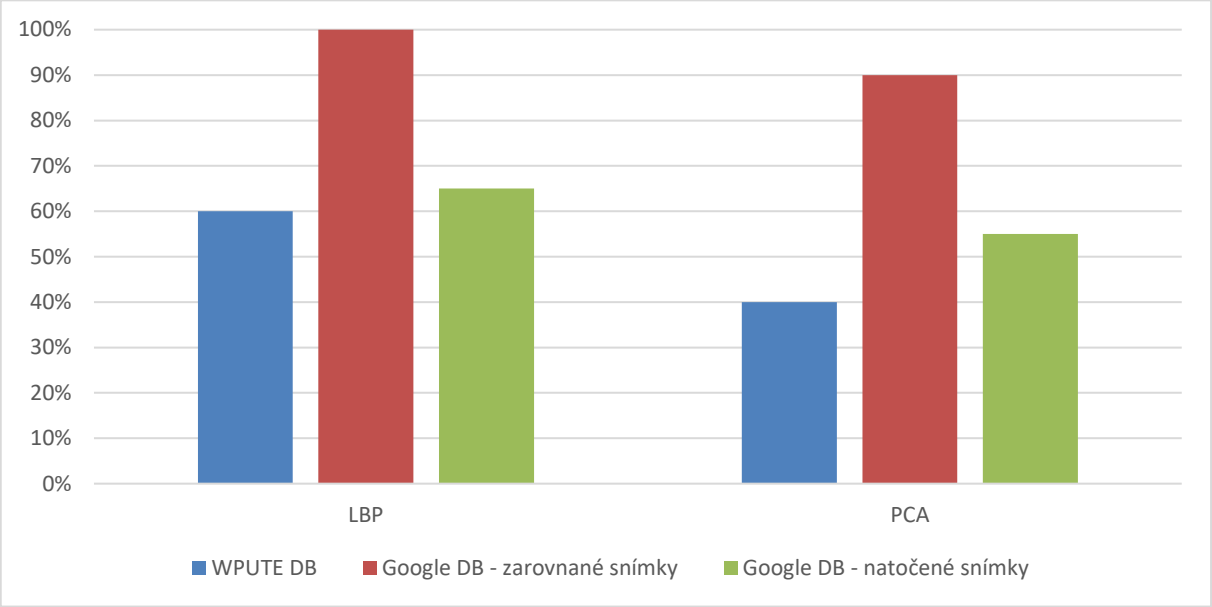
Obr. 30 Ukážka snímok z databázy WPUTE

Na testovanie sme použili dve databázy. Prvou bola naša zozbieraná databáza zo snímok voľne dostupných na internete. Snímky sme vyhľadávali pomocou vyhľadávača Google. Pozostáva z dvadsiatich snímok pravého ucha rôznych subjektov. Uši neboli ničím prekryté, ale rozlíšenie, poloha či natočenie sa odlišovali, vďaka čomu vieme do určitej miery touto databázou simulovať neriadené podmienky. Ukážka snímok z našej databázy je zobrazená na obrázku 29.
Druhou použitou databázou bola databáza WPUTE [Frej10]. V tomto prípade sa jedná o kontrolovanú databázu. Snímky majú rozlíšenie 380*500 pixelov, obsahuje 4 snímky na osobu, pričom niekedy sú časti uší prekryté. Ukážka z databázy je zobrazená na obrázku 30.
Podobne ako pri rozpoznávaní tváre, prvým krokom bolo predpracovanie snímok. Úlohou bolo získanie snímky, ktorú je možné použiť v procese rozpoznávania. Jednotlivým snímkam sme znížili rozlíšenie, previedli sme ich do odtieňov šedi a vykonali na nich histogramovú ekvalizáciu. Na detekciu oblasti ucha sme použili algoritmy Haar kaskády a LBP. Následne sme snímky orezali na oblasti ucha a znormalizovali rozlíšenie na veľkosť 80*160 pixelov.
Vykonali sme dve série testovaní, kde v prvom prípade sme príznaky extrahovali použitím algoritmu LBP a klasifikácia bola vykonaná pomocou metriky vzdialenosti Chi-square. V druhej sérii sme na extrakciu použili algoritmus PCA a metriku euklidovskej vzdialenosti na klasifikáciu. Vstupné dáta pre jednotlivé série testovaní sa odlišovali použitými databázami a spôsobom úpravy snímok:

- Databáza WPUTE s dvadsiatimi subjektami, kde každý subjekt mal tri snímky použité na natrénovanie a jednu snímku na testovanie (WPUTE DB)
- Zozbieraná databáza z Google s dvadsiatimi subjektami s rovnakou snímkou použitou na trénovanie aj testovanie bez natočenia (Google DB – zarovnané snímky)
- Zozbieraná databáza z Google s dvadsiatimi subjektami s rovnakou snímkou použitou na trénovanie aj testovanie s natočením 7-10 stupňov (Google DB – natočené snímky)

Výsledky testovania sú znázornené v grafe na obrázku 31. Metóda extrakcie príznakov pomocou LBP dosahovala všeobecne vyššiu úspešnosť ako metóda PCA, pri nenatočených snímkach a s použitím identických snímok sme dosiahli úspešnosť 100 %. Preto považujeme metodológiu za správnu. S použitím databázy WPUTE boli výsledky značne nižšie, kde pri metóde PCA klesla

úspešnosť pod 50 %. V prípade LBP bola úspešnosť okolo 60 %, čo môžeme pripísať horšej kvalite snímok z tejto databázy. Výsledky dosiahnutých úspešností rozpoznávania sme publikovali na konferencii IWSSIP 2016 [V8].



Obr. 31 Úspešnosť rozpoznávania ucha pomocou LBP a PCA algoritmov

4. Splnenie cieľov dizertačnej práce

1. Návrh metód s využitím strojového učenia na definíciu vhodnej trénovacej množiny pre rozpoznávanie podľa tváre tak, aby bola maximalizovaná vnútro-triedna variabilita.

Pre potrebu definície trénovacej množiny sme navrhli použitie zhlukovacích algoritmov. Dokázali sme použiteľnosť tohto návrhu na viacerých experimentoch, kde sa oproti náhodnému výberu snímok zvýšila úspešnosť rozpoznávania o 10 % až 15 %. Výsledky prvotného výberu pomocou zhlukovacích algoritmov sme uverejnili v publikáciách [V2] a [V4]. Výsledkom týchto experimentov bol záver, že použitie SOM považujeme za vhodné riešenie výberu trénovacej množiny. Zároveň sme sa zamerali na problematiku odstránenia nežiaducich účinkov variability osvetlenia. Vo viacerých vykonaných experimentoch sme použili histogramovú ekvalizáciu na jej zmiernenie. Overili sme vplyv použitia histogramovej ekvalizácie a adaptívnej histogramovej ekvalizácie na úspešnosť výberu trénovacej množiny a výslednú úspešnosť rozpoznávania. Zlepšenie úspešnosti bolo v rozmedzí 9 % až 10 % pri použití adaptívnej histogramovej ekvalizácie oproti klasickej histogramovej ekvalizácii alebo bez použitia ekvalizácie. Výsledky sme uverejnili v publikáciách [V2] a [V4].

V ďalších experimentoch sme otestovali a vyhodnotili úspešnosť rozpoznávania s použitím algoritmov SOM, K-means, C-means, DBSCAN. Hlavným algoritmom skúmania v prvých experimentoch sa stala neurónová sieť SOM. Dosahovali sme najlepšie výsledky úspešnosti práve s jej použitím, dosahujúc úspešnosti okolo 80 % (82,13 % a 81,33 % pre systémy založené na PCA a SVM nad databázou CMU-PIE a 86,73 % a 71,12 % pre systém založený na LBP a Manhattan metrike nad databázami CMU-PIE a EURECOM). C-means dosahoval úspešnosť v priemere o 3 % nižšiu pri použití databázy CMU-PIE, ale pri použití databázy EURECOM bola úspešnosť vyššia ako pre SOM o 0,9 %. Pripisujeme to nevyužitiu adaptívnej histogramovej ekvalizácie v danom experimente. Všetky ostatné algoritmy mali nižšiu úspešnosť ako neurónová sieť SOM. DBSCAN dosahoval pri použití databázy CMU-PIE úspešnosť nižšiu takmer o 4 % a pre databázu EURECOM o 1 %. K-means pre databázu CMU-PIE dosahoval úspešnosť nižšiu o 6 % a pre databázu EURECOM o takmer 3 %. Najhoršia úspešnosť dosiahnutá použitím náhodného výberu bola nižšia o 13 % až 15 % pre databázu CMU-PIE a 10 % pre databázu EURECOM. V publikáciách [V5] a [V7] sme uverejnili dosiahnuté úspešnosti jednotlivých systémov.

V záverečnom experimente sme sa zamerali na overenie úspešnosti rozpoznávania s použitím hlbokého konvolučného autokodéra. Pre databázu LFWcrop sme s použitím zhlukovania algoritmom K-means a rozpoznávaním pomocou VGG-Face algoritmu dosiahli zlepšenie rozpoznávania oproti náhodnému výberu o 8,59 % pre výber 9 trénovacích snímok, 7,71 % pre výber 12 snímok a 10,28 % pre výber 16 trénovacích snímok. Uskutočnili sme množstvo rôznych konfigurácií systému výberu trénovacích snímok, kde sme zistili ich výsledné úspešnosti. Najúspešnejšou sa stala konfigurácia AE-KMC-ORG pozostávajúca z krokov:

- Natrénovanie autokodéra na snímkach LFWcrop databázy
- Predspracovanie snímok subjektu pomocou CLAHE
- Zakódovanie snímok subjektu
- Aplikácia zhlukovania (K-means, SOM) na zakódovaných snímkach a získanie centier zhlukov
- Nájdenie najbližších zakódovaných snímok k centrám zhlukov
- Dekódovanie vybraných zakódovaných snímok na dekódované snímky
- Nájdenie najbližších originálnych snímok k dekódovaným snímkam s použitím euklidovskej vzdialenosti
- Vyhodnotenie na VGG-Face rozpoznávaní

Ďalšie konfigurácie sú popísané v kapitole 3.7. Výber snímok sme porovnali aj s postupmi navrhnutými v predošlých experimentoch. Z dosiahnutých výsledkov je zrejmé, že použitím autokodérov a zhlukovania pomocou K-means je možné dosiahnuť najlepšie výsledky. Zároveň je zrejmé, že pri algoritmoch založených na hlbokom učení je rozsah trénovacej množiny úmerný dosiahnutej úspešnosti.

2. Zameranie sa na rozpoznávanie podľa tváre v mobilných zariadeniach.

Nami navrhované postupy je možné úspešne použiť na dnešných výkonných mobilných zariadeniach bez väčších problémov. Pre prípady slabších zariadení sme navrhli použitie klient server architektúry, kde klient (mobilné zariadenie) má za úlohu získavanie snímok, automatický výber snímok pre trénovanie a extrakciu príznakov. Server (vzdialené zariadenie) má za úlohu ukladať vzory poslané klientom a klasifikáciu vzorov za účelom rozpoznávania. Navrhovanú architektúru sme uverejnili v publikácii [V6].

3. Zameranie sa na rozpoznávanie podľa tváre v reálnom čase.

V nami navrhnutej architektúre sme vďaka oddeleniu výpočtovo náročných operácií zabezpečili využitie v reálnom čase aj na mobilných zariadeniach. Súčasne vďaka obmedzenému výberu snímok je zabezpečená dostatočná rýchlosť algoritmov trénovania a rozpoznávania. Ako sme uviedli vyššie z výsledkov našich experimentov s využitím hlbokého učenia, je úspešnosť týchto systémov úmerná rozsahu trénovacej množiny. S jej rastom ale rastie aj výpočtový čas, najmä na trénovanie. Je preto vhodné zvoliť si optimálny kompromis medzi veľkosťou trénovacej množiny a časovej náročnosti. Funkčnosť navrhnutého konceptu aj s reálnym využitím zhlukovacích algoritmov sme overili a výsledky úspešností sme uverejnili v publikácii [V7]. V tejto publikácii sme ale ešte nepoužili postupy s hlbokým učením.

4. Vytvorenie vhodného systému snímania obrazov očných dúhoviek, na minimalizáciu neriadených podmienok osvetlenia.

Nami zostrojený systém snímania obrazov očných dúhoviek sa skladá zo špeciálnej kamery citlivej v infračervenom spektre, infračerveného filtra a osvetlenia a pomocných súčasti ako stojan, zdroj a pod. Porovnali sme kvalitu snímok jednak bez použitia infračerveného filtra, ako aj s filtrom a rovnako aj s osvetlením. Výsledné obrazy dúhoviek dosahovali vysokú kvalitu a dosiahli sme výrazné odfiltrovanie nepriaznivých vplyvov neriadeného osvetlenia. Náš systém sme uverejnili v publikácii [V3].

Z doteraz vykonaných experimentov a publikovaných výsledkov je zrejmé, že zhlukovanie ako také je vhodnou metódou výberu trénovacej množiny. Výsledná úspešnosť bola závislá na použitej databáze. Pre databázy CMU-PIE a EURECOM sme dokázali, že najvyššiu úspešnosť dosiahneme algoritmom SOM. Javí sa preto ako vhodný algoritmus na automatizovaný výber trénovacích vzoriek. Naproti tomu pre databázu LFWcrop s použitím autokodéra a algoritmu K-means sme dokázali, že práve ich využitím je možné dosiahnuť výrazné zlepšenie úspešnosti aj oproti algoritmu SOM. Súčasne sme overili prínos použitia adaptívnej histogramovej ekvalizácie na úpravu snímok pred samotným výberom. S jej použitím bola dosiahnutá úspešnosť zakaždým najvyššia. Vo výsledkoch experimentu publikovaných v [V7] sme z dôvodu zjednodušenia nemali implementovanú metódu adaptívnej histogramovej ekvalizácie, z čoho aj usudzujeme horší výsledok pre sieť SOM oproti algoritmu C-means nad databázou EURECOM.

Navrhovaný koncept sme overili aj s použitím navrhnutej architektúry pre mobilné zariadenia. Možnými smermi ďalšieho výskumu by bola samotná reálna implementácia v mobilnom zariadení a overenie na viacerých zariadeniach, s použitím adaptívnej histogramovej ekvalizácie a na reálnych snímkach zo zariadení.

Využiteľnosť navrhnutého riešenia vidíme pre akýkoľvek systém rozpoznávania založený na metódach pracujúcich v reálnom čase. Využitím zhlukovania akoukoľvek metódou a adaptívnej histogramovej ekvalizácie sa preukázateľne zvýšila úspešnosť rozpoznávania.

V oblasti rozpoznávania podľa očnej dúhovky sme sa zaoberali vytvorením vhodného snímacieho zariadenia, systému, vďaka ktorému by sme vedeli znížiť vplyv neriadených podmienok osvetlenia. Pomocou nášho systému sme boli schopní vďaka použitiu infračerveného osvetlenia a infračerveného filtra dosiahnuť vysokú kvalitu zosnímaných obrazov dúhoviek. S jeho použitím je možné zozbieranie vhodnej referenčnej databázy obrazov dúhoviek, použiteľnej na ďalší výskum a súčasne kvantifikáciu prínosu úspešnosti za použitia jednotlivých súčastí systému.

Riešiteľ projektov

[P1] VEGA 1/0529/13 Návrh pokročilých metód biometrického rozpoznávania na základe obrazov tváre a dúhovky, 2013-2016

[P2] BioDaT, Biometrické rozpoznávanie na základe obrazov Dúhovky a Tváre, 2013

[P3] MoBiFaIR, Metódy biometrického rozpoznávania tvárí a dúhoviek, Programu na podporu mladých výskumníkov 2014, vedúci projektu: Ing. Vojtěch Jirka

[P4] Spolupráca na projekte „Hybrid Broadcast Broadband Next Generation“ FP7-ICT-2011-7-287848

Publikácie autora

[V1] Jirka V.: E - learningová multimediálna aplikácia Biometria, Bakalárska práca. Bratislava: STU FEI, 2010. 64 s.

[V2] Ban, J., Féder, M., Jirka, V., Loderer, M., Omelina, Ľ., Oravec, M., Pavlovičová, J.: An Automatic Training Process Using Clustering Algorithms for Face Recognition System, 55th International Symposium ELMAR-2013, 25-27 September 2013, Zadar, Croatia, pp. 15-18, ISBN 978-953-7044-14-5

[V3] Jirka V., Oravec M.: Robust Iris Imaging System for Less Constrained Iris Recognition, REDŽÚR 2014, 8th International Workshop on Multimedia and Signal Processing, May 13, 2014, Dubrovnik, Croatia

[V4] Jirka V., Oravec M., Pavlovičová J., Féder M.: Experiments with Automatic Training Samples Selection Using Self-organizing Map, REDŽÚR 2014, 8th International Workshop on Multimedia and Signal Processing, May 13, 2014, Dubrovnik, Croatia

[V5] Jirka, V., Féder, M., Pavlovičová, J., Oravec, M.: Face Recognition System with Automatic Training Samples Selection Using Self-organizing Map, 56th International Symposium ELMAR-2014, 10-12 September 2014, Zadar, Croatia, pp. 23-26, ISBN 978-953-184-199-3

[V6] Oravec, M., Sopiak, D., Jirka, V.: Face recognition on mobile devices based on client-server architecture, REDŽÚR 2015, 9th International Workshop on Multimedia and Signal Processing, 22-23 April 2015, Smolenice, Slovakia, pp.15-18, ISBN 978-80-227-4346-4

[V7] Oravec, M., Sopiak, D., Jirka, V., Pavlovičová , J., Budiak, M.: Clustering algorithms for face recognition based on client-server architecture, In 22th International Conference on Systems, Signals and Image Processing IWSSIP 2015. 10-12 September, London, UK, pp. 241-244, ISBN 978-1-4673-8352-3

[V8] Oravec M., Pavlovičová J., Sopiak D., Jirka V., Loderer M., Lehota L., Vodička M., Fačkovec M., Mihalik M., Tomík M., Gerát J.: Mobile ear recognition application, 2016 International Conference on Systems, Signals and Image Processing (IWSSIP), Bratislava, 2016, pp. 1-4, doi: 10.1109/IWSSIP.2016.7502719.

Zoznam použitej literatúry

[Brad00] Bradski, G.: The OpenCV Library. Dr. Dobb`s Journal of Software Tools, 2000

[CMU02] Carnegie Mellon University - The Robotics Institute, Pose, Illumination, and Expression, CMU-PIE Database, 2002
[http://www.cs.cmu.edu/

[Ester96] Ester M., et al.: "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," In Second International Conference on Knowledge Discovery and Data Mining, ISSN 09758887, vol. 2, p. 226-231, 1996

[Frej10] Frejlichowski D., Tyszkiewicz N., The West Pomeranian University of Technology Ear Database – A Tool for Testing Biometric Algorithms. 2010, 6112. 227-234. 10.1007/978-3-642-13775-4_23

[Gupp13] Guppy F-038B NIR Datasheet, Allied Vission Technologies, dátum prístupu 2013

[Min18] Min E., Guo X., Liu Q., Zhang G., Cui J., Long J.: A Survey of Clustering With Deep Learning: From the Perspective of Network Architecture, in IEEE Access, vol. 6, pp. 39501-39514, 2018, doi: 10.1109/ACCESS.2018.2855437

[Mull01] Müller K.R., Mika S., Rätsch G., Tsuda K., Scholkopf B.: An Introduction to Kernel-Based Learning Algorithms, IEEE Transactions on neural networks, Vol. 12, No.2, March. 2001

[Orav12] Oravec M.: Metódy strojového učenia na extrakciu príznakov a rozpoznávanie vzorov I: Neurónové siete na extrakciu príznakov, kompresiu a rozpoznávanie obrazu. -1. vyd. -Bratislava: STU v Bratislave FEI, 2012. -150 s. -ISBN 978-80-227-3691-6

[Orav13] Oravec,M., Pavlovičová,J., Mazanec,J., Omelina,Ľ., Féder,M., Ban,J., Mazanec,J., Valčo,M., Zelina,M.: Metódy strojového učenia na extrakciu príznakov a rozpoznávanie vzorov II: Rozpoznávanie tvárí v biometrii, monografia, vydavateľstvo Felia, Bratislava, 2013, ISBN 978-80-971512-0-1

[Rah07] Rahman M. Md., Islam R. Md., Bhuiyan N. I., Ahmed B., Islam A. Md.: Person Identification Using Ear Biometrics, International Journal of The Computer, the Internet and Management Vol. 15: 2, May - August 2007, pp. 1 – 8

[Rui14] Rui Min, Neslihan Kose, Jean-Luc Dugelay, "KinectFaceDB: A Kinect Database for Face Recognition," Systems, Man, and Cybernetics: Systems, IEEE Transactions on , vol.44, no.11, pp.1534,1548, Nov. 2014, doi: 10.1109/TSMC.2014.2331215

[Sand09] Sanderson C., Lovell B.C.: Multi-Region Probabilistic Histograms for Robust and Scalable Identity Inference. ICB 2009, LNCS 5558, pp. 199-208, 2009

[Sing13] Singh R., Om h.: An Overview of Face Recognition in an Unconstrained Environment, Proceedings of the 2013 IEEE, Second International Conference on Image Information Proceedings, 2013, pp. 672-677

[Spath85] Spath H.: "The cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples," New York: Prentice-Hall, 1985

[Turk91] Turk M., Pentland A.: Eigenfaces for Recognition, Journal of Cognitive Neuroscience, vol. 3, 1991, pp. 71-86

[Yimi20] Yiming L., Jie S., Shiyang Ch., Maja P.: MobiFace: FT-RCNN: Real-time Visual Face Tracking with Region-based Convolutional Neural Networks, The IEEE International Conference on Automatic Face and Gesture Recognition (FG), 2020