VYSOKÁ ŠKOLA BÁŇSKÁ–TECHNICKÁ UNIVERZITA OSTRAVA
VŠB–TECHNICAL UNIVERSITY OF OSTRAVA

FAKULTA ELEKTROTECHNIKY A INFORMATIKY
FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

KATEDRA TELEKOMUNIKAČNÍ TECHNIKY
DEPARTMENT OF TELECOMMUNICATIONS

Využití umělé inteligence pro vícenásobnou bezkontaktní biometrickou autentizaci

Multimodal biometric contactless authentication using the artificial intelligence

DIZERTAČNÍ PRÁCE
DISSERTATION THESIS

AUTOR PRÁCE — Ing. Jaromír Továrek
AUTHOR

VEDOUCÍ PRÁCE — doc. Ing. Miroslav Vozňák, Ph.D.
SUPERVISOR

OSTRAVA, 2017

PODĚKOVÁNÍ

Rád bych tímto poděkoval vedoucímu dizertační práce, panu docentovi Ing. Miroslavu Vozňákovi, Ph.D., za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Ostrava ....

(podpis autora)

PROHLÁŠENÍ

Prohlašuji, že jsem svou dizertační práci vypracoval samostatně pod vedením vedoucího dizertační práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.
Jako autor uvedené dizertační práce dále prohlašuji, že v souvislosti s vytvořením této dizertační práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Ostrava ....
(podpis autora)

ABSTRAKT

Biometrické autentizační systémy slouží k ověření identity osoby pomocí jedinečných tělesných znaků (otisk prstu, geometrie obličeje, duhovka oka, sítnice oka, geometrie ruky, hlas atd.). Výhodou tohoto typu autentizace je, že si osoba nemusí pamatovat několikamístné heslo nebo s sebou neustále nosit snadno zcizitelný token (přihlašovací kartu). Biometrická autentizace je rychlou, pohodlnou a velice přesnou metodou. Mezi hlavní výhody biometrické autentizace patří vysoký stupeň spolehlivosti, nulové provozní náklady, rychlost, praktičnost a zřejmost. Oblast využití biometrických systémů můžeme rozdělit do dvou sfér a to do bezpečnostně-komerční (ochrana počítačů a dat, zajištění komfortu, vstup do objektů) a forenzní (soudní, kriminalistická a vyšetřovací). Podstatou všech biometrických systémů je automatizované snímání biometrických charakteristik a jejich následné porovnání s dříve získanými údaji. Jedním z cílů v oblasti bezpečnosti je vytvoření komplexních systémů založených na kombinaci měření více charakteristik. Z tohoto důvodu se tato práce zaměřuje na návrh biometrického autentizačního systému založeného na dvou charakteristikách a to na hlase a geometrii obličeje, kde hlavní roli v oblasti klasifikace bude hrát strojové učení. Vytvořením takového systému pracujícího na bázi vícenásobné autentizace s robustně navrženými klasifikátory vzniká unikátní nástroj pro bezdotykovou autentizaci, který může být v budoucnu využit jako přístupový systém v budovách či pro ověření přístupu osob k různým zařízením.

KLÍČOVÁ SLOVA

Autentizace, biometrické charakteristiky, biometrické systémy, fúze, neuronové sítě, rozpoznávání obličeje, rozpoznávání řečníka, strojové učení, umělá inteligence, vícenásobná biometrie.

ABSTRACT
Biometric authentication systems are used to verify the identity of the person using unique physical features (fingerprint, facial geometry, iris, retina, hand geometry, voice, etc.). The advantage of this type of authentication is that a person does not need to remember a password or always carry an easily stealable token (registration card). Biometric authentication is a fast, convenient and very precise method. Among the main benefits of biometric authentication include a high reliability, zero operating costs, speed, practicality, and clarity. The field of application of biometric systems can be divided into two spheres - security-commercial (security of computers and data, ensuring a comfort, entry into buildings) and forensic (judicial, forensic and investigative). The basic of all biometric systems is automated scanning of biometric characteristics and their subsequent comparison with previously collected data. One of the goal in the field of security is the realization of complex systems based on a combination of multiple characteristics measurements. For this reason, this work focuses on the design of a biometric authentication system based on two characteristics - voice and facial geometry, where the main role in the classification is played by machine learning. By creating such a multimodal autentication system with robustly designed classifiers, a unique contactless authentication tool is created, which can be used as an building access system or personal access system in the future.

KEYWORDS
Artificial intelligence, authentication, biometric characteristic, biometric systems, face recognition, fusion, machine learning, multimodal biometrics, neural networks, speaker recognition.

OBSAH

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

SEZNAM SYMBOLŮ A ZKRATEK

Seznam zkratek

2D	Dvourozměrný
2DLDA	Two-Dimensional Linear Discriminative Analysis
ANN	Artifiacial Neural Network
DET	Detection Error Tradeoff
DNA	Deoxyribonucleic Acid
EER	Equal Error Rate
EGM	Elastic Graph Matching
FAR	False Acceptance Rate
FRR	False Rejection Rate
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
HOG	Histogram of Oriented Gradients
K-NN	K-Nearest Neighbors
LBP	Local Binary Patterns
LPC	Linear Prediction Coefficient
MFCC	Mel Frequency Cepstral Coefficient
MLNN	Multilayer Neural Network
ROC	Receiver Operating Charasteristic
SVM	Support Vector Machines
UBM	Universal Background Model
VoIP	Voice over Internet Protocol
WAV	Waveform audio file format

Použité symboly

α_v	Váha slabého klasifikátoru
β_i	Váha podpůrného vektoru
Δw_i	Změna synaptických vah

$\Delta w_i^{'}$	Změna synaptické vahy z předchozího kroku
η	Koeficient učení
$\frac{\partial f}{\partial x}$	Gradient obrazové funkce ve směru x
$\frac{\partial f}{\partial y}$	Gradient obrazové funkce ve směru y
λ	Strmost sigmoidu
μ	Koeficient změny vah
∇f	Gradient obrazové funkce
ω_j	Klasifikační třída
ψ	Úhel gradientu
$\sum_i$	Kovarianční matice gaussovských rozdělení
Θ	Prahová hodnota slabého klasifikátoru
Θ'	Prahová hodnota silného klasifikátoru
$\vec{\mu_i}$	Vektor středních hodnot gaussovských rozdělení
$\vec{b^s}$	Vektor výstupních pravděpodobností markovova modelu
$\vec{I_i}$	Vektor excitací vstupní vrstvy
$\vec{O_i}$	Vektor excitací výstupní vrstvy
$\vec{w}_{SVM}$	Vektor vah SVM klasifikátoru
$\vec{x_n}$	Vektor extrahovaných příznaků
$\vec{X}_i$	Vektor trénovacích dat
$\vec{X}_{test}$	Vektor testovacích dat
$a(k)$	LPC koeficienty lineárního filtru
A^s	Matice pravděpodobností přechodů markovova modelu
a_{ij}	Pravděpodobnosti přechodu markovova modelu
b	Bias
c	Výsledná třída po fúzi klasifikátorů
$c(k)$	Kepstrální koeficinety LPC
c_m	Melovské kepstrální koeficienty MFCC
c_{lift}	Lifterované kepstrální koeficienty LPC

E	Celková chyba neuronové sítě
f	Obrazová funkce
$f[Hz]$	Frekvence v lineární škále
$f_m[Mel]$	Frekvence v nelineární melovské škále
f_w	Hodnota Haarova příznaku
H_w	Odezva silného klasifikátoru
h_w	Odezva slabého klasifikátoru
I	Počet mixů gaussových hustotních funkcí
I_c	Hodnota jasu středového pixelu
I_u	Hodnota jasu sousedního pixelu
iim	Hodnoty integrálního obrazu
im	Hodnoty pixelů původního obrazu
K	Počet trénovacích vektorů
K_P	Počet podpůrných vektorů
L	Váha lifteringu
LBP	Hodnoty lokálního binárního vzoru
M	Počet melovských kepstrálních koeficientů
m	Počet neuronů výstupní vrstvy
M^*	Počet pásem melovského filtru
M_{client}	Model/třída oprávněného uživatele
M_{world}	Model/třída neoprávněných uživatelů
N	Gaussovské hustoty
n	Počet vazeb vstupujících do neuronu
N_r	Celkový počet rámců řečového signálu
N_{EVA}	Počet pokusů oprávněných osob o identifikaci/verifikaci
N_{FA}	Počet chybných přijetí
N_{FR}	Počet chybných odmítnutí
N_{IVA}	Počet pokusů neoprávněných osob o identifikaci/verifikaci

o_j	Požadovaná odezva výstupní vrstvy
p	Počet vzorů trénovací množiny
p_w	Polarita Haarova příznaku
Q	Řád lineárního filtru
R	Celkový počet klasifikátorů
R_c	Výsledné pravděpodobnostní skóre
$S(z)$	Aktivační funkce neuronu
S_c	Pravděpodobnostní skóre pro model oprávněného uživatele
S_w	Pravděpodobnostní skóre pro model neoprávněného uživatele
$S_{j,k}$	Normalizované skóre klasifikátoru pro danou třídu
S_j	Výsledné skóre po fúzi klasifikátorů
T	Trénovací množina neuronové sítě
t_n	Třídy trénovacích dat SVM klasifikátoru
U	Počet sousedních pixelů
V_w	Počet slabých klasifikátorů daného stupně
w_i	Synaptické váhy neuronu
w_i^*	Váha mixu gaussových hustotních funkcí
w_k	Váha klasifikátoru
x_c	X souřadnice středového pixelu
x_w	Detekční okno
x_{image}	X souřadnice pixelu integrálního obrazu
x'_{image}	X souřadnice pixelu původního obrazu
x_i	Vstupní hodnoty do neuronu
y_c	Y souřadnice středového pixelu
y_j	Skutečná odezva neuronu výstupní vrstvy
y_m	Odezvy melovských filtrů
y_{ex}	Hodnota excitace neuronu
y_{image}	Y souřadnice pixelu integrálního obrazu
y'_{image}	Y souřadnice pixelu původního obrazu
z	Vnitřní potenciál neuronu

1 ÚVOD

Problematiku ověření identity osob můžeme rozdělit do tří skupin podle použité metody autentizace. První skupinu tvoří metody založené na prokázání identity pomocí vlastnictví (identifikační doklady, karty, čipy), druhá skupina využívá k prokázání identity znalostí (hesla, identifikační čísla), třetí skupinu představují metody využívající k ověření identity měřitelné biometrické charakteristiky (otisk prstu, geometrie obličeje, duhovka oka, sítnice oka, geometrie ruky, hlas atd.) Z principu jednotlivých metod můžeme odvodit největší nevýhody. U prokazování identity vlastnictvím je největší bezpečnostní riziko dáno možností odcizení nebo napodobení tokenu. Ověření identity pomocí znalostí nese riziko v odpozorování, uhodnutí, případně odvození hesla nebo identifikačního čísla. V praxi se ke snížení těchto bezpečnostních rizik používá kombinace obou přístupů. Přesto se jako nejbezpečnější způsob jeví ověření identity pomocí specifické biometrické charakteristiky člověka.

Přes vysokou bezpečnost ověřování identity pomocí biometrických charakteristik se i u této identifikace začínají objevovat první pokusy pachatelů o změnu nebo napodobení biometrické charakteristiky. Příkladem mohou být pokusy o změnu otisků prstů, plastické operace v oblasti tváře, napodobování hlasu apod. Pokud by se pachateli podařilo vytvořit napodobeninu některé z biometrických charakteristik, tak by to mohlo znamenat výraznou hrozbu pro všechny systémy založené na biometrické identifikaci. Z tohoto důvodu je snaha stálého zvyšování bezpečnosti biometrických systémů, tak aby k podobným pokusům pokud možno vůbec nedocházelo nebo se jim alespoň dalo v dostatečné míře čelit.

Jedním ze způsobů jak dosáhnout vyšší bezpečnosti biometrických aplikací je použití vícenásobné biometrie (Multiple Biometrics). Jedná se o kombinaci více biometrických charakteristik pro verifikaci v jednom systému. V současné době jsou nejpoužívanější kombinace otisk prstů - geometrie obličeje a geometrie oční duhovky - hlas. Lze očekávat, že v brzké době přibudou i další kombinace biometrických charakteristik.

Z tohoto důvodu je tato práce zaměřena na návrh vícenásobného biometrického autentizačního systému, který je založen na autentizace hlasem a autentizaci geometrií obličeje. Tyto biometrické metody byly vybrány s ohledem na jejich vlastnosti. Mezi tyto vlastnosti patří: přijatelnost pro uživatele (snímání biometrických dat je přirozené - promluva, fotka), dostatečná přesnost a jednoduchost snímání (stačí pouze fotoaparát a mikrofon). Jednoduchost snímání umožňuje využití systému v podstatě na jakémkoliv zařízení a v různých oblastech. Další výhodou této kombinace metod je to, že pohyb úst při verifikaci hlasem může být využit jako "Test živosti"(data pochází od skutečného uživatele) autentizující se osoby. Takto vzniklý systém může být nasazen jako vysoce bezpečný autentizační systém pro vstup do objektu nebo pro přihlášení k různým zařízením.

Z pohledu struktury dizertační práce se úvodní kapitola 2 zabývá současným stavem řešené problematiky. V této kapitole je podrobně popsán princip a metody využívající se jak pro autentizaci hlasem tak pro autentizaci geometrií obličeje. V poslední části této

kapitoly jsou uvedeny současné trendy v oblasti vícenásobné biometrie. V kapitole 3 jsou uvedeny cíle dizertační práce, tak jak byly schváleny státní komisí.

Jednotlivé kroky návrhu biometrického autentizačního systému založeného na verifikaci hlasem jsou podrobně popsány v kapitole 4. Kapitola 5 popisuje návrh systému založeného na autentizaci pomocí geometrie obličeje. Na základě navržených systémů v kapitolách 4 a 5 byl vytvořen komplexní vícenásobný biometrický autentizační systém, který je popsán v kapitole 6.

Kapitola 7 poskytuje srovnání navrženého vícenásobného biometrického systému s aktuálně používanými systémy biometrické autentizace. Zároveň je uvedeno srovnání přesnosti navrženého vícenásobného systému v kapitole 6 se systémy navrženými v kapitolách 4 a 5. V poslední kapitole jsou poté shrnuty dosažené výsledky a možnost nasazení systému v reálné aplikaci.

2 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

Výzkum v rámci této dizertační práce se zabýval návrhem vícenásobného biometrického autentizačního systému s využitím umělé inteligence, který slouží jako unikátní nástroj pro vysoce bezpečnou autentizaci. Návrh systému probíhal na základě kombinace/fúze dvou biometrických charakteristik (hlas a geometrie obličeje).

Návrh biometrického systému založeného na verifikaci hlasem spočíval v nalezení vhodných parametrů, které v kombinaci s přesným klasifikátorem poskytují nejnižší chybovost, vysokou přesnost a zároveň je tento systém dostatečně rychlý. Obecným požadavkem na systém byla jeho textová závislost, která z pohledu přesnosti a bezpečnosti přináší výrazné zvýšení.

Princip návrhu biometrického systému založeného na rozpoznávání pomocí geometrie obličeje byl obdobný jako v případě hlasové autentizace. Opět bylo úkolem nalézt ideální parametry a klasifikátor za účelem dosažení vysoké přesnosti a bezpečnosti autentizace. Pro tento biometrický systém byly specifikovány následující požadavky: forma zpracovávaného obrazu 2D , typ spektra obrazu černobílé i barevné, způsob snímání obrazu - čelní pohled.

Výsledný vícenásobný biometrický autentizační systém vznikl vhodnou úpravou a kombinací/fúzí předchozích dvou navržených systémů.

Následující podkapitoly shrnují teorii k výše uvedené problematice.

2.1 Biometrická identifikace/verifikace

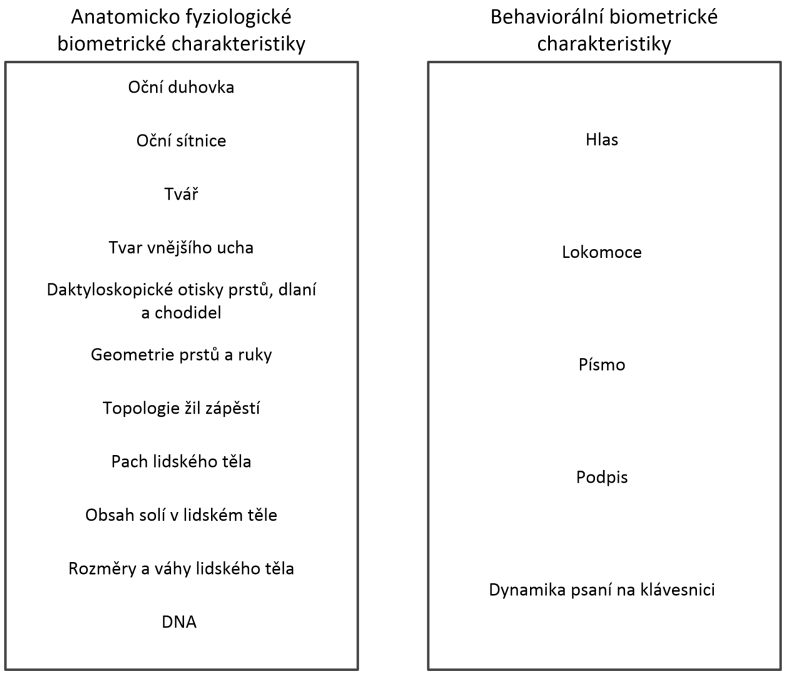
Základní myšlenka biometrické identifikace/verifikace je založena na tom, že naše fyzické (a psychické) charakteristiky jsou jedinečné a mohou být tedy použity pro efektivní identifikaci/verifikaci osoby. Jednotlivé charakteristiky tvořící identitu osoby je pak téměř nemožné napodobit nebo pozměnit. Biometrickou identitu nelze ani odcizit, protože identifikační charakteristiky jsou bezprostředně spojené s identifikovanou osobou. Mezi hlavní výhody biometrické identifikace/verifikace patří: nelze ji zapomenout nebo ztratit, je těžké ji odcizit nebo napodobit, je nepřenositelná, vysoká přesnost a rychlost identifikace, je lidsky přirozená [1].

2.1.1 Biometrické charakteristiky

Biometrická identifikace/verifikace je využití jedinečných, měřitelných, fyzikálních nebo fyziologických znaků (markantů) nebo projevů člověka k jednoznačnému zjištění (identifikace) nebo ověření (verifikace) jeho identity. Pro její účely se používají dvě skupiny biometrických charakteristik:

- fyziologické biometrické charakteristiky,
- behaviorální biometrické charakteristiky.

V praxi obecně platí, že behaviorální charakteristiky jsou používány méně často něž objektivnější fyziologické charakteristiky. Nároky kladené na jednotlivé charakteristiky jsou: jedinečnost, stálost, praktická měřitelnost a technologická možnost dalšího zpracování. Rozdělení charakteristik zobrazuje obrázek 2.1, [1], [4].



Obr. 2.1: Rozdělení biometrických charakteristik.

Fyziologické biometrické charakteristiky

Fyziologické charakteristiky jsou pro každého člověka unikátní a časově neměnné. Tyto charakteristiky jsou využívány na základě vědeckých poznatků o oční duhovce, oční sítnici, tváři, otiscích prstů, dlaní a chodidel, lidském tělesném pachu, skladbě DNA, atd.

Behaviorální biometrické charakteristiky

Behaviorální charakteristiky jsou dány specifickými rysy lidského chování. Využívají poznatků o lidském hlase, pohybu těla, o znalostech a dovednostech psaní. Behaviorální biometrické charakteristiky jsou unikátní a mohou být časově nestálé (změna frekvence základního hlasivkového tónu způsobená stárnutím).

2.1.2 Oblasti použití biometrické identifikace/verifikace

Biometrická identifikace nalézá uplatnění v mnoha oblastech běžného života. Podle přesnosti, spolehlivosti a objektivnosti metody identifikace můžeme tyto metody rozdělit do následujících oblastí [1]:

- policejně-soudní (forenzní),
- bezpečnostně-komerční,
- ezoterickou.

Policejně-soudní identifikace

Policejně-soudní (forenzní) identifikace je nejnáročnější oblastí biometrické identifikace. Případná chyba při identifikaci může mít fatální vliv na lidský život (chybné usvědčení nevinné osoby, chybné osvobození skutečného pachatele). Z tohoto důvodu je kladen velký důraz na vyloučení jakékoliv chyby. Mezi nejčastěji používané biometrické metody identifikace patří identifikace na základě otisků prstů, dlaní a chodidel, analýza DNA, popřípadě fonetická analýza lidského hlasu. Ve forenzních aplikacích převažuje identifikace nad verifikaci. Výsledek každého identifikačního procesu na konci zhodnocuje člověk (specialista - soudní znalec).

Bezpečnostně-komerční identifikace

Hlavním rozdílem mezi forenzní a bezpečnostně-komerční oblastí je v tom, že bezpečnostně-komerční biometrické aplikace jsou plně automatické (rozhodnutí aplikace o přijetí/zamítnutí je plně postačující a zásah člověka není potřeba). V této oblasti převládá verifikace nad identifikací. Většinou jsou aplikace využívány pro povolení/zamítnutí přístupu do chráněných "objektů"(fyzické objekty, mobilní telefony, bankomaty, servery, startování automobilů atd.). Mezi používané metody identifikace patří identifikace na základě geometrie obličeje, projevech hlasu, poznatků o oční duhovce, sítnici atd..

Ezoterická identifikace

Do oblasti ezoterické identifikace řadíme metody, které nejsou zdaleka tak rozšířeny a používány jako metody v předchozích oblastech. Do této oblastí patří metody jako lokomoce (rysy chůze), tvar vnějšího ucha, topografie žil, pach lidského těla, obsah solí v těle apod..

2.1.3 Kritéria pro biometrické technologie

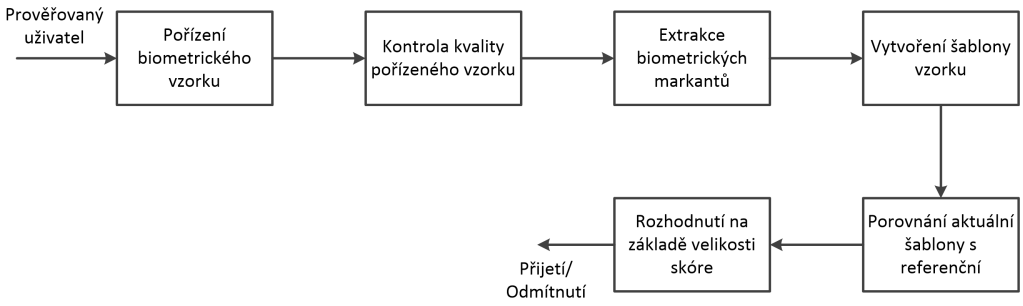
Pomocí kritérií lze popsat libovolný biometrický systém. Kritéria biometrických technologií rozhodují mimo jiné o samotné funkčnosti systému, efektivitě systému, praktičnosti, finanční přijatelnosti, bezpečnosti a celkové úspěšnosti. Kritéria mohou být rozdělena do následujících skupin [1]:

- operační kritéria,
- metodologická, algoritmická a bezpečnostní kritéria,
- technická kritéria,
- finanční kritéria,
- výrobní kritéria.

2.1.4 Obecné principy biometrických technologií

Při pohledu na biometrické systémy obecně, lze říci, že většina z nich pracuje podle stejného technologického postupu 2.2. Pro konkrétní biometrický systém se potom liší pouze použité metody v jednotlivých krocích. Obecný technologický postup se skládá z následujících kroků [1]:

- pořízení biometrického vzorku,
- kontrola kvality pořízeného vzorku,
- extrakce biometrických markantů,
- vytvoření šablony vzorku,
- porovnání aktuální šablony s referenční a vygenerování skóre,
- rozhodnutí (přijetí/odmítnutí) na základě velikosti skóre.



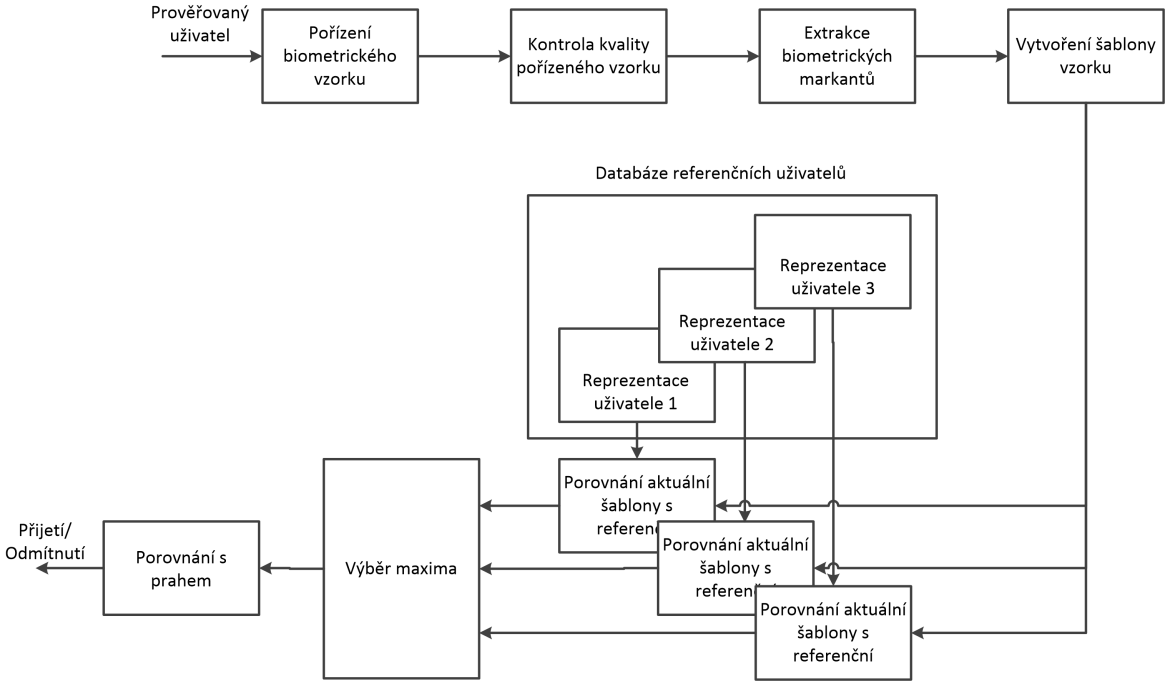
Obr. 2.2: Obecný technologický postup biometrické autentizace.

Identifikace

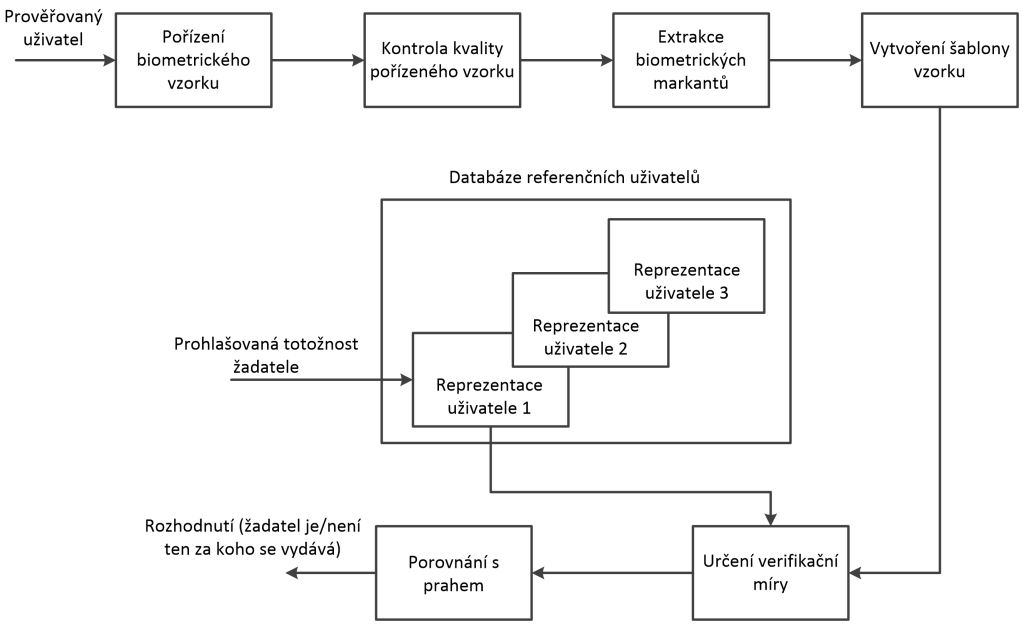
U identifikace je úkolem určit, komu ze známé množiny uživatelů patří (je nejpodobnější) daný biometrický vzorek. Proces porovnání probíhá v tomto případě mezi šablonou nasnímaného biometrického vzorku a mezi všemi referenčními šablonami v databázi. Proces identifikace je zobrazen na obrázku 2.3.

Verifikace (autentizace)

Verifikaci si lze představit jako ověření identity člověka. K dispozici máme biometrický vzorek a víme komu by měl patřit. Úkolem tedy je, ověřit zda opravdu biometrický vzorek patří osobě, za kterou se vydává. Proces porovnání se v tomto případě omezuje pouze na porovnání šablony nasnímaného biometrického vzorku s referenční šablonou prověřované osoby. Z toho vyplývá, že je verifikace podmnožinou identifikace. Proces verifikace je zobrazen na obrázku 2.4.



Obr. 2.3: Blokové schéma procesu identifikace.



Obr. 2.4: Blokové schéma procesu verifikace.

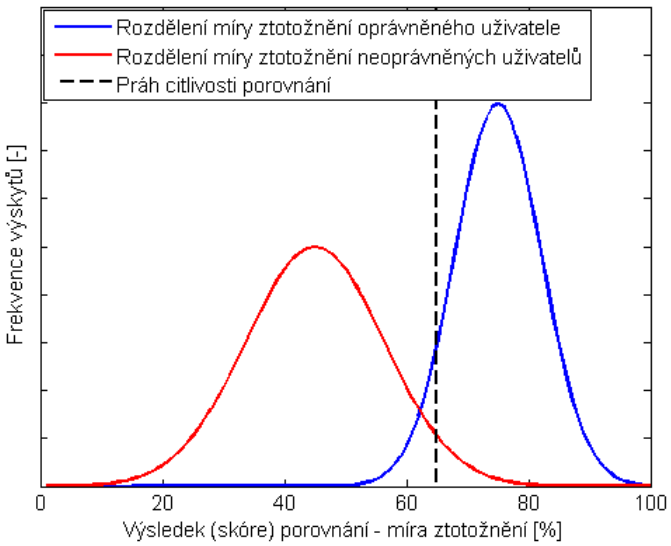
2.1.5 Měření výkonnosti biometrických metod

Pro porovnání efektivity různých biometrických metod je možné využít celou řadu statistických koeficientů. Charakteristickými výkonnostními parametry jsou pravděpodob-

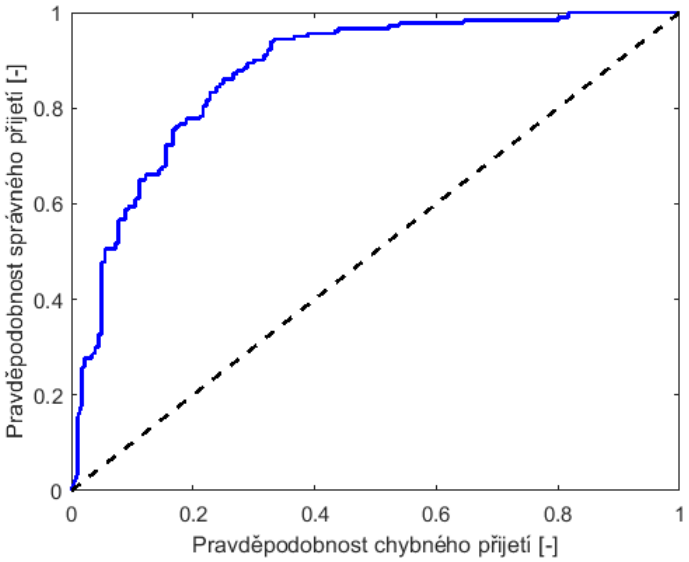
nost chybného přijetí a pravděpodobnost chybného odmítnutí. Často se také využívají parametry pravděpodobnost správného přijetí a pravděpodobnost správného odmítnutí. Biometrické metody jsou založeny na statistickém vyhodnocování podobnosti šablony biometrického vzorku a referenční šablony. Výsledkem vyhodnocení je poté míra ztotožnění takzvané skóre. V praxi se používá grafické vyjádření závislosti četnosti skóre osob, které se podrobují identifikaci/verifikaci. Tyto závislosti jsou zobrazeny v histogramu 2.5 a vychází z takzvané kontingenční tabulky 2.1. Na základě velikosti skóre a nastaveného prahu citlivosti je rozhodnuto o výsledku identifikace/verifikace (odmítnutí/přijetí). Ohodnocení činnosti systému může být také provedeno pomocí křivky ROC, která je zobrazena na obrázku 2.6, [1], [3].

Tab. 2.1: Kontingenční tabulka.

Skutečný výstup	*Oprávněný uživatel*	Skutečně pozitivní (TP)	Falešně pozitivní Chyba I. druhu (FP)	Prediktivní hodnota pozitivního testu
	Neoprávněný uživatel	Falešně negativní Chyba II. druhu (FN)	Skutečně negativní (TN)	Prediktivní hodnota negativního testu
		Senzitivita	Specificita	**Přesnost**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		



Obr. 2.5: Histogram rozdělení skóre oprávněných a neoprávněných uživatelů.



Obr. 2.6: ROC křivka.

Pravděpodobnost chybného přijetí (False Acceptance Rate - FAR)

Pravděpodobnost chybného přijetí udává pravděpodobnost, že bude neoprávněná osoba přijata jako oprávněná. Chybné přijetí může často vést ke vzniku škody (lupič je vpuštěn do objetu atd.). Koeficient FAR tedy udává míru bezpečnosti a je definován vztahem 2.1.

$$FAR = \frac{N_{FA}}{N_{IVA}}, \tag{2.1}$$

kde N_{FA} je počet chybných přijetí a N_{IVA} počet pokusů neoprávněných osob o identifikaci/verifikaci.

Pravděpodobnost chybného odmítnutí (False Rejection Rate - FRR)

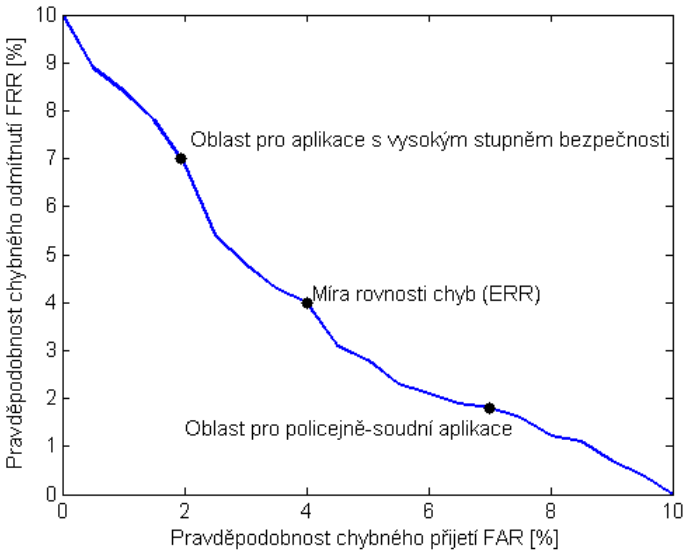
Pravděpodobnost chybného odmítnutí udává pravděpodobnost, že bude oprávněný uživatel odmítnut jako neoprávněný. Chybné odmítnutí znamená především omezení komfortu uživatele. Oprávněný uživatel musí opětovně prokazovat svoji identitu. Z bezpečnostního hlediska nemá tato chyba velký význam. Naopak má velký význam z marketingového hlediska, protože vznik této chyby má za následek nespokojeného uživatele. Koeficient FRR je definován vztahem 2.2.

$$FRR = \frac{N_{FR}}{N_{EVA}}, \tag{2.2}$$

kde N_{FR} je počet chybných odmítnutí a N_{EVA} počet pokusů oprávněných osob o identifikaci/verifikaci.

Vztah FRR a FAR

V ideálním případě je po biometrickém systému požadováno, aby nedělal žádné chyby. To ovšem v praxi není možné. Podle oblasti použití aplikace je nutné volit, která chyba může být větší. Například při použití biometrické verifikace v bance je výhodnější, aby systém dělal více FRR chyb, ale nedělal žádné FAR chyby (tzn. systém nepozná oprávněného uživatele i když je to opravdu on a vyzve ho k zopakování autentizace - v tomto případě je lepší pokusit se o autentizaci vícekrát, než aby byl k účtu připuštěn neoprávněný člověk). Naopak při aplikaci systému u bezpečnostních složek je výhodnější, aby systém produkoval více FAR chyb než FRR chyb (lepší prověřit více nesprávně obviněných lidí, než nechat utéct pachatele). Nastavení toho, která chyba bude "tolerována", se provádí pomocí rozhodovacího prahu (treshold). Závislost chyb FRR a FAR představuje křivka DET 2.7. Bod v němž se obě chyby rovnají se nazývá EER .



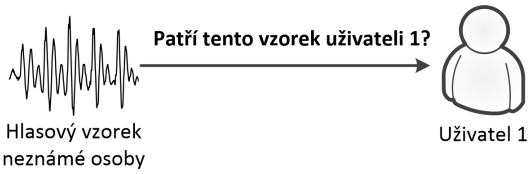
Obr. 2.7: DET křivka.

Zvyšování bezpečnosti biometrických metod

Snaha zvyšování bezpečnosti biometrických metod vyplývá z toho, že všechny reálné biometrické metody pracují s určitou chybovostí a to ve všech aplikacích nevyhovuje. V neposlední řadě se začínají objevovat případy pokusů o napadení biometrického systému (pokus o změnu otisků prstů, plastické operace tváře atd.). Jedním ze způsobů zvýšení bezpečnosti je využití ezoterické identifikace, zvýšení bezpečnosti je v tomto případě zaručeno obtížným nebo nemožným napodobení ezoterických metod. Druhým způsobem zvýšení bezpečnosti je použití vícenásobné biometrie [1], [20], [23], [26], [27].

2.2 Verifikace řečníka

Autentizaci řečníka si lze představit, jako ověření totožnosti člověka na základě jeho hlasu. K dispozici je záznam hlasu neznámé osoby a je známo, za koho se daná osoba vydává. Úkolem je tedy ověřit, zda hlas neznámé osoby je dostatečně podobný hlasu člověka, za kterého se vydává. Z toho také vyplývají oblasti použití této biometrické metody (řízení přístupu do objektů, autentizace transakcí a tak dále). Mezi hlavní výhody této biometrické metody patří: nízká cena, vysoká spolehlivost, akceptace metody uživateli, široké spektrum použití. Naopak mezi nevýhody patří: střední přesnost (nepatří mezi nejpřesnější biometrické metody), pro zkvalitnění referenční šablony je potřeba velký počet biometrických vzorků [1], [8], [30], [31].



Obr. 2.8: Verifikace (autentizace) řečníka.

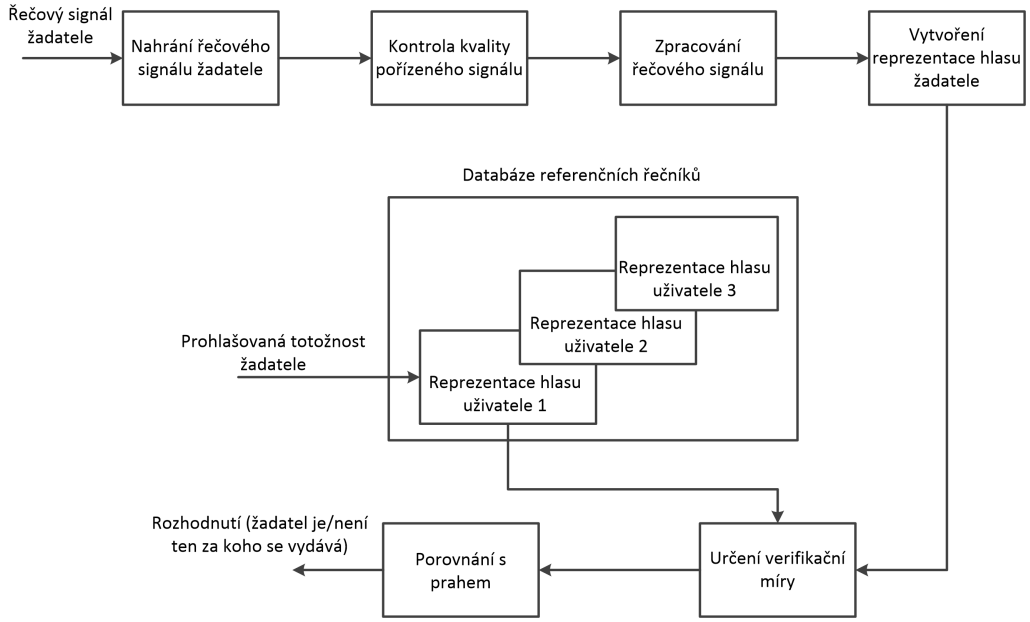
2.2.1 Princip činnosti verifikačního systému řečníka

Před procesem samotné autentizace musí žadatel potvrdit svou prohlašovanou totožnost (identitu). Následně je žadatel systémem vyzván k pronesení určitého textu (libovolný, předem určený). Řečový signál je zaznamenám a podle jeho kvality následně zpracováván příslušnými metodami zpracování řečového signálu. Z takto předzpracovaného signálu jsou extrahovány signifikantní příznaky (markanty), které jsou následně použity k vhodné reprezentaci hlasu žadatele. Reprezentace hlasu žadatele je porovnána s reprezentací hlasu prohlašované totožnosti a určí se míra podobnosti (skóre). Porovnáním skóre s předem nastavenou hodnotou verifikačního prahu je rozhodnuto o tom, zda je žadatel opravdu ten člověk, za kterého se vydává. Pokud je skóre větší než verifikační práh, tak se opravdu jedná o člověka, za kterého se žadatel vydává, v opačném případě nikoliv. Blokové schéma principu činnosti je zobrazeno na obrázku 2.9, [8], [30], [32], [33].

2.2.2 Rozdělení systémů pro rozpoznávání řečníka

Systémy pro rozpoznávání řečníka mohou být rozděleny podle typu textu, který má uživatel vyslovit, aby byl autentizován do tří skupin [1], [8], [31]:

- textově závislé - systém v tomto případě "zná"sekvenci, kterou má uživatel vyslovit. Tato sekvence je použita při trénování systému. Jedná se tedy o dvojitou ochranu, první ochrana je dána promluvou, kterou má uživatel vyslovit. Tato promluva je poté považována za heslo. Druhá ochrana je dána verifikací řečníka. Textově závislé



Obr. 2.9: Princip činnosti systému pro verifikaci řečníka.

systémy mají největší výhodu v tom, že jsou přesnější, protože jsou natrénovány na danou promluvu (pokrývají menší prostor parametrů). Hlavní nevýhoda těchto systémů spočívá v hrozbě útoku přehráním (pachatel si nahraje pokus referenčního uživatele o autentizaci). Nejvíce se tyto systémy využívají v oblasti řízení přístupu [14], [30],

- systémy s textovou výzvou - u těchto systémů není sekvence, která má být vyslovena předem známa uživateli. Místo toho je systémem náhodně vybrána jedna z předem připravených sekvencí a uživatel ji musí pronést. Tyto systémy pracují ve dvou krocích. Nejprve je ověřeno zda uživatel pronesl správnou sekvenci a poté teprve dochází k verifikaci. Tento princip odstraňuje hrozbu útoku přehráním a může být považován za test živosti. Další výhodou je, že si uživatel nemusí pamatovat žádné heslo. Nevýhodou může být dialog s uživatelem (delší doba verifikace). Oproti textově závislým systémům se snižuje i výkonnost verifikace. Oblast využití těchto systémů je především v zabezpečení přístupu k telefonním službám [1], [31],
- textově nezávislé - systém nezná danou promluvu, musí být schopný rozpoznat řečníka nezávisle na tom, co říká (jiný jazyk, nemoc, emoce a tak dále). Z tohoto důvodu jsou tyto systémy méně přesné než systémy textově závislé. Hlavní využití těchto systémů je u bezpečnostních složek, kde se například snaží určit, kdo hovoří na dané nahrávce [31], [33].

2.2.3 Řečové příznaky používané pro rozpoznávání řečníka

Před samotnou extrakcí řečových parametrů je nejprve nutné provést takzvané předzpracování řečového signálu. Předzpracování se skládá ze čtyř kroků [8], [33], [Tov01], [Tov02]:

- odstranění stejnosměrné složky - stejnosměrná složka nenese žádnou užitečnou informaci a naopak může být pro další zpracování rušivá,
- preemfáze - slouží k vyrovnání kmitočtové charakteristiky řeči (energie klesá směrem k vyšším frekvencím),
- segmentace - z důvodu nestacionarity řečového signálu je nutné provést rozdělení řečového signálu na menší úseky (segmenty), kde lze předpokládat stacionaritu signálu. Délka rámce se zpravidla volí v rozmezí 20-30 ms,
- váhování oknem - segmentací signálu vznikají na okrajích rámců ostré přechody, které jsou nežádoucí při frekvenční analýze. Vynásobením rámců okenní funkcí jsou tyto ostré přechody potlačeny.

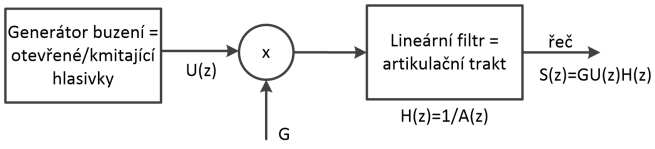
Z takto předzpracovaného signálu lze následně provést extrakci signifikantních příznaků. Na řečové příznaky používané pro rozpoznávání řečníka jsou kladeny následující požadavky [8], [32]:

- robustnost - příznak by se neměl měnit v čase, při nemoci nebo ve stresové situaci,
- bezpečnost - příznak by měl jednoznačně definovat daného řečníka, nemělo by být jednoduché někoho napodobit,
- praktičnost - příznak by mělo být jednoduché zjistit a měl by se v řeči běžně vyskytovat.

Mezi nejpoužívanější příznaky, které splňují předchozí požadavky jsou koeficienty LPC případně z nich odvozené kepstrální koeficienty a především melovské kepstrální koeficienty MFCC [6], [8], [30], [33].

2.2.4 Kepstrální analýza

Kepstrální analýza vychází z matematického modelu generování řeči (řeč je dána jako konvoluce budící funkce a impulzní odezvy hlasového ústrojí v časové oblasti a jako násobení obrazů ve frekvenční oblasti). Zjednodušený model generování řeči je zobrazen na obrázku 2.10. Cílem kepstrální analýzy je vydělit jednotlivé členy konvolutorního součinu. Procesem kepstrální analýzy vzniká takzvané kepstrum. Část výsledného kepstra poté nese informaci o budící funkci a druhá část o impulzní odezvě hlasového ústrojí [8].



Obr. 2.10: Model generování řeči.

Kepstrální koeficienty LPC

Základní výhodou kepstrálních koeficientů LPC je to, že jsou obecně méně korelované než například samotné koeficienty LPC. Kepstrální koeficienty LPC jsou vztaženy ke spektrální obálce odvozené analýzou LPC [8], [30], [Tov03]. Při výpočtu kepstrálních koeficientů LPC je postupováno v souladu s pravidly homomorfního zpracování řeči. Výpočet kepstrálních koeficientů LPC je proveden podle následujících vztahů 2.3, 2.4, 2.5:

$$c(1) = -a_1, \tag{2.3}$$

$$c(k) = -a_k - \sum_{i=1}^{k-1} \left(\frac{i}{k}\right) c(i) a_{k-1}, \; pro \; 2 \leq k \leq Q, \tag{2.4}$$

$$c(k) = \sum_{i=1}^{Q} \left(\frac{k-i}{k}\right) c(k-i) a_i, \; pro \; k > Q, \tag{2.5}$$

kde $c(k)$ jsou samotné kepstrální koeficienty LPC, $a(k)$ jsou odhadnuté koeficienty LPC lineárního filtru a Q je řád tohoto filtru (artikulačního traktu).

Jelikož kepstrální koeficienty LPC s vyššími indexy nabývají standardně nižších hodnot, je z praktických důvodů užitečné provést tzv. liftering podle vztahu 2.6, [8].

$$c_{lift}(n) = \left[1 + \frac{L}{2} \sin\left(\frac{\pi n}{L}\right)\right] c(n), \tag{2.6}$$

kde L je váha lifteringu s typickou hodnotou $L = 22$. Díky lifteringu získáme parametry s přibližně podobně velkými hodnotami.

Melovské kepstrální koeficienty MFCC

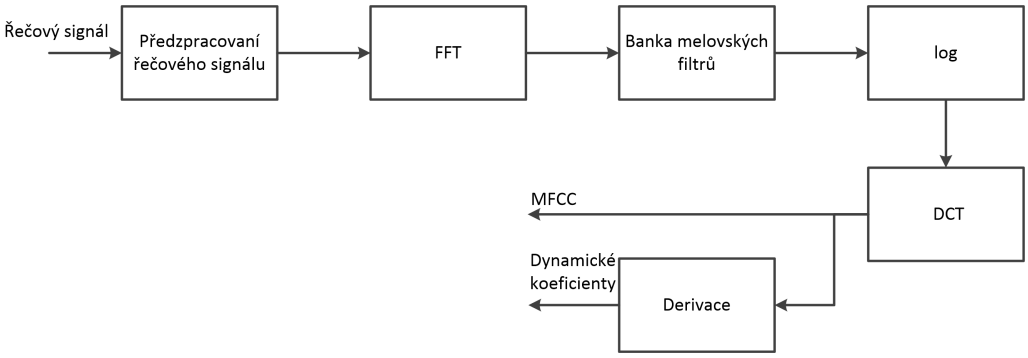
Zpracování řečových signálů pomocí melovských kepstrálních koeficientů přináší největší výhodu v tom, že do jisté míry respektuje nelineární vlastnosti vnímání zvuků lidským uchem (lidské ucho má na nízkých frekvencích větší rozlišení než na vysokých). Této výhody se využívá hlavně u rozpoznávání řeči, kde chceme co nejvíce přiblížit kepstrum slyšení [8], [10], [13], [Tov01].

Melovské kepstrální koeficienty se snaží kompenzovat nelineární vnímání frekvencí pomocí banky trojúhelníkových pásmových filtrů s lineárním rozložením frekvencí v takzvané melovské frekvenční škále [8]. Tato škála je definována vztahem 2.7.

$$f_m = 2595 \log\left(1 + \frac{f}{700}\right), \tag{2.7}$$

kde f je frekvence v lineární škále a f_m je odpovídající frekvence v nelineární melovské škále.

Postup výpočtu melovských kepstrálních koeficientů je zobrazen na obrázku 2.11 a samotný výpočet provedeme pomocí vztahu 2.8, [8], [Tov01].



Obr. 2.11: Postup výpočtu MFCC.

$$c_m(j) = \sum_{i=1}^{M^*} \log y_m(i) \cos\left(\frac{\pi j}{M^*}(i - 0,5)\right), \; pro \; j = 0, 1, ..., M, \tag{2.8}$$

kde $y_m(i)$ jsou odezvy jednotlivých filtrů, M^* je počet pásem melovského filtru a M je počet melovských kepstrálních koeficientů.

2.2.5 Používané metody klasifikace pro autentizaci řečníka

Používané přístupy k automatickému rozpoznávání řečníka lze rozdělit na metody využívající vzorové reprezentace a metody využívající pravděpodobnostních modelů. V praxi se můžeme nejčastěji setkat s metodami, které využívají pravděpodobnostních modelů. Mezi tyto metody patří [1], [8], [29], [30], [Tov04], [32]:

- autentizace řečníka na základě směsi Gaussových hustotních funkcí (GMM),
- autentizace řečníka s využitím skrytých Markovových modelů (HMM),
- autentizace řečníka pomocí umělých neuronových sítí (ANN).
- autentizace řečníka pomocí podpůrných vektorů (SVM).

Autentizace řečníka na základě směsi Gaussových hustotních funkcí (GMM)

Rozdělení pravděpodobností příznakových vektorů získaných z řeči konkrétního řečníka lze popsat směsí Gaussových hustotních funkcí [9], [34], [Tov05], [Tov06], to znamená váženou lineární kombinací hustotních funkcí normálních rozdělení jednotlivých akustických tříd ve tvaru 2.9, [1], [9], [Tov04].

$$p(\vec{x_n}|M_{client}) = \sum_{i=1}^{I} w_i^* N\left(\vec{x_n}, \vec{\mu_i}, \sum_i\right), \tag{2.9}$$

kde $p(\vec{x_n}|M_{client})$ je pravděpodobnost vektoru charakteristiky (likelihood) $\vec{x_n}$ v modelu oprávněného uživatele M_{client}, I je počet mixů, w_i^* je váha pro mix i a gaussovské hustoty N jsou parametrizovány střední hodnotou $Dx1$ vektorem $\vec{\mu_i}$ a DxD kovarianční maticí $\sum_i$.

Při hypotéze nezávislosti charakteristik je globální pravděpodobnostní skóre pro celou sekvenci rámců nahrávky spočítáno jako součin pravděpodobností jednotlivých rámců podle vzorce 2.10, [1], [9].

$$S_c = p(X|M_{client}) = \prod_{n=1}^{N_r} p(\vec{x_n}|M_{client}), \tag{2.10}$$

kde $\vec{x_n}$ je příznakový vektor n-tého rámce a N_r je celkový počet rámců.

Skóre hypotézy, že nahrávka nepochází od oprávněného uživatele, je určeno pomocí univerzálního modelu okolí [9], [32], [34], tento model je trénován maximálním počtem ostatních řečníků [7]. Skóre je dáno vztahem 2.11.

$$S_w = p(X|M_{world}) = \prod_{n=1}^{N_r} p(\vec{x_n}|M_{world}). \tag{2.11}$$

Výsledné skóre se poté v praxi často počítá v logaritmické doméně, jedná se tedy o rozdíl log-pravděpodobností 2.12, [1].

$$R_c = \log(S_c) - \log(S_w). \tag{2.12}$$

Autentizace řečníka s využitím skrytých Markovových modelů (HMM)

Využití skrytých Markovových modelů při rozpoznávání řečníka spočívá v představě o způsobu vytváření řeči člověkem. Předpokladem je, že hlasové ústrojí mluvícího člověka je během mikrosegmentu v jednom z konečného počtu stavů a při tom je generován krátký řečový signál odpovídající aktuálnímu nastavení hlasového ústrojí. Při promluvě pak hlasové ústrojí přechází z jednoho stavu do jiného. Proces tvorby promluvy může být modelován pomocí podpůrného Markovova řetězce, který představuje časovou posloupnost stavů hlasového ústrojí a s ním spojeným řetězcem vektorů příznaků, které reprezentují spektrální charakteristiky řečového signálu v daném stavu [1], [8], [30], [32]. Přechody mezi stavy jsou popsány pravděpodobnostmi přechodu a_{ij} a příslušnost vektorů k jednotlivým stavům je dána výstupní pravděpodobností b_j. Markovův model poté může být popsán rovnicí 2.13.

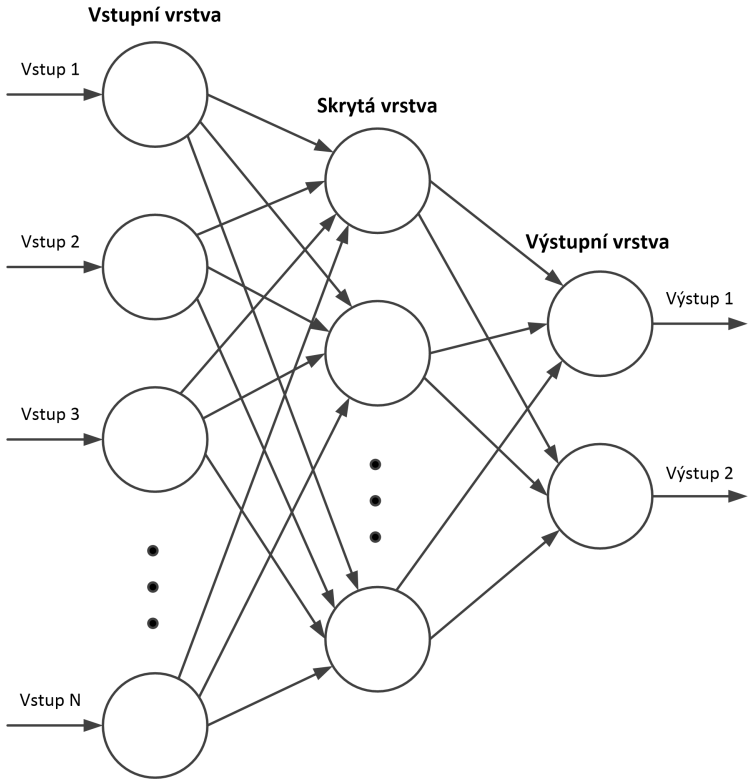
$$M_{client} = \left\{A^s, \vec{b^s}\right\}, \tag{2.13}$$

kde $A^s = [a_{ij}^s]$ je matice pravděpodobností přechodů pro řečníka s a $\vec{b^s} = [b_j^s]$ je vektor výstupních pravděpodobností. Jednotlivé prvky se pak určují stejně jako u GMM v průběhu trénování systému.

Pravděpodobnost $p(X|M_{client})$ potřebnou pro verifikaci řečníka lze určit tak, že je spočítána pravděpodobnost, se kterou mohla být posloupnost X generována všemi možnými posloupnostmi stavů délky Markovova modelu M_{client}. Pro výpočet této pravděpodobnosti se často používá například Viterbiův algoritmus [30], [32].

Autentizace řečníka pomocí umělých neuronových sítí (ANN)

Umělé neuronové sítě představují další prostředek pro autentizaci řečníka. V poslední době se stávají neuronové sítě nejčastěji používaným klasifikátorem v této oblasti díky své přesnosti a možnosti jednoduché adaptace na konkrétní data [1], [11], [12]. Nejčastěji se používají takzvané vícevrstvé neuronové sítě (MLNN) , které se skládají nejméně ze třech vrstev (vstupní vrstva, skrytá vrstva a výstupní vrstva) [30], [33]. Mezi těmito vrstvami existuje takzvané úplné propojení neuronů (každý neuron nižší vrstvy je spojen se všemi neurony vrstvy vyšší). Každá vazba je ohodnocena příslušnou synaptickou vahou. Struktura takové neuronové sítě je zobrazena na obrázku 2.12. Hodnota excitace neuronu y_{ex} (hodnota výstupu z neuronu) je dána aktivační funkcí neuronu (aktivační funkce mohou mít různý průběh, jako základní se používá tvar sigmoidy), funkce sigmoidy je popsána vztahem 2.14.



Obr. 2.12: Vícevrstvá neuronová síť.

$$y_{ex} = S(z) = \frac{1}{1 + e^{-\lambda z}}, \tag{2.14}$$

kde $S(z)$ je aktivační funkce neuronu, z je vnitřní potenciál neuronu a λ je strmost sigmoidu.

$$z = \sum_{i=0}^{n} w_i x_i, \tag{2.15}$$

kde w_i je i-tá hodnota váhy, x_i je i-tá hodnota signálu a n je počet vazeb do neuronu.

Počet neuronů vstupní vrstvy je dán počtem extrahovaných parametrů pro daný řečový segment. Výstupní neurony poté představují jednotlivé klasifikační třídy (pro autentizaci bude mít neuronová síť vždy dva výstupní neurony (autorizovaný/neautorizovaný uživatel). Výstup každého neuronu výstupní vrstvy poté představuje posteriorní pravděpodobnost dané třídy [27]. Počet neuronů skryté vrstvy je volen náhodně, ale vždy by měl být počet neuronů skryté vrstvy nižší než počet neuronů vstupní vrstvy a vyšší než počet neuronů výstupní vrstvy [30], [Tov04], [33].

Princip činnosti neuronové sítě lze rozdělit do dvou částí: šíření signálu a adaptace neuronové sítě. Níže jsou popsány pouze základní metody jednotlivých částí [Tov01]. Ostatní metody jako například QuickProp, Quasi-Newton a tak dále vychází právě z těchto základních metod.

- Šíření signálu - metoda dopředného šíření (feedforward).
  V prvním kroku metody dopředného šíření jsou excitovány na odpovídající úroveň (hodnoty 0 až 1) neurony vstupní vrstvy. Následně jsou tyto excitace pomocí příslušných vazeb přivedeny k následující vrstvě a upraveny (zesíleny nebo zeslabeny) pomocí synaptických vah. Každý neuron vyšší vrstvy provede sumaci upravených signálů od neuronů nižší vrstvy a je excitován na úroveň danou svou aktivační funkcí. Tento proces probíhá přes všechny vnitřní vrstvy až k vrstvě výstupní, kde pak získáme excitační stavy všech jejich neuronů [Tov01].
- Adaptace neuronové sítě - metoda zpětného šíření (backpropagation).
  K adaptaci neuronové sítě se používá tzv. trénovací množina, která obsahuje vektory extrahovaných příznaků s příslušným ohodnocením (1 pokud se jedná o daného řečníka, 0 pokud se nejedná o daného řečníka) [Tov01]. Trénovací množinu poté může být definována jako množina vzorů T 2.16.

$$T = \left\{ \left\{ \vec{I_1}, \vec{O_1} \right\} \left\{ \vec{I_2}, \vec{O_2} \right\} ... \left\{ \vec{I_i}, \vec{O_i} \right\} ... \left\{ \vec{I_p}, \vec{O_p} \right\} \right\}, \tag{2.16}$$

  kde $\vec{I_i}$ je vektor excitací vstupní vrstvy, $\vec{O_i}$ je vektor excitací výstupní vrstvy a p je počet vzorů trénovací množiny.
  V prvním kroku metody zpětného šíření je použit první vektor trénovací množiny $\vec{I_1}$, kterým jsou excitovány neurony vstupní vrstvy na odpovídající úroveň. Metodou dopředného šíření je provedeno šíření signálu až k výstupní vrstvě. Následně je srovnán požadovaný stav daný vektorem $\vec{O_1}$ se skutečnou odezvou neuronové sítě. Rozdíl mezi skutečnou a požadovanou odezvou definuje chybu neuronové sítě. Tato chyba je pak v určitém poměru (koeficient učení - learning rate) vracena zpět do neuronové sítě formou úprav synaptických vah 2.18 mezi jednotlivými vrstvami směrem od horních vrstev k vrstvám nižším tak, aby chyba při následující odezvě byla menší. Po vyčerpání celé trénovací množiny se vyhodnotí celková chyba E 2.17 přes všechny vzory trénovací množiny a pokud je tato chyba vyšší než požadovaná, tak

se celý proces opakuje znovu [Tov01].

$$E = \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{m}(y_j - o_j)_i^2, \tag{2.17}$$

kde y_j je skutečná odezva j-tého neuronu výstupní vrstvy, o_j je požadovaná odezva j-tého neuronu výstupní vrstvy daná vzorem trénovací množiny, p je celkový počet vzorů trénovací množiny a m je počet neuronů výstupní vrstvy.

$$\Delta w_i = -\eta\frac{\partial E}{\partial w_i} + \mu\Delta w_i^{'}, \tag{2.18}$$

kde η je koeficient učení, μ je koeficient změny vah z předchozího kroku (0 až 1) a $\Delta w_i^{'}$ je změna synaptické váhy z předchozího kroku.

Autentizace řečníka pomocí podpůrných vektorů (SVM)

Metoda Support vector machine patří stejně jako neuronové sítě do metod strojového učení. Tato klasifikační metoda je jednou z nejpoužívanějších a je určena především pro binární klasifikaci (klasifikaci do dvou tříd) [35]. Úkolem při trénování klasifikátoru je nalezení nadroviny, která v prostoru příznaků optimálně rozděluje trénovací data. V případě nalezení optimální nadroviny leží body jednotlivých tříd v opačných poloprostorech a zároveň je vzdálenost bodů od nadroviny co největší. To znamená, že na obě strany od nadroviny je co nejširší pruh (takzvaný maximal margin), ve kterém se body nenacházejí. Pro popis této nadroviny se používají pouze nejbližší body (podpůrné vektory - support vectors). Princip metody je zobrazen na obrázku 2.13. Počet těchto vektorů může být maximálně roven dimenzi problému navýšené o jedna. Rozdělující nadrovina je lineární funkcí v prostoru příznaků. Metoda SVM je ve své základní verzi určena pro řešení lineárně separetovatelných úloh. Aby byla metoda schopna řešit původně lineárně neseparetovatelné úlohy využívá jádrové transformace (kernel methods), které provedou převod původní lineárně neseparetovatelné úlohy na lineárně separetovatelnou pomocí transformace příznaků do vyšší dimenze.

Trénování je založeno na minimalizaci velikosti vektoru vah 2.19, kde musí být splněny podmínky uvedené v rovnici 2.20.

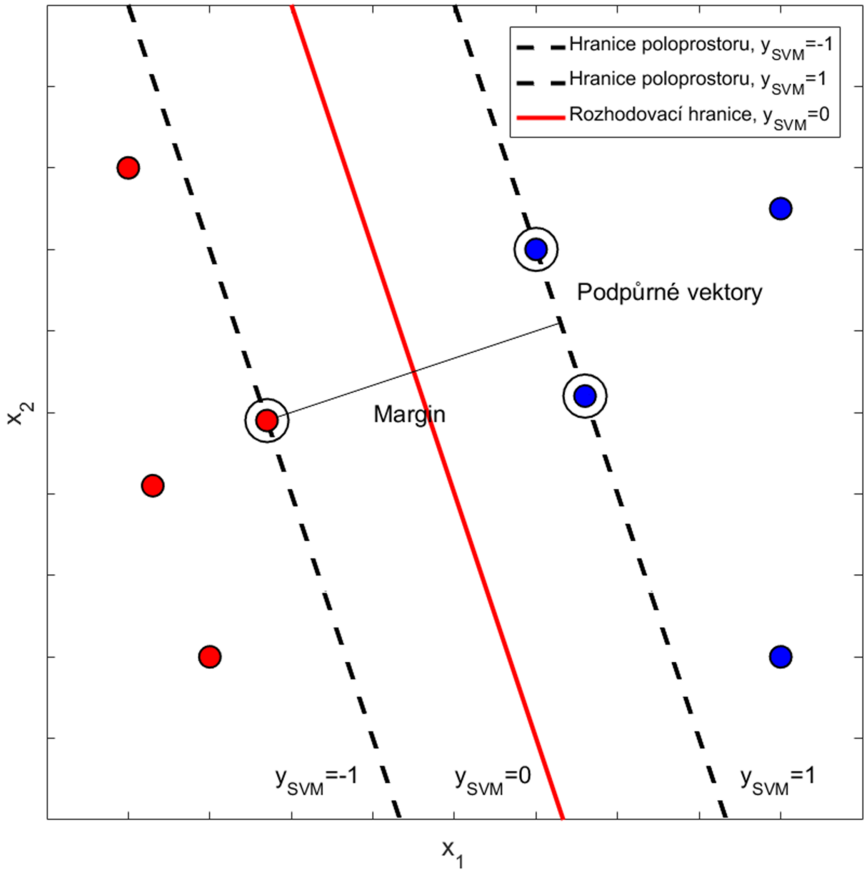
$$|\vec{w}_{SVM}| = \frac{1}{2}\vec{w}_{SVM}^{T}\vec{w}_{SVM}, \tag{2.19}$$

kde $\vec{w}_{SVM}$ je vektor vah SVM klasifikátoru.

$$t_n(\vec{w}_{SVM}^{T}\vec{X}_i + b) \geq 1,\ pro\ i \in \{1, 2, ..., K\}, \tag{2.20}$$

kde $t_n \in \{-1, 1\}$ jsou třídy pro trénovací data, b je bias, $\vec{X}_i$ je vektor trénovacích dat a K je počet trénovacích vektorů.

Proces minimalizace v tomto případě představuje takzvanou optimalizaci s omezením. Jelikož lze vektor vah definující rozhodovací hranici vyjádřit jako lineární kombinaci



Obr. 2.13: Princip SVM metody

podpůrných vektorů 2.21, je možné na trénování klasifikátoru pohlížet jako na počítání skalárních součinů mezi trénovacími vektory. Těmito skalárními součiny je následně definovaná rozhodovací nadrovina. V případě vysoce dimenzionálního prostoru se nepočítají skalární součiny explicitně, ale pomocí jádrové funkce.

$$\vec{w}_{SVM} = \sum_{i \in K_P} t_i \beta_i \vec{X}_i, \ kde \ \beta_i > 0, \tag{2.21}$$

kde K_P je počet podpůrných vektorů, β_i je váha pro daný podpůrný vektor a t_i je třída daného podpůrného vektoru.

Klasifikace natrénovaným klasifikátorem se následně provádí pomocí výpočtu skalárního součinu (kernel transformace) mezi podpůrnými vektory a vektory testovacích dat 2.22.

$$y_{SVM} = \sum_{i \in K_P} t_i \beta_i \vec{X}_i \vec{X}_{test} + b, \tag{2.22}$$

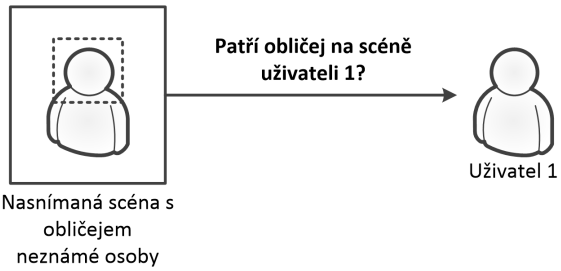
kde y_{SVM} je skóre SVM klasifikátoru a $\vec{X}_{test}$ je vektor testovacích dat.

Výstupem SVM klasifikátoru není přímo pravděpodobnostní interpretace jako například u neuronových sítí, ale takzvané měkké skóre, na základě kterého je provedeno roz-

hodnutí a zařazení do dané třídy. Toto měkké skóre může být na pravděpodobnostní interpretaci transformováno [29], [35], [51].

2.3 Verifikace geometrií obličeje

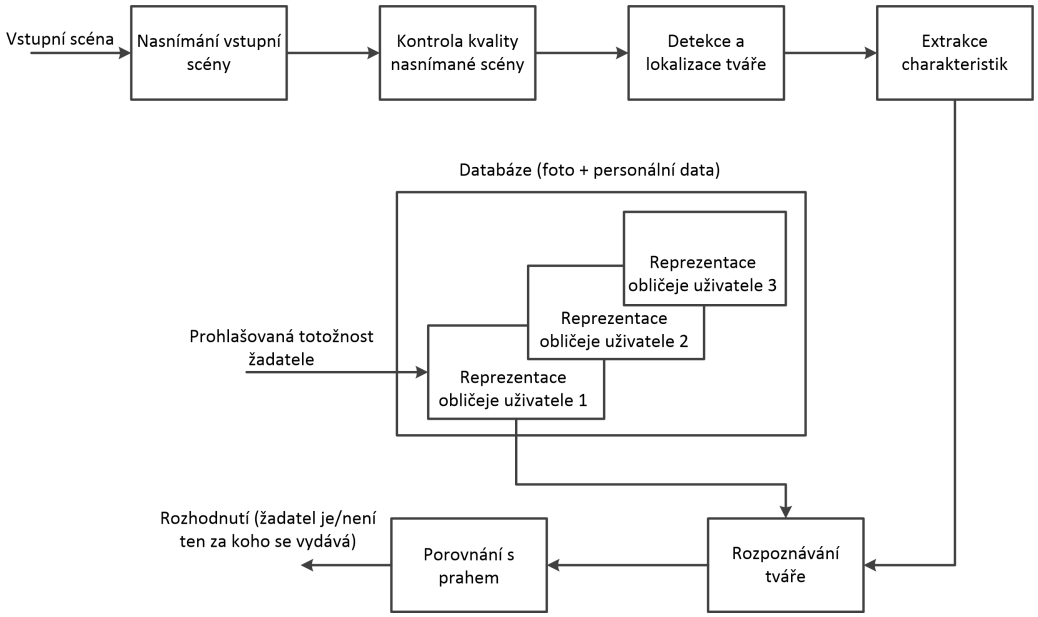
Autentizace obličejem představuje ověření identity člověka na základě geometrie jeho tváře. K dispozici je fotografie (video záznam) lidské tváře neznámé osoby a úkolem je určit, zda neznámá tvář náleží osobě, za kterou se vydává. Mezi hlavní oblasti využití této biometrické metody patří řízení přístupu do objektů a autentizace pro přístup k výpočetním prostředkům. Naopak identifikace pomocí obličeje se nejčastěji využívá u bezpečnostních složek k identifikaci podezřelých osob. Největší výhoda této metody spočívá v bezkontaktním snímání (příjemné pro uživatele), stačí pouze kamera/fotoaparát. Nevýhodou je naopak nižší přesnost v porovnání například s metodou otisků prstů [1], [35], [36], [Tov09].



Obr. 2.14: Verifikace (autentizace) obličejem.

2.3.1 Princip činnosti verifikačního systému geometrií obličeje

Stejně jako u autentizace řečníka, je nutné před samotným procesem autentizace, aby žadatel potvrdil svou prohlašovanou totožnost (identitu). Poté je uživatel informován o snímání své tváře (video nahrávka, foto snímek). Obrazový signál je uložen a následně předzpracován příslušnými metodami zpracování obrazu. V další fázi probíhá detekce a lokalizace tváře v obraze. Pro detekovanou tvář extrahujeme signifikantní příznaky (markanty), které jsou následně použity k vhodné reprezentaci obličeje žadatele. Následně probíhá samotné rozpoznávání tváře, kde je reprezentace obličeje žadatele porovnána s reprezentací obličeje prohlašované totožnosti a určí se míra podobnosti (skóre). Porovnáním skóre s předem nastavenou hodnotou verifikačního prahu je rozhodnuto o tom, zda je žadatel opravdu ten člověk, za kterého se vydává. Pokud je skóre větší než verifikační práh, tak se opravdu jedná o člověka, za kterého se žadatel vydává, v opačném případě nikoliv. Blokové schéma principu činnosti je zobrazeno na obrázku 2.15, [19], [35], [36].



Obr. 2.15: Princip činnosti verifikačního systému geometrií obličeje.

2.3.2 Rozdělení systémů pro verifikaci geometrií obličeje

Systémy pro verifikaci geometrií obličeje je možné rozdělit do několika kategorií podle parametrů zpracovávaného obrazu [1], [16]:

- forma zpracovávaného obrazu - dvourozměrný obraz, třírozměrný obraz,
- typ spektra obrazu - černobílé obrazy, barevné obrazy, infračervené obrazy,
- způsob snímání obrazu - čelní pohled, pohled z boku, obecný pohled,
- časové hledisko obrazu - statické, dynamické.

2.3.3 Detekce a lokalizace tváře

Úkolem detekce a lokalizace tváře je nalezení tváře v daném obraze a určení jejích souřadnic. Princip spočívá v tvorbě modelu lidské tváře, která je následně porovnávána s jednotlivými objekty obrazu. Z výsledku porovnání je následně určeno, zda se jedná o tvář či nikoli. Součástí této fáze bývá zpravidla také předzpracování neboli normalizace nalezeného obrazu tváře, které vede ke zvýšení úspěšnosti samotného rozpoznávání. Mezi metody předzpracování můžeme zařadit rotaci obrazu, translaci obrazu, změnu velikosti obrazu a škálování barev (šedé, barevné) [1], [36], [37].

Používané metody detekce a lokalizace tváře mohou být rozděleny do čtyř hlavních kategorií [36], [37]:

- znalostní metody - tyto metody jsou založené na nadefinovaných pravidlech, která jsou vytvořena na základě znalostí o typickém obličeji, pravidla obvykle popisují vztah mezi typickými rysy obličeje,

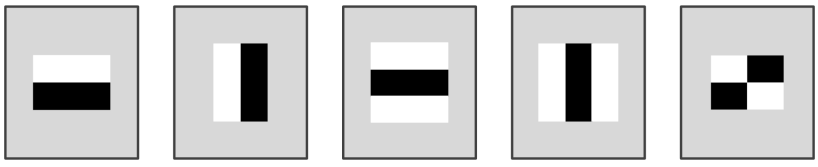
- metody založené na hledání neměnných rysů obličeje - dochází ke hledání typických rysů obličeje, mezi tyto rysy patří například hrany obličejových markantů, barva kůže nebo změna intenzity odstínu mezi markanty,
- metody porovnávající vzory obličeje - hledání tváře spočívá v porovnání vzoru obličeje nebo jeho části s oblastmi nasnímané scény, lze využít deformační model obličeje nebo nadefinovaný model obličeje,
- statisticky orientované metody - tyto metody nahlíží na problém detekce tváře jako na klasifikační problém dvou tříd, první třída reprezentuje obrazy tváře a druhá obrazy, na kterých se tvář nenachází. Tyto metody využívají strojového učení.

Samotné metody spadající do čtyř hlavních kategorií zpravidla neposkytují dostatečnou přesnost detekce a lokalizace. Z tohoto důvodu se většinou využívají kombinace výše uvedených přístupů [37]. V současnosti patří mezi nejpoužívanější metody: metoda Viola-Jones, metoda histogramu orientovaných gradientů a metody využívající konvoluční neuronové sítě [38], [39], [40], [41], [42], [43], [15], [44].

Detekční metoda Viola-Jones

Metoda Viola-Jones je první detekční metoda, která může být použita pro detekci a lokalizaci tváře v reálném čase. Metoda se vyznačuje třemi hlavními rysy (integrální obraz, učící algoritmus AdaBoost, kaskádní struktura). Díky těmto rysům je dosaženo velice rychlé a přesné detekce [41]. Princip metody spočívá v trénování detektoru pomocí algoritmu AdaBoost, který využívá Haarovy příznaky jako slabé klasifikátory. Trénování se provádí pomocí pozitivních a negativních snímků (obsahující/neobsahující tvář), ze kterých jsou vypočítány Haarovy příznaky pomocí integrálního obrazu. Tyto příznaky jsou selektovány pomocí algoritmu AdaBoost a následně složeny do kaskády klasifikátorů. Natrénovaná kaskáda poté představuje samotný detektor [38], [41], [44].

- Haarovy příznaky - hlavní motivací pro využití těchto jednoduchých příznaků je, že systémy založené na příznacích pracují mnohem rychleji a efektivněji než systémy pracující se samotnými pixely. Haarovy příznaky jsou odvozeny od takzvaných Haarových vzorů, které můžeme rozdělit do několika skupin podle typu informace, která má být detekována (hranové, čárové, středové atd.) 2.16.



Obr. 2.16: Ukázka Haarových vzorů vzhledem k detekčnímu oknu

Výpočet samotných příznaků je potom dán jako rozdíl ploch černé a bílé oblasti. Plocha dané oblasti je vypočtena jako skalární součin obrazu a příslušného Haarova

vzoru. Díky integrálnímu obrazu je výpočet příznaků mnohem efektivnější. Počet vypočtených příznaků je dán nastavením velikosti detekčního okna. Zpravidla je toto okno nastaveno na velikost 24x24 pixelů. Hodnoty příznaků jsou počítány pro všechny možné velikosti a posunutí daného vzoru [38], [44].

- Integrální obraz - metoda Viola-Jones využívá integrální obraz z důvodu výrazného urychlení výpočtu Haarových příznaků. Výpočet integrálního obrazu se provádí ze vstupního obrazu tak, že každý pixel integrálního obrazu je roven sumě všech pixelů původního obrazu nacházejících se nad tímto pixelem a vlevo od pixelu. Princip výpočtu je zobrazen na obrázku 2.17 a samotný výpočet se provádí podle rovnice 2.23.

	y_{image1}	y_{image2}	y_{image3}	y_{image4}	y_{image5}
x_{image1}	11	12	13	14	15
x_{image2}	21	22	23	24	25
x_{image3}	31	32	33	34	35
x_{image4}	41	42	**43**	44	45
x_{image5}	51	52	53	54	55

Obr. 2.17: Princip výpočtu integrálního obrazu pro pixel o souřadnicích 4x3.

$$iim(x_{image}, y_{image}) = \sum_{\substack{x'_{image} \leq x_{image} \\ y'_{image} \leq y_{image}}} im(x'_{image}, y'_{image}), \tag{2.23}$$

kde iim jsou hodnoty pixelů integrálního obrazu pro souřadnice x_{image} a y_{image}, im jsou hodnoty pixelů původního obrazu pro souřadnice x'_{image} a y'_{image}.

- Algoritmus AdaBoost - pro každé detekční okno je vypočítáno velké množství Haarových příznaků (více jako samotných pixelů), toto množství je redukováno algoritmem AdaBoost, který provede selekci pouze signifikantních příznaků. Tyto signifikantní příznaky jsou poté použity jako takzvané slabé klasifikátory, které jsou složeny do kaskádní struktury. Princip rozhodování slabého klasifikátoru je popsán rovnicí 2.24, [38], [44].

$$h_w(x_w, f_w, p_w, \Theta) = \begin{cases} 1 & p_w f_w(x_w) < p_w \Theta \\ 0 & \text{jinak,} \end{cases} \tag{2.24}$$

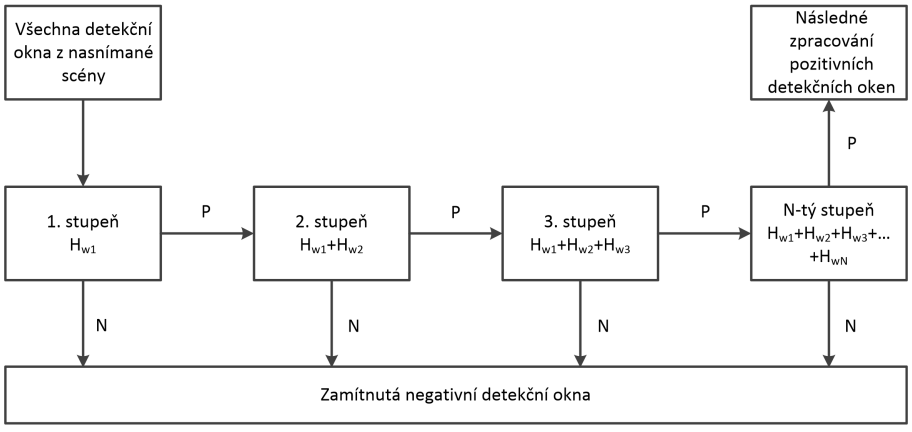
kde h_w je odezva slabého klasifikátoru, f_w je hodnota Haarova příznaku pro detekční okno x_w, p_w je kladná nebo záporná polarita a Θ je prahová hodnota.

Pokud je detekční okno klasifikováno jako pozitivní je odezva klasifikátoru rovna jedné v opačném případě nule. Metoda AdaBoost provádí selekci na základě nejlépe separovaných pozitivních a negativních případů (nejnižší chyba při klasifikaci trénovacích dat).

- Kaskáda klasifikátorů - hlavním úkolem kaskády je urychlení doby detekce obličeje. Kaskáda je složena z několika stupňů, kde každý stupeň obsahuje určitý počet slabých klasifikátorů. Tyto slabé klasifikátory jsou vybrány algoritmem AdaBoost. Jeden stupeň kaskády představuje silný klasifikátor. Funkce silného klasifikátoru je popsána rovnicí 2.25, [41].

$$H_w(x_w) = \begin{cases} 1 & \sum_{v=1}^{V_w} \alpha_v h_{wv}(x_w) \geq \Theta' \\ 0 & \text{jinak,} \end{cases} \tag{2.25}$$

kde H_w je odezva silného klasifikátoru pro detekční okno x_w, α_v je váha slabého klasifikátoru h_{wv} a Θ' je prahová hodnota daného stupně.
Princip činnosti kaskády je založen na předpokladu, že každá nasnímaná scéna obsahuje mnohem větší množství negativních oken než oken obsahujících tváře. Úkolem každého stupně kaskády je zamítnout co největší množství negativních oken a poslat všechna pozitivní okna do dalšího stupně. Následující stupeň už pracuje pouze s pozitivními okny z předchozího stupně. Tento postup probíhá přes všechny stupně kaskády. Na výstupu z kaskády se následně objeví pouze pozitivní okna. Princip kaskádového zapojení je zobrazen na obrázku 2.18, [41].



Obr. 2.18: Kaskádové zapojení klasifikátorů.

Při samotné detekci je nejprve nastavena velikost a pozice detekčního okna, následně je detekční okno klasifikováno kaskádou klasifikátorů a je rozhodnuto, zda se jedná o okno obsahující tvář či nikoli. V dalších krocích je detekční okno posouváno přes celou nasnímanou scénu. Velikost detekčního okna je zvětšována až do té doby, než je detekční okno větší jak nasnímaná scéna. Na všechny nalezené detekce je následně aplikované

takzvané následné zpracování, kde dochází ke sjednocení vícenásobných detekcí jedné tváře [38], [41], [44].

Detekční metoda histogramu orientovaných gradientů - HOG

Metoda je založená na nalezení a spočítání výskytů orientovaných gradientů hran v pozitivních a negativních částech obrazu. Takto získané příznaky jsou poté použity pro natrénovaní detektoru pomocí některého z algoritmů strojového učení.

Nasnímaná scéna je rozdělena do menších oblastí takzvaných buněk. Buňky zachycují určitou oblast nasnímané scény a zpravidla mají velikost 8x8 pixelů. Pro každou buňku je vypočítán histogram orientovaných gradientů, který je normalizován vzhledem k takzvanému bloku, což je seskupení několika sousedních buněk. Histogram je vypočítán ze všech pixelů dané buňky. Orientace gradientů se může pohybovat v rozsahu 0°-360° což představuje velké množství směrů. Z tohoto důvodu je vhodné tento rozsah rozdělit do několika kanálů (9 kanálů, rozsah pro jeden kanál 40°). Histogram poté nese informaci o počtu gradientů v daném kanále. Velikost histogramu odpovídá počtu kanálů. Vektor normalizovaných histogramů pro jeden blok poté představuje příznakový vektor. Pro každý blok získáme jeden příznakový vektor. Celý proces detekce lze rozdělit do následujících čtyř kroků [42], [45], [46], [Tov09].

- Předzpracování nasnímané scény - tento krok vede ke zvýšení úspěšnosti samotné detekce. V rámci předzpracování se v této metodě může využít převedení obrazu do úrovní šedi a normalizace [42].

- Výpočet gradientů - pro každý pixel v nasnímané scéně je vypočítán gradient. Gradient můžeme definovat jako vektor parciálních derivací obrazových dat pro osy x a y 2.26.

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right), \tag{2.26}$$

  kde $\frac{\partial f}{\partial x}$ je gradient obrazové funkce f ve směru x a $\frac{\partial f}{\partial y}$ je gradient obrazové funkce f ve směru y.
  Výpočet gradientů pro jednotlivé směry můžeme provést jako konvoluci obrazové funkce a takzvané masky, která představuje filtr 2.27. Maska se pro jednotlivé směry liší pouze v orientaci, pro osu x se jedná o řádkový vektor, pro osu y se jedná o sloupcový vektor stejné velikosti [45].

$$\frac{\partial f}{\partial x} = [-1, 0, 1] * f, \tag{2.27}$$

  Pro každý bod nasnímané scény je také vypočítán modul gradientu 2.28.

$$\|\nabla f\| = \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial x} \right)^2}, \tag{2.28}$$

Na závěr je vypočítán úhel (směr) gradientu podle rovnice 2.29.

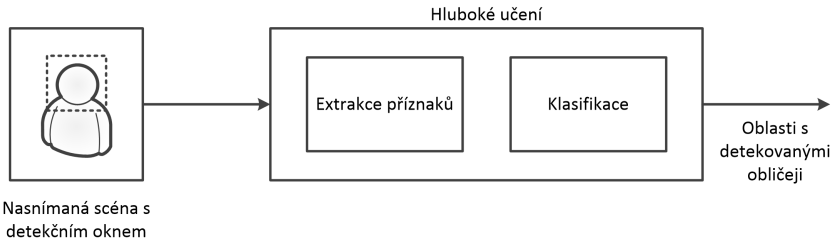
$$\psi = \arctan \frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial x}}, \tag{2.29}$$

- Normalizace bloků - po výpočtu gradientu je nasnímaná scéna rozdělena na buňky a pro každou buňku je určen histogram jejích gradientů. Velikost histogramu je dána počtem kanálů, do kterých je rozdělen celkový rozsah možných směrů (0°-360°). Optimální počet kanálů je podle literatury [42] nastaven na hodnotu 9 (kanály jsou rozděleny po 40°). Četnost jednotlivých kanálů histogramu je dána počtem odpovídajících úhlů gradientů. Buňky jsou následně seskupeny do bloků. Blok je zpravidla o velikosti 3x3 buňky. Všechny buňky jsou následně normalizovány v rámci bloku do kterého patří. Tato normalizace vede ke zvýšení přesnosti detekce. Hodnoty normalizovaných buněk představují výsledné extrahované parametry [42], [45].

- Klasifikace - pro samotnou klasifikaci na pozitivní a negativní části obrazu se využívají algoritmy strojového učení. Mezi nejpoužívanější algoritmy v tomto případě patří Support vector machines, popřípadě vícevrstvé neuronové sítě popsané v podkapitole 2.2.5, [45], [46].

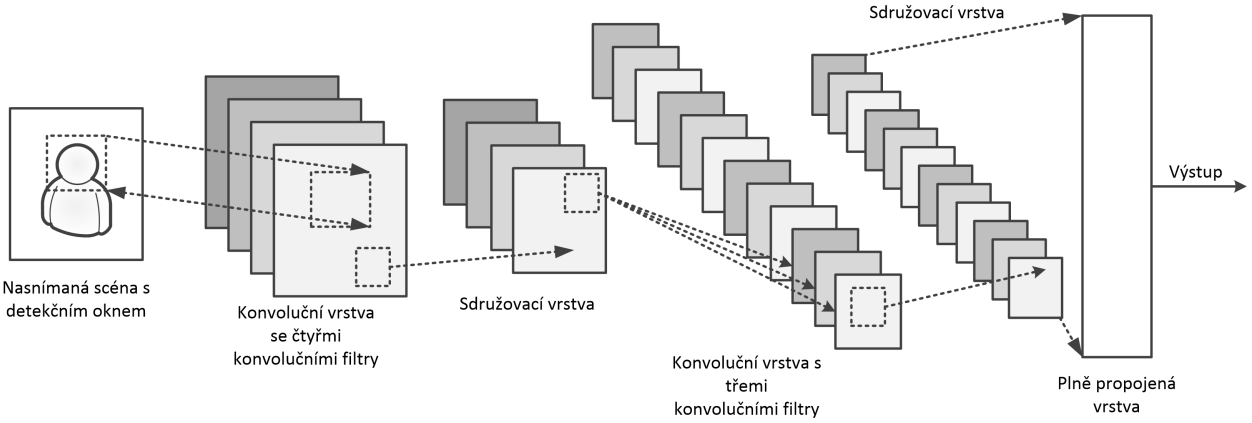
Detekční metody využívající konvoluční neuronové sítě

Metody využívající konvoluční neuronové sítě v případě detekce a lokalizace obličeje nezajišťují pouze samotnou detekci, ale jsou využity pro řešení komplexního problému, kterého je detekce součástí. Tyto sítě mohou být také využity pro samotnou autentizaci uživatele [48]. Konvoluční neuronové sítě jsou nejpoužívanějším algoritmem takzvaného hlubokého učení (Deep Learning). Největší výhodou těchto sítí je, že celý problém je řešen komplexně a to s velikou přesností. Princip hlubokého učení je zobrazen na obrázku 2.19. V případě



Obr. 2.19: Princip hlubokého učení.

detekce obličeje je vstupem do sítě přímo nasnímaná scéna bez předchozího předzpracování a výstupem jsou oblasti s detekovanými obličeji a jejich souřadnicemi [47]. To znamená, že samotná síť provede všechny kroky nutné pro detekci (extrakce parametrů, detekce a lokalizace obličeje). Schéma konvoluční neuronové sítě je zobrazeno na obrázku



Obr. 2.20: Schéma konvoluční neurnové sítě.

2.20. Síť se skládá ze vstupní vrstvy, výstupní vrstvy, konvoluční vrstvy, sdružovací vrstvy a plně propojené vrstvy. Změna vah se provádí stejně jako u vícevrstvých neuronových sítí pomocí metody zpětného šíření, která byla popsána v podkapitole 2.2.5, [39], [40].

- Vstupní vrstva - u konvoluční neuronové sítě je vstupní vrstva reprezentována samotnou nasnímanou 2D scénou (matice pixelů).
- Konvoluční vrstva - tyto vrstvy jsou základem celé sítě. Každá konvoluční vrstva obsahuje definovaný počet konvolučních filtrů. Jednotlivé filtry představují samotné neurony. Každý filtr je určen pro extrakci jiného příznaku. Výstupem této vrstvy jsou mapy příznaků. Počet těchto map je dán počtem neuronů v dané vrstvě.
- Sdružovací vrstva - tyto vrstvy bývají zpravidla umístěny za vrstvami konvolučními a to z důvodu redukce (zmenšení rozlišení) počtu prvků zpracovávané části obrazu. Jedná se o takzvané podvzorkování, které se provádí například pomocí funkcí MAX nebo MEAN. V případě funkce MAX je v definované oblasti vybrán pixel s maximální hodnotou a tímto pixelem je následně reprezentována celá oblast.
- Plně propojená vrstva - tato vrstva slouží pro propojení všech výstupů předchozí vrstvy do jednoho výstupu, stejně jako u vícevrstvých neuronových sítí popsaných v podkapitole 2.2.5.
- Výstupní vrstva - tato vrstva je reprezentována pozitivními a negativními částmi obrazu (obsahující/neobsahující tvář).

Největší využití nalézají tyto sítě v oblastech, kde je nutné detekovat a rozpoznávat velké množství tváří v nasnímané scéně a při různých podmínkách (světlo a poloha). Nevýhodou je vysoká výpočetní náročnost, která je způsobena potřebou velkého množství trénovacích dat (řádově miliony) [47] [48] [49]. Samotné trénování poté může zabrat i několik týdnů. Z tohoto důvodu není praktické využít tento přístup pro autentizaci, kde nemáme k dispozici tak velké množství trénovacích dat a může docházet k častému přetrénování neuronové sítě v závislosti na požadavcích uživatele.

2.3.4 Obrazové příznaky používané pro rozpoznávání obličeje

Obrazové příznaky musí stejně jako hlasové parametry splňovat určité požadavky. Tyto požadavky jsou uvedeny v podkapitole 2.2.3. U obrazových příznaků je navíc kladen důraz na invarianci vůči osvětlení popřípadě na invarianci vůči rotaci [35], [36], [Tov09].

Většina z používaných parametrů takzvaných deskriptorů je založena na vyjádření textury (organizace obličejové oblasti - hrany, odstíny šedi atd.), která je pro každý obličej odlišná. Mezi nejpoužívanější deskriptory můžeme zařadit histogramy orientovaných gradientů, lokální binární vzory nebo lze využít konvoluční neuronové sítě [15], [50], [48], [Tov09]. Postup extrakce histogramů orientovaných gradientů je obdobný jako při detekci obličeje a je podrobně popsán v podkapitole 2.3.3.

Konvoluční neuronové sítě v tomto případě opět nezajišťují pouze extrakci parametrů, ale mohou být využity přímo pro samotnou autentizaci, kde je extrakce parametrů pouze jedním z kroků. Princip činnosti neuronové sítě je obdobný jako v podkapitole 2.3.3 s tím rozdílem, že na vstup sítě je přiveden přímo obrázek detekované tváře a na výstupu sítě je rozhodnutí, zda se jedná o referenčního uživatele či nikoli [47], [48].

Lokální binární vzory

Extrakce lokálních binárních vzorů se využívá hlavně pro analýzu obrazových dat obsahujících texturu. Výhodou těchto parametrů je robustnost vůči osvětlení, invariance vůči rotaci a výpočetní nenáročnost. Metoda pracuje s obrazy v odstínech šedi [50].

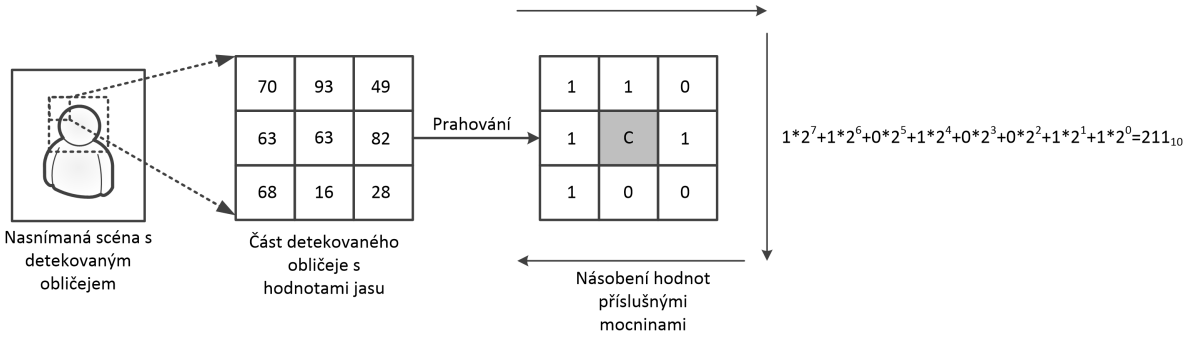
Princip výpočtu jednotlivých vzorů je založen na porovnávání jasu středového pixelu s jasem sousedních pixelů. Výpočet jednoho vzoru se provádí pro okolí o velikosti 3x3 pixely. Prvním krokem pro získání lokálního binárního vzoru je takzvané prahování hodnot okolí vzhledem k hodnotě středového pixelu (pixelům s vyšší nebo stejnou úrovní jasu než má středový pixel je přiřazena hodnota 1, ostatním 0). Výpočet LBP vzoru pro daný bod je proveden pomocí rovnice 2.30, [50].

$$LBP(x_c, y_c) = \sum_{u=0}^{U} s(I_u - I_c)2^u, \tag{2.30}$$

kde x_c a y_c jsou souřadnice pixelu pro který je vzor počítán, I_c je hodnota jasu středního pixelu, I_u je hodnota jasu sousedního pixelu, $s(I_u - I_c)$ je prahovací funkce 2.31 a U je počet sousedních pixelů.

$$s(I_u - I_c) = \begin{cases} 1 & (I_u - I_c) \geq 0 \\ 0 & (I_u - I_c) < 0. \end{cases}, \tag{2.31}$$

Tímto způsobem jsou vypočítány LBP vzory pro všechny pixely v nasnímané scéně a následně je z nich vytvořen histogram, který představuje extrahované příznaky. Postup výpočtu jednoho vzoru je zobrazen na obrázku 2.21, [50], [52], [53].



Obr. 2.21: Postup výpočtu jednoho LBP vzoru.

2.3.5 Používané metody klasifikace pro autentizaci geometrií obličeje

Pro účely autentizace pomocí geometrie obličeje je možné využít obdobné klasifikátory jako u rozpoznávání řečníka. Mezi nejčastěji používané patří vícevrstvé neuronové sítě, konvoluční neuronové sítě nebo Support vector machines [18], [35], [36].

Princip klasifikace pomocí vícevrstvé neuronové sítě je totožný s využitím tohoto klasifikátoru pro autentizaci řečníka s tím rozdílem, že neuronová síť je trénována pomocí parametrů extrahovaných pro daný obličej. Výstupní vrstva opět obsahuje dvě třídy (oprávněný/neoprávněný uživatel) [18]. Popis vícevrstvé neuronové sítě je uveden v podkapitole 2.2.5. V případě využití SVM klasifikátoru je princip klasifikace také obdobný jako u verifikace řečníka. Popis klasifikace pomocí podpůrných vektorů je uveden v podkapitole 2.2.5. Využití konvolučních neuronových sítí pro klasifikaci je popsáno v podkapitole 2.3.4, [48].

2.4 Vícenásobné biometrické systémy

Vícenásobné (multimodální) biometrické systémy se vyznačují tím, že využívají několik zdrojů informací na rozdíl od jednoduchých biometrických systémů, které využívají zdroj pouze jeden. Tato skutečnost vede ke zvýšení přesnosti těchto systémů. Použitím více zdrojů dojde k omezení nebo potlačení nežádoucích vlastností jednoduchých systémů (šum v nasnímaných datech, variabilita mezi uživateli, variabilita v rámci jednoho uživatele atd.). Propojení informací z více zdrojů může probíhat na několika úrovních a zároveň mohou být využity různé strategie získávání informací. Podle použitých zdrojů informací lze definovat jednotlivé typy (strategie) vícenásobných systémů [20], [23]:

- vícenásobné senzory - jedna biometrická charakteristika je snímána více senzory,
- vícenásobná biometrie - více biometrických charakteristik je využito v jednom systému,
- vícenásobné snímání - jedna biometrická charakteristika je několikrát nasnímána,

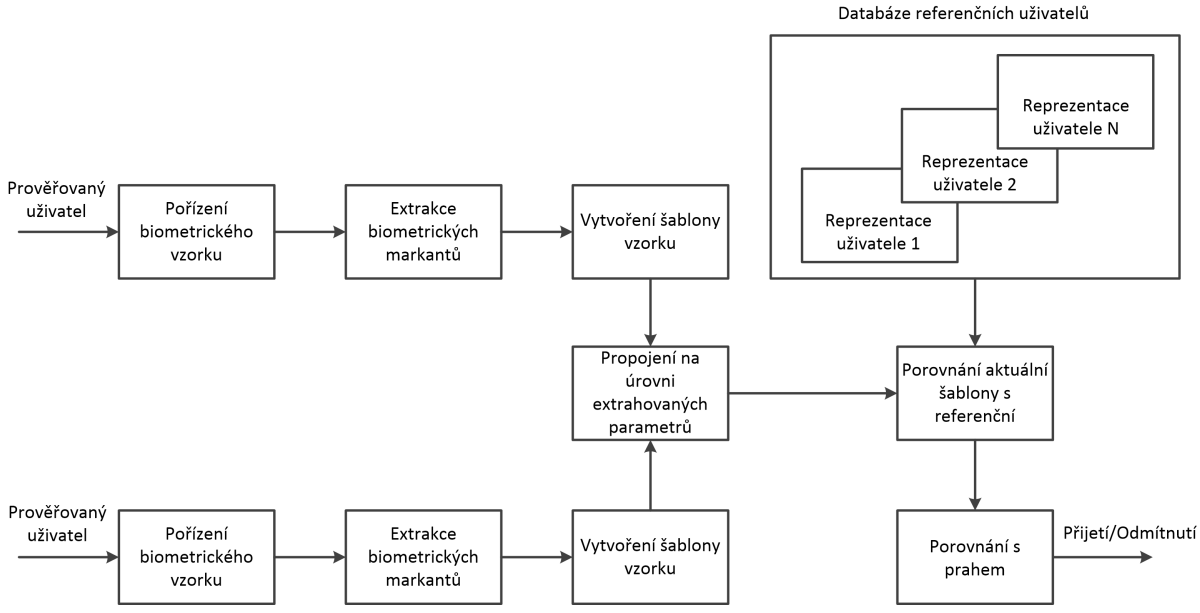
- vícenásobné parametry - pro jednu biometrickou charakteristiku extrahujeme několik různých parametrů,
- vícenásobná klasifikace - jedna biometrická charakteristika je vyhodnocována více klasifikátory.

Jednotlivé typy vícenásobných systémů lze libovolně kombinovat a propojovat. Složitější kombinace mohou sice přinášet lepší výsledky v oblasti přesnosti, ale na úkor efektivity celého systému [20]. V další části je podrobně věnována pozornost pouze vícenásobné biometrii, která byla stěžejně použita v praktické části dizertační práce.

2.4.1 Vícenásobná biometrie

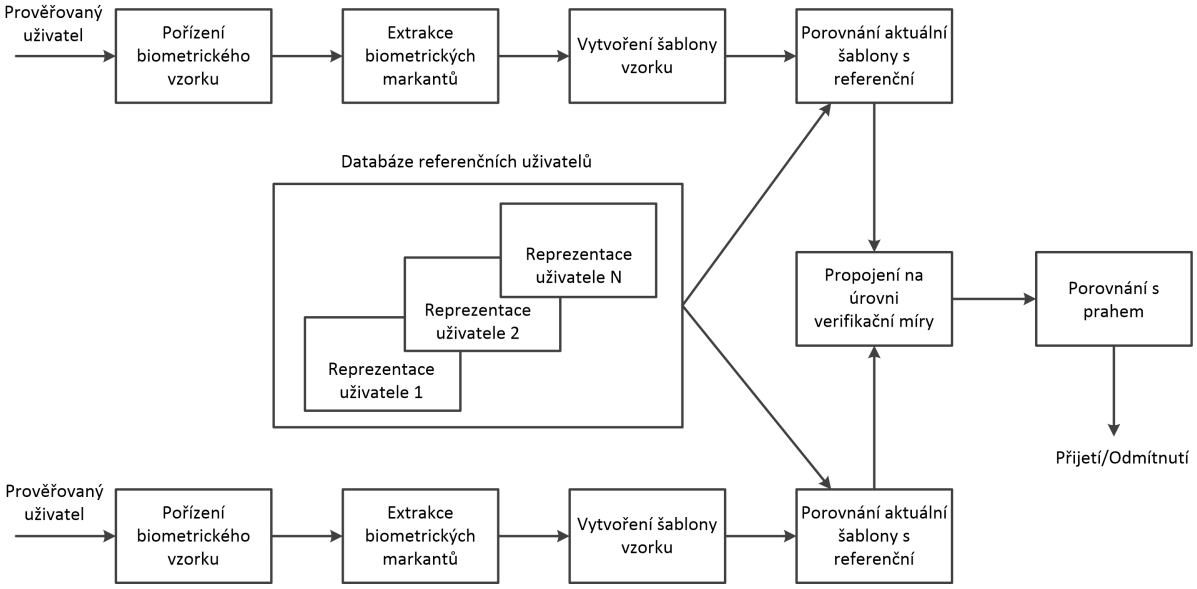
Vícenásobná biometrie představuje kombinaci více biometrických charakteristik pro verifikaci v jednom systému. Samotnou kombinací dojde ke snížení nežádoucích vlastností jednoduchých biometrických systémů (vyhodnocování zašuměných dat, nízká variabilita mezi uživateli, vysoká variabilita pro jednoho uživatele, nedostatečná univerzálnost, možnost podvržení identity), které negativně ovlivňují jak přesnost, tak bezpečnost systému. Velkou výhodou této strategie je nekorelovatelnost jednotlivých biometrických charakteristik. Z tohoto důvodu je potlačení nežádoucích vlastností ještě markantnější než u ostatních strategií vícenásobných systémů. Kombinaci neboli fúzi charakteristik můžeme provést na několika úrovních [20], [23], [26], [28], [29]:

- propojení na úrovni extrahovaných parametrů - dochází k propojení extrahovaných parametrů z jednotlivých snímačů 2.22, výsledkem je poté jeden vektor parametrů,



Obr. 2.22: Blokové schéma vícenásobného biometrického autentizačního systému - propojení na úrovni extrahovaných parametrů.

- propojení na úrovni verifikační míry - dochází ke kombinaci skóre, která jsou generována jednotlivými klasifikátory 2.23, v tomto případě je nutné provést normalizaci obou skóre,



Obr. 2.23: Blokové schéma vícenásobného biometrického autentizačního systému - propojení na úrovni verifikační míry.

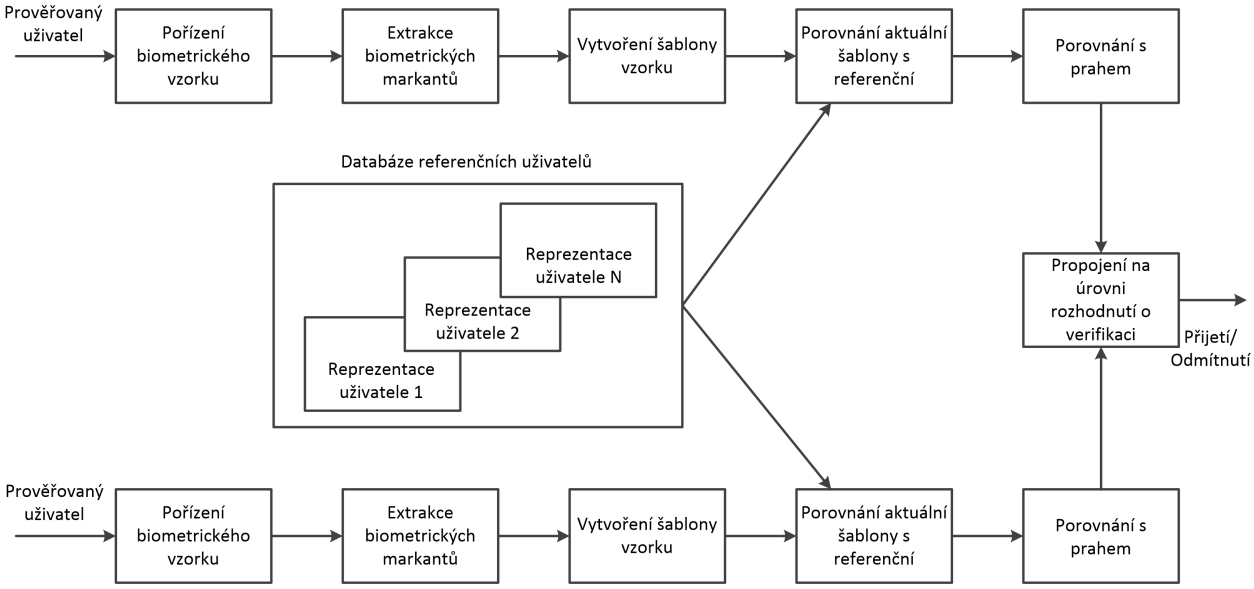
- propojení na úrovni rozhodnutí o verifikaci - k výslednému rozhodnutí dochází na základě rozhodnutí jednotlivých podsystémů 2.24, můžeme například využít AND/OR pravidlo [21], [28], [29].

V současné době patří mezi nejpoužívanější kombinace charakteristik otisk prstů - geometrie obličeje [23], [24] a geometrie oční duhovky - hlas [2], [5]. Lze očekávat, že v brzké době přibudou i další kombinace biometrických charakteristik. Z pohledu úrovně propojení se nejčastěji využívá varianta propojení na úrovni verifikační míry [20], [23], [25], [28], [29]. Z tohoto důvodu je podrobný popis problematiky propojení věnován pouze této variantě.

2.4.2 Propojení na úrovni verifikační míry

Důležitým aspektem při propojení na úrovni verifikační míry je normalizace skóre. Úkolem normalizace je namapování skóre jednotlivých klasifikátorů do jedné domény (například rozsah od 0 do 1). K tomuto úkolu lze využít dostupné normalizační techniky jako jsou min-max normalizace, Z-normalizace, decimální škálování atd. [28]. Z pohledu verifikace může být na propojení na úrovni verifikační míry pohlíženo jako na dva odlišné přístupy. V prvním případě se jedná o klasifikační problém, kdežto v druhém případě o kombinační problém.

Klasifikační přístup vytvoří příznakový vektor pomocí vygenerovaných skóre z jednotlivých klasifikátorů. Takový vektor je poté klasifikován do dvou tříd (přijetí nebo odmít-



Obr. 2.24: Blokové schéma vícenásobného biometrického autentizačního systému - propojení na úrovni rozhodnutí o verifikaci.

nutí). Pro klasifikaci lze využít například rozhodovací strom, SVM , K-NN [Tov07] nebo neuronové sítě [22], [28], [29].

V případě kombinačního přístupu jsou jednotlivá skóre zkombinována tak, že z nich vznikne jedno skalární skóre, na základě kterého je provedeno finální rozhodnutí o přijetí nebo odmítnutí. Kombinaci lze provést buď pro skóre vyjádřené pomocí posteriorních pravděpodobností nebo přímo pro číselné skóre generované jednotlivými klasifikátory. V případě skóre vyjádřeného posteriorními pravděpodobnostmi můžeme využít následující kombinační metody [28]:

- pravidlo o násobení pravděpodobností (product rule) - toto pravidlo je založeno na předpokladu, že příznakové vektory, které vstupují do jednotlivých klasifikátorů jsou statisticky nezávislé. Vstupní vzor (uživatel) je poté přiřazen do výsledné třídy c podle vzorce 2.32.

$$c = \underset{j}{\operatorname{argmax}} \prod_{n=1}^{R} P(\omega_j|\vec{x_n}), \tag{2.32}$$

  kde $P(\omega_j|\vec{x_n})$ je posteriorní pravděpodobnost, že příznakový vektor $\vec{x_n}$ pro n-tý klasifikátor patří do j-té třídy ω_j, v případě verifikace uvažujeme pouze dvě třídy (přijetí/odmítnutí), R je celkový počet klasifikátorů,
- pravidlo o sčítání pravděpodobností (sum rule) - na rozdíl od předchozího pravidla, je u tohoto pravidla brán v úvahu další předpoklad a to, že posteriorní pravděpodobnosti jednotlivých klasifikátorů se moc neliší od apriorních pravděpodobností

jednotlivých tříd. Vstupní vzor poté přiřadíme do výsledné třídy následovně 2.33.

$$c = \underset{j}{\operatorname{argmax}} \sum_{n=1}^{R} P(\omega_j|\vec{x_n}), \tag{2.33}$$

kde význam proměnných odpovídá proměnným z rovnice 2.32,

- pravidlo o maximální pravděpodobnosti (max rule) - toto pravidlo aproximuje průměr posteriorních pravděpodobností pomocí maximální hodnoty. Přiřazení do třídy provedeme pomocí rovnice 2.34.

$$c = \underset{j}{\operatorname{argmax}} \max_{n} P(\omega_j|\vec{x_n}), \tag{2.34}$$

kde max provede výběr maximální hodnoty posteriorní pravděpodobnosti ze všech klasifikátorů pro danou třídu ω_j,

- pravidlo o minimální pravděpodobnosti (min rule) - obdoba předchozího pravidla s tím rozdílem, že využíváme minimální hodnoty posteriorních pravděpodobností. Výslednou třídu získáme pomocí rovnice 2.35.

$$c = \underset{j}{\operatorname{argmax}} \min_{n} P(\omega_j|\vec{x_n}), \tag{2.35}$$

kde min provede výběr minimální hodnoty posteriorní pravděpodobnosti ze všech klasifikátorů pro danou třídu ω_j.

V případě lineární kombinace číselných skóre máme k dispozici následující metody [24], [28], [29]:

- jednoduchý součet skóre (simple sum of score) - v případě jednoduchého součtu je proveden součet jednotlivých hodnot skóre 2.36, která byla vygenerována jednotlivými klasifikátory a následně normalizována některou z uvedených metod (min-max normalizace, Z-normalizace atd.).

$$S_j = \sum_{k=1}^{R} S_{j,k}, \forall j, \tag{2.36}$$

kde S_j je výsledné skóre po fúzi pro j-tou třídu a $S_{j,k}$ je normalizované skóre k-tého klasifikátoru pro j-tou třídu,

- vážený součet skóre (weighted sum of score) - vážený součet skóre je založen na tom, že každému skóre (klasifikátoru) lze přiřadit určitou váhu. Váha může být přidělena buď fixně (hodnota vah je nastavena vždy stejně) nebo adaptivně (nastavení vah probíhá například na základě kvality pořízených vzorků) [29]. Vzorec pro výpočet výsledného skóre poté vypadá následovně 2.37.

$$S_j = \sum_{k=1}^{R} w_k S_{j,k}, \forall j, \tag{2.37}$$

kde S_j je výsledné skóre po fúzi pro j-tou třídu, $S_{j,k}$ je normalizované skóre k-tého klasifikátoru pro j-tou třídu a w_k je váha k-tého klasifikátoru,

- maximální skóre (max-score) - kombinace skóre je v tomto případě dána pouze výběrem největšího skóre z jednotlivých klasifikátorů pro danou třídu. Výběr je dán vztahem 2.38.

$$S_j = \max(S_{j,1}, S_{j,2}, ..., S_{j,R}), \forall j, \tag{2.38}$$

  kde S_j je výsledné skóre po fúzi pro j-tou třídu, $S_{j,k}$ jsou skóre jednotlivých klasifikátorů pro j-tou třídu a R je celkový počet klasifikátorů,
- minimální skóre (min-score) - stejně jako v předchozím případě je kombinace skóre dána výběrem pouze jediné hodnoty z dostupných skóre v tomto případě minimální 2.39.

$$S_j = \min(S_{j,1}, S_{j,2}, ..., S_{j,R}), \forall j, \tag{2.39}$$

kde hodnoty proměnných odpovídají proměnným z rovnice 2.38.

Výsledné skóre je následně porovnáno s vhodně nastaveným rozhodovacím prahem. Na základě výsledku porovnání je provedeno finální rozhodnutí o přijetí/zamítnutí uživatele. Ohodnocení činnosti vícenásobného biometrického autentizačního systému poté probíhá stejně jako u unimodálního (jednoduchého) biometrického systému, jak bylo uvedeno v podkapitole 2.1.5.

3 CÍLE DIZERTAČNÍ PRÁCE

Z předchozí kapitoly, která byla zaměřena na současný stav problematiky biometrických systémů vyplývá řada problémů spojených především s bezpečností a přesností těchto systémů. Na jejich základě byly stanoveny dílčí cíle pro vypracování dizertační práce. Tyto cíle jsou z důvodu větší přehlednosti zformulovány do 4 hlavních bodů:

- Návrh biometrického autentizačního systému založeného na verifikaci hlasem s využitím umělé inteligence.
- Návrh biometrické autentizace založené na rozpoznávání pomocí geometrie obličeje s využitím pokročilých metod vyhodnocování extrahovaných příznaků.
- Vytvoření komplexního vícenásobného biometrického autentizačního systému využívajícího verifikaci hlasem a verifikaci geometrií obličeje.
- Experimentální ověření funkčnosti navrženého vícenásobného autentizačního systému a porovnání přesnosti s aktuálně používanými systémy biometrické autentizace.

Zpracování jednotlivých bodů vede k naplnění hlavního dizertabilního cíle, kterým je vytvoření komplexního vícenásobného biometrického autentizačního systému, který je určen pro bezkontaktní autentizaci a může být využit jako přístupový systém v budovách či pro ověřování přístupu osob k různým zařízením.

4 NÁVRH HLASOVÉHO AUTENTIZAČNÍHO SYSTÉMU

Tato kapitola popisuje návrh biometrického autentizačního systému, který využívá k ověření totožnosti uživatele jeho hlas.

4.1 Princip návrhu hlasového autentizačního systému

Návrh systému probíhal jak z pohledu vhodných parametrů, tak z pohledu vhodného klasifikátoru. V rámci návrhu byly porovnány následující parametry: MFCC, delta MFCC, delta-delta MFCC, LPC a jejich kombinace. Bylo použito 13 koeficientů MFCC, ze kterých byly odvozeny jejich deriváty (delta MFCC a delta-delta MFCC) a 13 LPC koeficientů. Popis použitých parametrů je uveden v podkapitole 2.2.4. Tyto parametry byly následně klasifikovány pomocí dvou klasifikátorů (MLNN a SVM). Struktura MLNN se skládala ze tří vrstev. Počet neuronů vstupní vrstvy odpovídal délce příznakového vektoru, počet neuronů ve skryté vrstvě byl nastaven na 10 a výstupní vrstva byla tvořena dvěma neurony. Jako aktivační funkce neuronů byla použita funkce sigmoid se strmostí 0,5. Struktura a nastavení parametrů neuronové sítě bylo provedeno na základě poznatků z předchozího výzkumu [Tov01]. Pro klasifikaci pomocí SVM byla použita polynomiální funkce konkrétně kubická. Popis použitých klasifikátorů je uveden v podkapitole 2.2.5. Volba signifikantních parametrů a klasifikátorů byla provedena na základě rešerše provedené v kapitole 2 s přihlédnutím ke konkrétním požadavkům na systém. Tyto požadavky vychází ze skutečnosti, že systém má být součástí komplexního vícenásobného biometrického systému.

Prvním krokem hlasové autentizace je předzpracování řečového signálu, které se obvykle skládá ze čtyř fází: odstranění stejnosměrné složky, preemfáze, segmentace (20 ms), váhování oknem (Hammingovo okno). V dizertační práce byl tento krok rozšířen o jednu fázi, která se nazývá odstranění nízkoenergetických segmentů. Druhým krokem je extrakce signifikantních parametrů, po které je aplikována min-max normalizace. Po klasifikaci jednotlivých segmentů je provedena evaluace segmentů v rámci jedné nahrávky pomocí fúzní metody Majority voting [26]. Výsledky evaluace jsou porovnány s verifikačním prahem a následně je provedeno finální rozhodnutí o autentizaci. Jednotlivé přístupy jsou mezi sebou porovnávány pomocí FAR, FRR, EER, ROC a DET. Výsledky byly získány pro databázi Comtech, která vznikla v rámci dizertační práce. Implementace použitých metod pro návrh systému probíhala v prostředí Matlab. Zdrojové kódy včetně komentářů jsou přiloženy na SD kartě.

4.2 Databáze Comtech

V rámci dizertační práce byla vytvořena na Katedře telekomunikační techniky, VŠB-TU Ostrava databáze řečových vzorků s názvem Comtech, která je určena pro návrh a eva-

luaci algoritmů použitých v systémech rozpoznávání řečníka (textově závislých i textově nezávislých). Jedná se o databázi českých řečových vzorků a tudíž je určena především pro tuzemské použití. Databáze byla rozdělena do dvou částí. První část je tvořena nahrávkami 18 referenčních řečníků (13 mužů, 5 žen) a skládá se z 1080 foneticky vyvážených vět a 1080 přístupových frází (hesel). Druhá část obsahuje nahrávky neoprávněných uživatelů, takzvaných podvodníků (uživatelé, kteří se snaží o neoprávněný přístup). Tato část se skládá ze 190 foneticky vyvážených vět a 190 přístupových frází. Věk řečníků se pohyboval v rozmezí 25-50 let.

Řečové vzorky byly pořizovány během jednoho měsíce (03/2017). Nahrávání databáze (část referenční řečníci) bylo rozděleno do 6 sezení. Každý z řečníků byl posazen do kanceláře, kde byl umístěn VoIP telefon (Grandstream GXP2140). Uživatel zadal volbu pro vytočení IVR, které bylo nakonfigurováno pro provedení uživatele nahrávacím procesem. V první části je uživatel vyzván pro zadání svého identifikačního čísla, tímto číslem jsou označeny všechny nahrávky daného uživatele. Následně uživatel pronesl 10 vět a 10 přístupových frází, které byly zaznamenány. Tento proces se opakoval pro všechny referenční řečníky a všechna sezení. Část databáze tvořena neoprávněnými uživateli/podvodníky byla nahrána stejným způsoben, jen s tím rozdílem, že nahrávání probíhalo pouze v jednom sezení. Řečový signál byl zaznamenán ve WAV formátu se vzorkovací frekvencí 8kHz s 32 bitovým rozlišením.

Jako promluvy jednotlivých řečníků byly zvoleny foneticky vyvážené věty a krátké dvojslovné přístupové fráze. Každé sezení obsahovalo 10 různých vět. Doba trvání každé z vět byla přibližně 10 sekund. Jednotlivé věty byly vybírány z elektronických novin. Texty vět jsou uvedeny v příloze B. Přístupová fráze byla stejná během všech sezení a pro všechny řečníky. Doba trvání přístupové fráze byla přibližně 2 sekundy a její znění bylo "Jaromír Továrek".

V rámci dizertační práce byly využity vzorky přístupových frází, které jsou určeny především pro textově závislé rozpoznávání řečníka (textově závislou autentizaci hlasem). Struktura a popis vzorků databáze je uveden v příloze A. Samotná Comtech databáze je dostupná na přiložené SD kartě.

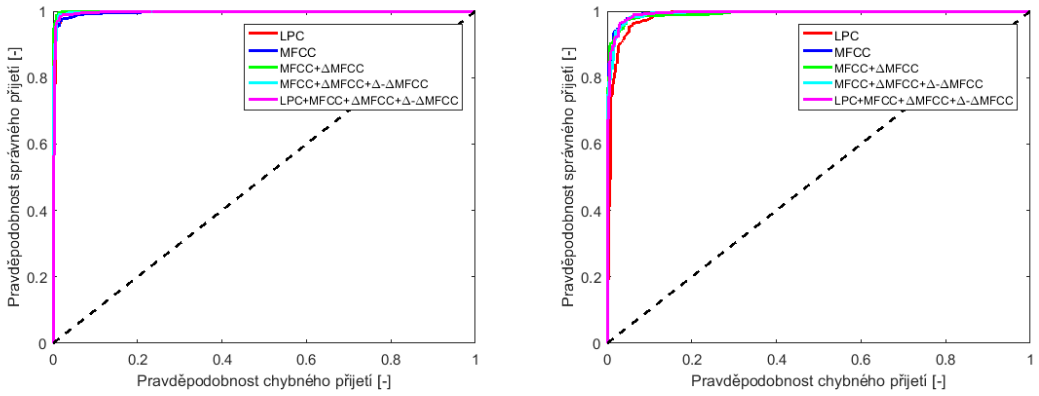
4.3 Experimentální výsledky

Pro každého z 18 referenčních řečníků byly natrénovány dva klasifikátory (SVM model a MLNN klasifikátor). Oba klasifikátory byly natrénovány k rozpoznávání mezi dvěma třídami (referenční uživatel, podvodník). Třída referenční uživatel byla trénována pomocí 40 přístupových frází (4 sezení databáze). Zbývajících 20 nahrávek bylo použito pro testování referenčního uživatele. Třída reprezentující podvodníky byla trénována jako univerzální model pozadí. Tato třída byla trénována 40 nahrávkami od zbývajících 17 referenčních řečníků. Poměr trénovacích dat pro obě třídy byl přibližně stejný. Jako testovací data pro třídu podvodníků byly použity přístupové fráze z druhé části databáze (19 nahrávek od

19 podvodníků). Tyto nahrávky nepochází od referenčních řečníků, to znamená, že model pozadí není trénován nahrávkami od těchto uživatelů. Tento proces trénování klasifikátorů se opakoval pro různé řečové parametry a jejich kombinace. Výsledky pro oba klasifikátory a všechny typy parametrů jsou uvedeny v tabulce 4.1. Tabulka obsahuje hodnoty FAR, FRR pro nastavený rozhodovací práh 0.5 (50%) a hodnotu EER. ROC křivky pro oba klasifikátory jsou zobrazeny na obrázku 4.1.

Tab. 4.1: Výsledky dosažené pomocí SVM a MLNN klasifikátorů pro použité řečové parametry (rozhodovací práh 50%).

Parametry	SVM			MLNN		
	FAR [%]	FRR [%]	EER [%]	FAR [%]	FRR [%]	EER [%]
LPC	5	0.3	1.7	13	1.4	5.3
MFCC	5	1.3	2.3	11	0.8	3.8
MFCC + ΔMFCC	2.3	0.27	**0.85**	7.8	1.6	4.3
MFCC + ΔMFCC + Δ-ΔMFCC	3.5	0.55	1.14	8.5	1.6	3.9
LPC + MFCC + ΔMFCC + Δ-ΔMFCC	2.9	1	1.7	6	1	**3.3**



(a) ROC křivky - SVM klasifikátor

(b) ROC křivky - MLNN klasifikátor

Obr. 4.1: ROC křivky - hlasová autentizace

Nejlepších výsledků bylo dosaženo pro příznakový vektor tvořený parametry MFCC a delta MFCC při použití SVM klasifikátoru. Hodnoty FAR a FRR byly v tomto případě 2.3% a 0.27% pro rozhodovací práh nastavený na 50%. Tyto výsledky odpovídají přesnosti systému 98.7%. Pro tento příznakový vektor byla hodnota EER 0.85%, což odpovídalo nastaveného rozhodovacímu prahu 55%. Pro klasifikátor MLNN bylo dosaženo nejnižší hodnoty EER (3.3%) s použitím příznakového vektoru složeného ze všech použitých parametrů (LPC + MFCC + delta MFCC + delta-delta MFCC). Z pohledu přesnosti klasifikátorů dosáhl lepších výsledků klasifikátor SVM a to pro všechny varianty příznakových vektorů. Z tohoto důvodu jsou níže uvedeny podrobné výsledky pouze pro SVM klasifikátor s využitím příznakového vektoru složeného z MFCC a delta MFCC.

Na obrázku 4.2 je zobrazen histogram rozdělení skóre oprávněných a neoprávněných uživatelů s vyznačenými hodnotami rozhodovacího prahu. Z pohledu verifikace (autentizace) je snahou dosáhnout co nejnižší hodnoty FAR, v ideálním případě hodnoty nulové. Aby bylo dosaženo snížení hodnoty FAR je potřeba zvýšit hodnotu rozhodovacího prahu. Obrázek 4.3 ukazuje závislost hodnot FAR a FRR na velikosti rozhodovacího prahu. Nulové hodnoty FAR bylo dosaženo po nastavení rozhodovacího prahu na hodnotu 65%. Tabulky 4.2, 4.3, 4.4 zobrazují kontingenční tabulky pro různé hodnoty rozhodovacího prahu. Jak je vidět na obrázku 4.3, hodnota FAR klesá se zvyšujícím se rozhodovacím prahem, naopak hodnota FRR v tomto případě roste. Obrázek 4.4 zobrazuje DET křivku s vyznačeným bodem EER.

Tab. 4.2: Kontingenční tabulka - SVM klasifikátor (MFCC + delta MFCC), rozhodovací práh 50%.

Skutečný výstup	*Oprávněný uživatel*	359 51.1%	8 1.2%	97.8%
	Neoprávněný uživatel	1 0.1%	334 47.6%	99.7%
		99.7%	97.7%	**98.7%**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		

Tab. 4.3: Kontingenční tabulka - SVM klasifikátor (MFCC + delta MFCC), rozhodovací práh 60%.

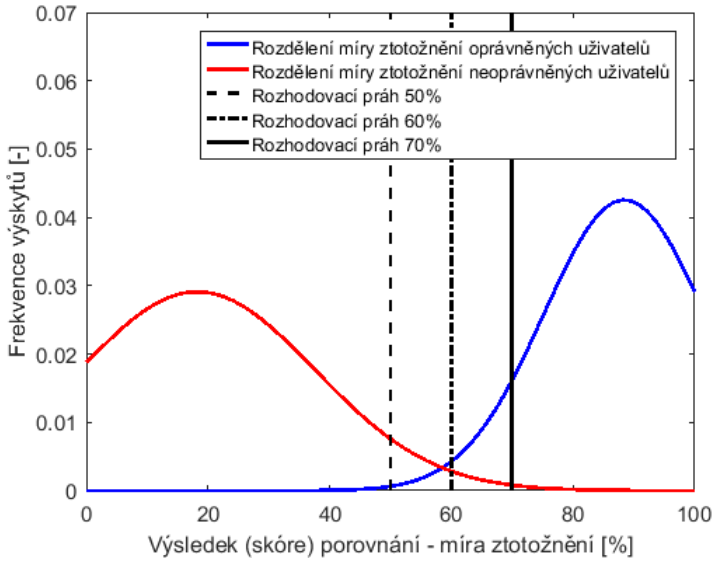
Skutečný výstup	*Oprávněný uživatel*	355 50.6%	3 0.4%	99.2%
	Neoprávněný uživatel	5 0.7%	339 48.3%	98.5%
		98.6%	99.1%	**98.9%**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		

4.4 Zhodnocení dosažených výsledků

Z výše uvedených výsledků vyplývá, že z pohledu klasifikátorů je vhodnější pro hlasovou autentizaci využít klasifikátor SVM, který potřebuje pro dosažení nejlepších výsledků pouze minimální množství parametrů, na rozdíl od klasifikátoru MLNN, který pro dosažení minimální chybovosti vyžaduje všechny testované parametry. Tato skutečnost je

Tab. 4.4: Kontingenční tabulka - SVM klasifikátor (MFCC + delta MFCC), rozhodovací práh 70%.

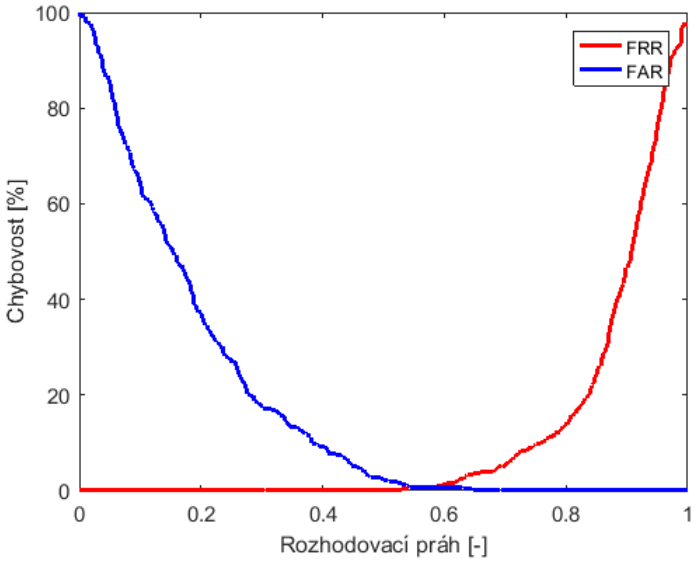
Skutečný výstup	*Oprávněný uživatel*	340 48.5%	0 0.0%	100.0%
	Neoprávněný uživatel	20 2.8%	342 48.7%	94.5%
		94.4%	100.0%	**97.2%**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		



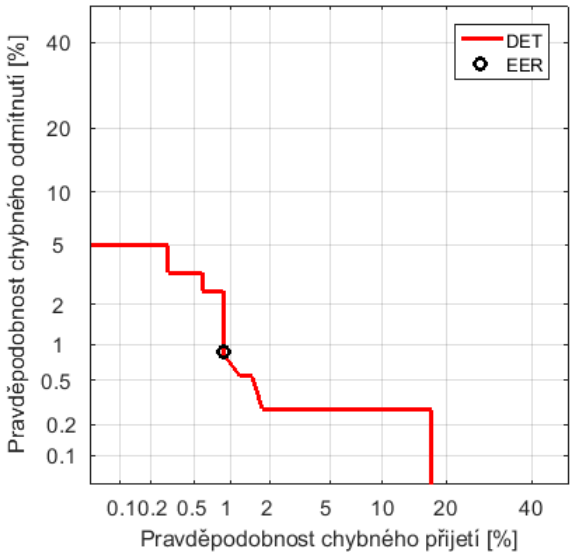
Obr. 4.2: Histogram rozdělení skóre oprávněných a neoprávněných uživatelů s vyznačenými hodnotami rozhodovacího prahu - hlasová autentizace.

dána vlastnostmi jednotlivých klasifikátorů. Při trénování klasifikátoru MLNN dochází k promíchání trénovacích dat tak, aby nedocházelo k přetrénování klasifikátoru. Promíchání trénovacích dat způsobí porušení časové sekvence v rámci jedné nahrávky a tudíž klasifikátor vyžaduje i parametry popisující časové změny mezi jednotlivými segmenty (delta-delta MFCC). Při trénování SVM klasifikátoru nedochází k promíchávání trénovacích dat a tudíž není potřeba při trénování využít parametry popisující časové změny mezi segmenty.

Výstupem této kapitoly je tedy textově závislý hlasový autentizační systém postavený na klasifikátoru SVM s využitím příznakového vektoru složeného z MFCC a delta MFCC parametrů. Takto navržený systém produkuje nejnižší počet chyb pro databázi Comtech. Systém je následně využit jako část vícenásobného biometrického autentizačního systému



Obr. 4.3: Závislost hodnot FAR a FRR na velikosti rozhodovacího prahu - hlasová autentizace.



Obr. 4.4: DET křivka - hlasová autentizace.

popsaného v kapitole 6.

5 NÁVRH AUTENTIZAČNÍHO SYSTÉMU ZALOŽENÉHO NA OVĚŘENÍ IDENTITY POMOCÍ GEOMETRIE OBLIČEJE

V této kapitole je popsán návrh unimodálního biometrického autentizačního systému založeného na ověření totožnosti uživatele pomocí geometrie tváře.

5.1 Princip návrhu biometrického systému pro autentizaci geometrií obličeje

Princip návrhu systému byl obdobný jako v případě systému navrženého v kapitole 4. Stejně jako v předchozím případě byl návrh proveden jak z pohledu nejvhodnějších parametrů, tak z pohledu vhodného klasifikátoru. Z pohledu parametrů byly porovnávány tyto parametry: LBP, HOG a jejich kombinace. Pro extrakci HOG parametrů byly stanoveny následující podmínky: velikost buňky 8x8 pixelů, počet buněk v bloku 4, překryv buněk mezi sousedními bloky 1, počet binů orientovaného histogramu 9. Velikost okolí v případě extrakce LBP parametrů byla nastavena na 3x3 pixelů. Tyto podmínky byly stanoveny na základě předchozího výzkumu [Tov09]. Popis použitých parametrů je uveden v podkapitolách 2.3.3 a 2.3.4. Tyto parametry byly vyhodnocovány pomocí klasifikátorů MLNN a SVM. Nastavení klasifikátorů MLNN a SVM bylo obdobné jako v kapitole 4. Volba stejných klasifikátorů jako v případě návrhu systému hlasové autentizace byla provedena na základě jejich vhodnosti pro řešení problému binární klasifikace a zároveň na snaze zjednodušit finální vícenásobný biometrický autentizační systém. Princip činnosti klasifikátorů je uveden v kapitole 2.2.5.

Prvním krokem autentizace pomocí geometrie obličeje je detekce tváře. Problematika detekce tváře tvoří velkou část v oblasti výzkumu rozpoznávání tváří. V případě autentizace ovšem dochází ke snížení požadavků na detekci (invariance vůči rotaci, osvětlení atd.), to je způsobeno tím, že uživatel chce být rozpoznán. Z tohoto důvodu není této problematice v práci věnován větší prostor. Při návrhu systému je využita detekční metoda Viola-Jones, která je podrobně popsána v podkapitole 2.3.3. Druhým krokem je předzpracování obrazu detekované tváře. Tento krok se skládá ze dvou částí: změna velikosti detekované tváře (nastavení fixní velikosti pro všechny obrazy - 120x120 pixelů) a převod obrazu do odstínů šedi (vyžadováno pro správnou extrakci parametrů). Další krokem je samotná extrakce parametrů, která je doplněna o min-max normalizaci. Extrahované parametry jsou následně klasifikovány příslušným klasifikátorem a na základě výsledků klasifikace je provedeno finální rozhodnutí a autentizaci. Jednotlivé přístupy jsou stejně jako v předchozí kapitole porovnávány pomocí FAR, FRR, ROC a DET. Výsledky byly získány pro databázi AR Face Database [54]. Návrh a implementace jednotlivých metod probíhala v prostředí Matlab. Zdrojové kódy včetně komentářů jsou přiloženy na SD kartě.

5.2 AR Face Database

Databáze slouží pro návrh a evaluaci algoritmů používaných v systémech rozpoznávání tváří. Databáze obsahuje přes 4000 barevných čelních pohledů na tváře od 126 respondentů (70 mužů, 56 žen). Nasnímané obrázky byly pořízeny během dvou sezení. Mezi jednotlivými sezeními byl časový rozestup 14 dní. Během každého sezení bylo pořízeno 13 čelních pohledů na tvář jednotlivých respondentů. Na respondenty nebyly kladeny žádné požadavky ohledně oblečení, účesů, brýlí atd.. V rámci snímání byly pořízeny obrazy v za předem definovaných podmínek:

- výraz tváře - neutrální výraz, úsměv, zlost, křik,
- změna osvětlení - osvětlení zleva, osvětlení zprava, osvětlení z obou stran,
- částečné zakrytí tváře - sluneční brýle, šátek přes obličej.

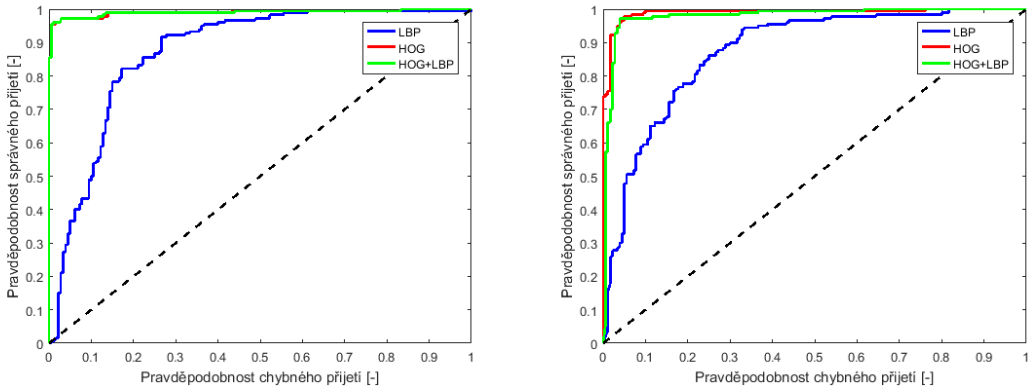
V rámci návrhu systému bylo vybráno 18 referenčních uživatelů (13 mužů, 5 žen) a dalších 10 uživatelů, kteří představují neoprávněné uživatele/podvodníky. Tento počet byl zvolen s ohledem na počet uživatelů řečové databáze Comtech a následný návrh vícenásobného biometrického systému. Ukázka vzorků pro jednoho uživatele je uvedena v příloze C.

5.3 Experimentální výsledky

Stejně jako v případě autentizace hlasem byly natrénovány klasifikátory SVM a MLNN pro každého z 18 referenčních uživatelů. Klasifikátory byly natrénovány tak, aby rozpoznávaly mezi dvěma třídami (třída referenční uživatel, třída podvodník). Každý z klasifikátorů byl trénován pomocí 32 digitálních obrazů detekovaných tváří. Pro trénování třídy referenčního uživatele bylo použito 16 obrazů tváře daného uživatele (mix obrazů z obou sezení). Jako testovací data pro třídu referenčního uživatele bylo použito 10 zbývajících obrazů. Třída podvodníků byla trénována jako univerzální model pozadí. Pro model pozadí každého referenčního uživatele bylo použito 16 obrazů tváří, které pocházejí od zbývajících referenčních uživatelů. Jako testovací data byla v tomto případě použita data od 10 vybraných uživatelů z databáze (tito uživatelé netvoří univerzální model pozadí). Tento proces trénování klasifikátorů se opakoval pro různé obrazové parametry a jejich kombinaci. Validační výsledky obou klasifikátorů pro všechny použité parametry jsou uvedeny v tabulce 5.1. Tabulka obsahuje hodnoty FAR, FRR pro rozhodovací práh 0.5 (50%) a hodnotu EER. ROC křivky pro jednotlivé klasifikátory jsou zobrazeny na obrázku 5.1.

Tab. 5.1: Výsledky dosažené pomocí SVM a MLNN klasifikátorů pro použité obrazové parametry (rozhodovací práh 50%)

Parametry	SVM			MLNN		
	FAR [%]	FRR [%]	EER [%]	FAR [%]	FRR [%]	EER [%]
LBP	35.0	6.6	17.8	26.6	13.9	21.7
HOG	2.7	3.3	**2.8**	6.0	2.2	**3.9**
LBP + HOG	2.7	3.3	**2.8**	7.2	2.7	3.9



(a) ROC křivka - SVM klasifikátor

(b) ROC křivka - MLNN klasifikátor

Obr. 5.1: ROC křivky - autentizace tváří

Jak je patrné z tabulky 5.1 nejlepších výsledků bylo dosaženo pro obrazové parametry HOG při použití SVM klasifikátoru. Hodnoty FAR a FRR byly pro tuto kombinaci klasifikátoru a parametrů 2.7% a 3.3% pro rozhodovací práh 50%. Systém tedy dosáhl přesnosti 96.9%. Hodnota EER dosáhla v tomto případě minimální hodnoty 2.8%. Stejných výsledků bylo dosaženo také pro kombinaci parametrů HOG a LBP. Vzhledem k tomu, že kombinací parametrů nebylo dosaženo zvýšení přesnosti a snížení chyb, lze tvrdit, že parametry LBP nepřináší žádnou novou informaci popisující rozdíly mezi tvářemi. Pro klasifikátor MLNN bylo dosaženo nejlepších výsledků pro parametry HOG, hodnota EER byla v tomto případě 3.9%. V následující části jsou uvedeny podrobné výsledky pro SVM klasifikátor využívající parametry HOG.

Obrázek 5.2 zobrazuje histogram rozdělení skóre s vyznačenými hodnotami prahu. Výsledky klasifikace a změny přesnosti pro jednotlivé verifikační prahy zobrazují tabulky 5.2, 5.3, 5.4. Z tabulek je patrné, že zvyšováním verifikačního prahu může být dosaženo snížení hodnoty FAR, ale na druhou stranu dojde ke zvýšení hodnoty FRR a v důsledku toho klesne i přesnost celého systému. Pro verifikační práh 70% dosáhne systém hodnoty FRR 67.8%, což je v praxi nepoužitelné, protože by byli dva ze tří oprávněných uživatelů odmítnuti. Z tohoto důvodu je nutné volit kompromis při nastavení verifikačního prahu. Na obrázku 5.3 je zobrazen průběh hodnot FAR a FRR v závislosti na verifikačním prahu. Obrázek 5.4 zobrazuje DET křivku se znázorněným bodem EER.

5.4 Zhodnocení dosažených výsledků

Na základě uvedených výsledků se jeví jako nejvhodnější přístup pro návrh systému klasifikátor SVM s využití HOG parametrů. Při využití těchto přístupů bylo dosaženo nejnižších hodnot FAR, FRR a EER. Tento přístup je také následně využit pro návrh vícenásobného biometrického systému v kapitole 6. Dosažení lepších výsledků pro SVM klasifikátor je

Tab. 5.2: Kontingenční tabulka - SVM klasifikátor (parametry HOG), rozhodovací práh 50%.

Skutečný výstup				
	Oprávněný uživatel	174 48.3%	5 1.4%	97.2%
	Neoprávněný uživatel	6 1.7%	175 48.6%	96.6%
		96.6%	97.2%	**96.9%**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		

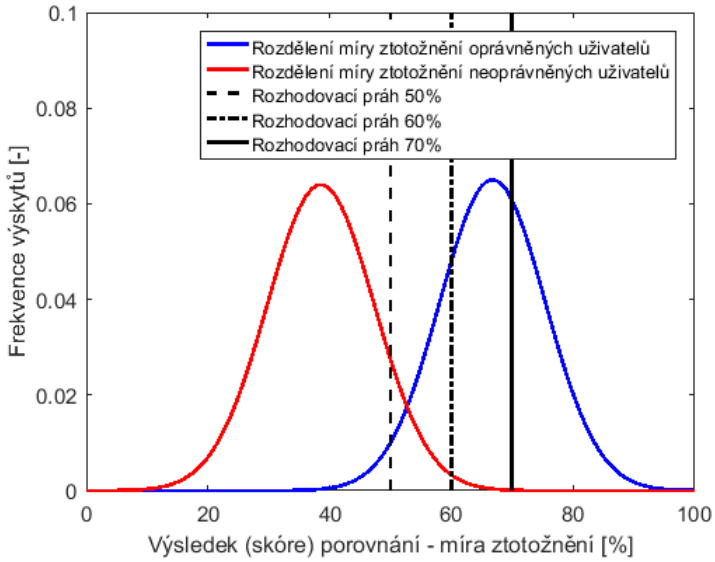
Tab. 5.3: Kontingenční tabulka - SVM klasifikátor (parametry HOG), rozhodovací práh 60%.

Skutečný výstup				
	Oprávněný uživatel	166 46.1%	1 0.3%	99.4%
	Neoprávněný uživatel	14 3.9%	179 49.7%	92.7%
		92.2%	99.4%	**95.8%**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		

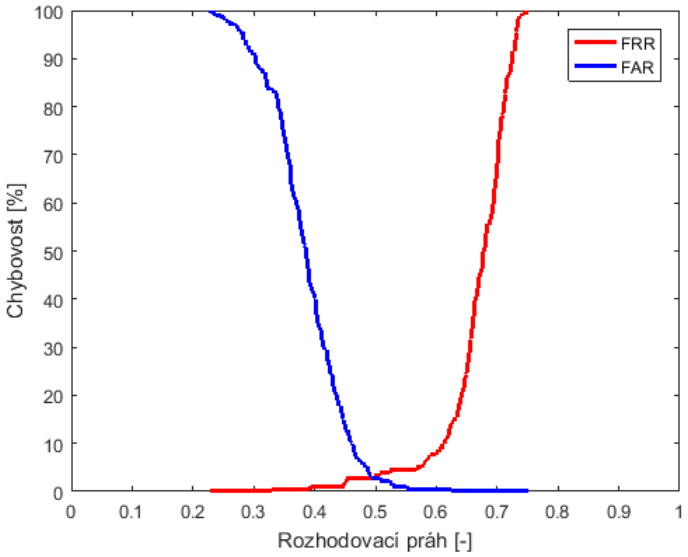
Tab. 5.4: Kontingenční tabulka - SVM klasifikátor (parametry HOG), rozhodovací práh 70%.

Skutečný výstup				
	Oprávněný uživatel	58 16.1%	0 0.0%	100.0%
	Neoprávněný uživatel	122 33.9%	180 50.0%	56.6%
		32.2%	100.0%	**66.1%**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		

dáno vlastnostmi klasifikátorů. V případě SVM jsou použity pro klasifikaci pouze podpůrné vektory. To znamená, že SVM klasifikátor nepotřebuje velkou trénovací sadu, pokud trénovací sada obsahuje vhodné podpůrné vektory. V případě MLNN klasifikátoru je důležitá velká trénovací sada, pomocí které jsou nastaveny váhy jednotlivých neuronů sítě. Z pohledu parametrů dosáhly nejlepších výsledků parametry HOG a to jak pro klasifikátor SVM, tak i pro klasifikátor MLNN. Dosažení lepších výsledků pro HOG parametry je dáno tím, že tyto parametry popisují soubor (histogram) změn (gradientů) v celém obraze, kdežto LPB parametry popisují pouze lokální vzory. Při detailní analýze výsledků bylo zjištěno, že nejčastější chyba klasifikace (SVM i MLNN klasifikátor) vzniká při klasi-

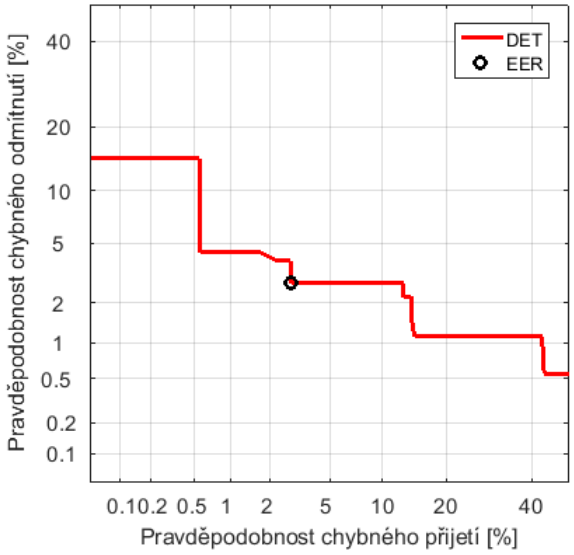


Obr. 5.2: Histogram rozdělení skóre oprávněných a neoprávněných uživatelů s vyznačenými hodnotami rozhodovacího prahu - autentizace tváří.



Obr. 5.3: Závislost hodnot FAR a FRR na velikosti rozhodovacího prahu - autentizace tváří.

fikaci obrázků, kde je osoba částečně zahalena do šátku. Dá se tedy předpokládat, že při odstranění těchto snímků z trénovací i testovací množiny bude dosaženo nižších hodnot FAR, FRR, EER a vyšší přesnosti. Odstranění těchto snímků je i v souladu z principem autentizace, kde uživatel chce být rozpoznán a tudíž nemá důvod zahalovat se šátkem.



Obr. 5.4: DET křivka - autentizace tváří.

6 NÁVRH KOMPLEXNÍHO VÍCENÁSOBNÉHO BIOMETRICKÉHO AUTENTIZAČNÍHO SYSTÉMU

Tato kapitola je věnována hlavní části dizertační práce a to návrhu vícenásobného biometrického autentizačního systému. Tento systém je založen na vhodné kombinaci hlasového autentizačního systému a systému využívajícího geometrii obličeje pro ověření totožnosti daného uživatele. Tyto systémy jsou postaveny na výsledcích a poznatcích z kapitol 4 a 5.

6.1 Princip návrhu vícenásobného biometrického autentizačního systému

Při návrhu vícenásobného autentizačního systému bylo vycházeno ze strategie vícenásobné biometrie (v jednom systému je využito více biometrických charakteristik). V rámci práce byly porovnány dva typy propojení (fúze) biometrických charakteristik: propojení na úrovni rozhodnutí o verifikaci a propojení na úrovni verifikační míry. V případě fúze na úrovni rozhodnutí o verifikaci byly ověřeny dva přístupy (AND a OR pravidla). Pro fúzi na úrovni verifikační míry bylo použito pravidlo o maximální pravděpodobnosti (Max rule) a pravidlo o sčítání pravděpodobností (Sum rule). Popis jednotlivých fúzních strategií včetně blokových schémat je uveden v podkapitole 2.4. V práci není uvažováno s klasifikačním přístupem pro fúzi na úrovni verifikační míry a to z důvodu nízkého počtu testovacích dat. Implementace jednotlivých typů fúzí včetně jejich vyhodnocení probíhala v prostředí Matlab. Zdrojové kódy jsou dostupné na SD kartě.

V experimentu byla použita data pocházející z databází Comtech a AR Face Database. Referenčním řečníkům (18) z databáze Comtech byly přiřazeny referenční obličeje pocházející z AR Face Database. Totéž platilo pro 10 uživatelů vydávajících se za podvodníky. K testování systému bylo použito 10 vzorků od každého referenčního uživatele a 10 vzorků od 10 různých podvodníků. Pod pojmem vzorek je v tomto případě myšleno spojení 1 nahrávka + 1 obraz tváře. Pro část systému založeného na autentizaci hlasem byl využit přístup poskytující nejlepší výsledky uvedené v kapitole 4. Jedná se o přístup využívající parametry MFCC + delta MFCC a klasifikátor SVM. Část systému sloužící pro ověření identity na základě obrazu tváře byla postavena na výsledcích z kapitoly 5. Tento přístup zahrnoval parametry HOG s využitím SVM klasifikátoru.

6.2 Propojení na úrovni rozhodnutí o verifikaci

V rámci propojeni na úrovni rozhodnutí o verifikaci byly ověřeny dvě strategie: AND a OR pravidlo. V případě použití AND pravidla je totožnost referenčního uživatele potvrzena pouze v případě, že obě části systému (autentizace hlasem, autentizace geometrií obličeje) rozhodly, že se jedná o referenčního uživatele. V ostatních případech je uživatel označen

jako podvodník. Toto pravidlo je vhodné v případě autentizace, kdy má být dosaženo nízké hodnoty FAR (přijetí podvodníků). Naopak v případě OR pravidla stačí, aby alespoň jedna část systému označila uživatele jako referenčního a ten je následně opravdu označen jako referenční. To znamená, že z pohledu autentizace bude dosaženo nízké hodnoty FRR (odmítnutí referenčních uživatelů). Vzhledem k tomu, že k finálnímu rozhodnutí dochází až po rozhodnutí dílčích podsystémů, kde je již nastaven verifikační práh jsou výsledky těchto fúzí popsány pouze pomocí kontingenční tabulky.

6.2.1 Experimentální výsledky pro fúzi pomocí AND pravidla

Aplikací AND pravidla na finální rozhodnutí dílčích podsystémů bylo dosaženo snížení počtu přijetí neoprávněných uživatelů (podvodníků) na nula. Přístup je povolen uživateli pouze v případě, že oba podsystémy rozhodly, že se jedná o oprávněného uživatele. Tento typ fúze je možné využít v případě, kdy chceme opravdu striktně odmítat přístup neoprávněných uživatelů (hodnota FAR je rovna nule). Výsledky pro AND pravidlo jsou uvedeny v kontingenční tabulce 6.1. Použitím této fúze je dosaženo celkové přesnosti systému 98.3%. Porovnání jednotlivých typů fúzí z pohledu hodnot FAR a FRR je uvedeno v tabulce 6.5.

Tab. 6.1: Kontingenční tabulka - fúze pomocí AND pravidla, rozhodovací práh podsystémů 50%.

Skutečný výstup	*Oprávněný uživatel*	174 48.3%	0 0.0%	100.0%
	Neoprávněný uživatel	6 1.7%	180 50.0%	96.7%
		96.6%	100.0%	**98.3%**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		

6.2.2 Experimentální výsledky pro fúzi pomocí OR pravidla

V případě použití OR pravidla je kladen důraz na co pokud možno největší komfort oprávněných uživatelů (stačí aby alespoň jeden z podsystémů označil uživatele jako oprávněného a uživatel je následně označen za oprávněného i vícenásobným systémem). Na druhé straně pro podvodníka je mnohem jednodušší systém prolomit. Aplikováním OR pravidla je sice dosaženo nízké hodnoty FRR, ale naproti tomu vzroste hodnota FAR. Výsledky jsou uvedeny v kontingenční tabulce 6.2. Celková přesnost systému v tomto případě je 96.7%.

Tab. 6.2: Kontingenční tabulka - fúze pomocí OR pravidla, rozhodovací práh podsystémů 50%.

Skutečný výstup	*Oprávněný uživatel*	180 50.0%	11 3.0%	94.2%
	Neoprávněný uživatel	0 0.0%	169 47.0%	100.0%
		100.0%	93.8%	**96.7%**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		

6.3 Propojení na úrovni verifikační míry

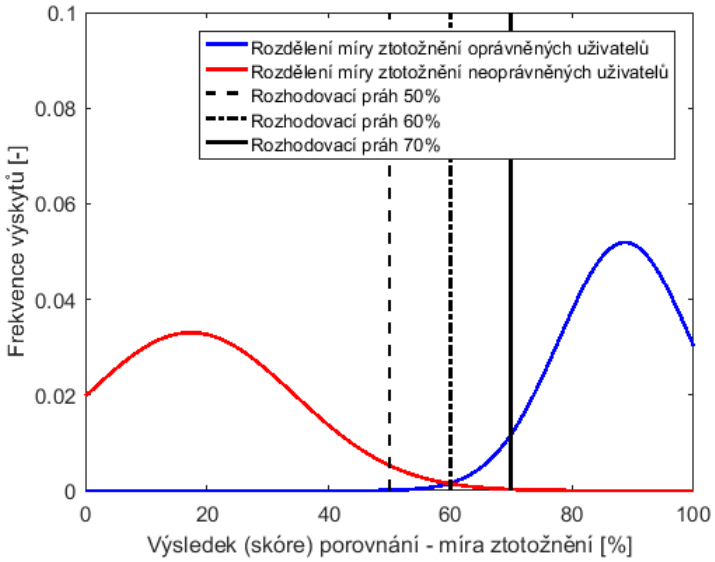
V případě fúze na úrovni verifikační míry, byl výzkum zaměřen na dvě kombinační metody: pravidlo o maximální pravděpodobnosti a pravidlo o sčítání pravděpodobností. V případě využití pravidla o maximální pravděpodobnosti je fúze dána pouze výběrem největší verifikační míry (skóre) z jednotlivých klasifikátorů pro danou třídu. Matematické vyjádření této fúze je popsáno rovnicí 2.34. Při použití pravidla o sčítání pravděpodobností je proveden součet pravděpodobností z jednotlivých klasifikátorů pro danou třídu. Zároveň může být každému klasifikátoru přiřazena určitá váha. Výpočet se provádí pomocí vzorce 2.33.

6.3.1 Experimentální výsledky pro fúzi pomocí pravidla o maximální pravděpodobnosti

Při použití pravidla o maximální pravděpodobnosti je rozhodnutí provedeno na základě porovnání maximální pravděpodobnosti jednoho z podsystémů a nastaveného rozhodovacího prahu. Výběrem maximální pravděpodobnosti je upřednostněn podsystém, který si je více "jistý"o přiřazení do dané třídy. Histogram rozdělení skóre po aplikaci fúze je zobrazen na obrázku 6.1. Z tabulky 6.3 je patrné, že se pomocí fúze podařilo snížit počet chybných rozhodnutí. S tím koresponduje i zvýšení přesnosti systému na 99.7%. ROC křivka systému je zobrazena na obrázku 6.2. Průběh hodnot FAR a FRR v závislosti na rozhodovacím prahu je uveden na obrázku 6.3. Hodnota EER byla v tomto případě 0.55% a je vyznačena v DET křivce na obrázku 6.4.

6.3.2 Experimentální výsledky pro fúzi pomocí pravidla o sčítání pravděpodobností

V případě fúze pomocí sčítání pravděpodobností je proveden součet posteriorních pravděpodobností jednotlivých podsystémů pro odpovídající třídy. Tento součet je následně porovnán s verifikačním prahem a na základě porovnání je provedeno finální rozhodnutí o autentizaci. Histogram rozdělení skóre pro tento typ fúze je zobrazen na obrázku 6.5. Z



Obr. 6.1: Histogram rozdělení skóre oprávněných a neoprávněných uživatelů s vyznačenými hodnotami rozhodovacího prahu - Max rule.

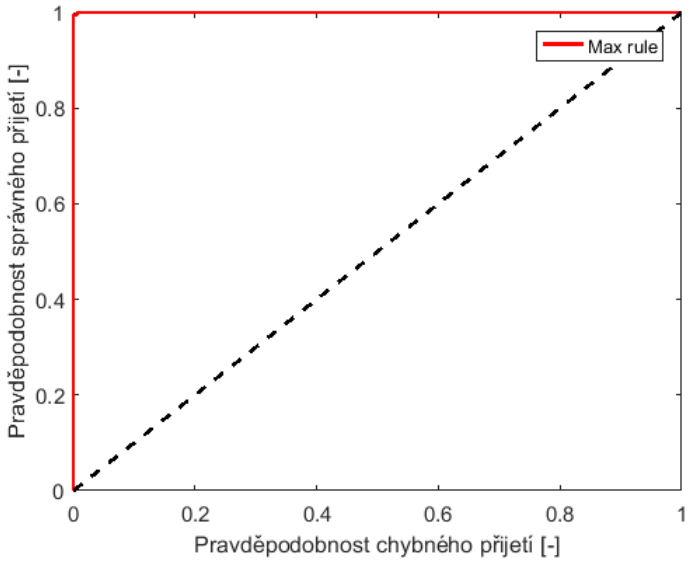
Tab. 6.3: Kontingenční tabulka - fúze pomocí Max rule, rozhodovací práh 50%.

Skutečný výstup	*Oprávněný uživatel*	180 50.0%	1 0.3%	99.4%
	Neoprávněný uživatel	0 0.0%	179 49.7%	100.0%
		100.0%	99.4%	**99.7%**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		

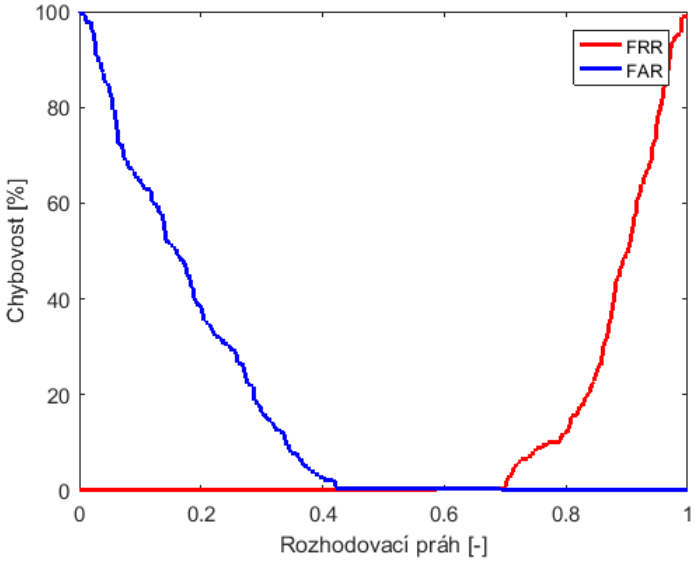
kontingenční tabulky 6.4 je patrné, že aplikací tohoto typu fúze je dosaženo zvýšení přesnosti oproti fúzi Max rule (systém stále produkuje 1 chybu). ROC křivka je zobrazena na obrázku 6.6. Zlepšení sytému v případě Sum rule je dosaženo z pohledu velikosti EER, která je v tomto případě 0.0%. To znamená, že pro vhodně nastavený rozhodovací práh (0.55 nebo 55%) je možné dosáhnout nulových hodnot FAR a FRR na použité testovací sadě. Tento fakt je patrný z obrázku 6.7. DET křivka není uvedena z důvodu nulové hodnoty EER.

6.4 Zhodnocení dosažených výsledků

Na základě výše uvedených výsledků lze konstatovat, že každý typ fúze přináší určité výhody z pohledu oblasti uplatnění systému. Porovnání jednotlivých typů fúzí je uvedeno



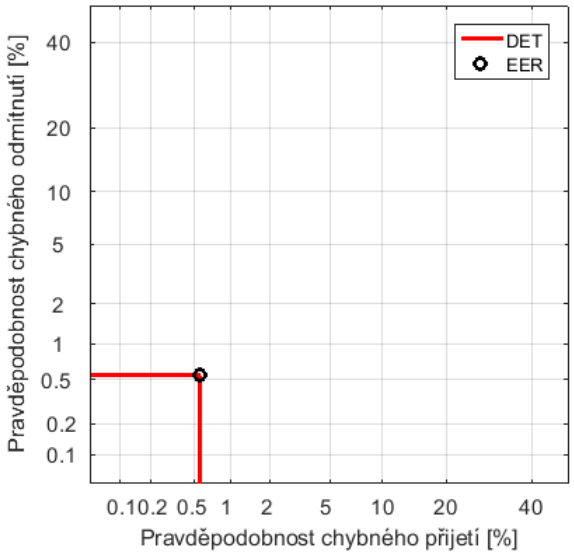
Obr. 6.2: ROC křivka - Max rule.



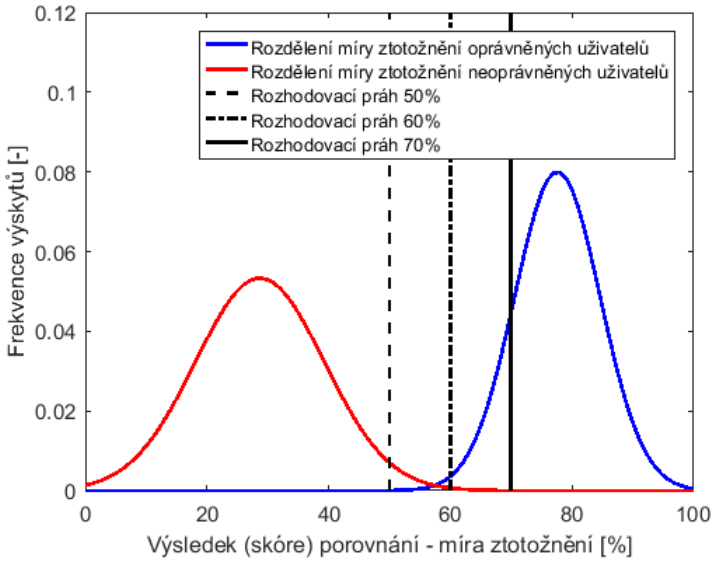
Obr. 6.3: Závislost hodnot FAR a FRR na velikosti rozhodovacího prahu - Max rule.

v tabulce 6.5.

Použitím AND pravidla je dosaženo nejnižší hodnota FAR (0.0%). Z praktického hlediska to znamená, že v rámci testovací sady není označen žádný podvodník jako referenční uživatel. Na druhé straně je v tomto případě dosaženo nejvyšší hodnoty FRR (3.3%) což znamená, že určitý počet oprávněných uživatelů je označen za podvodníky. Vícenásobný systém s využitím tohoto typu fúze je vhodné využít v případě, kdy je kladen obrovský



Obr. 6.4: DET křivka - Max rule.

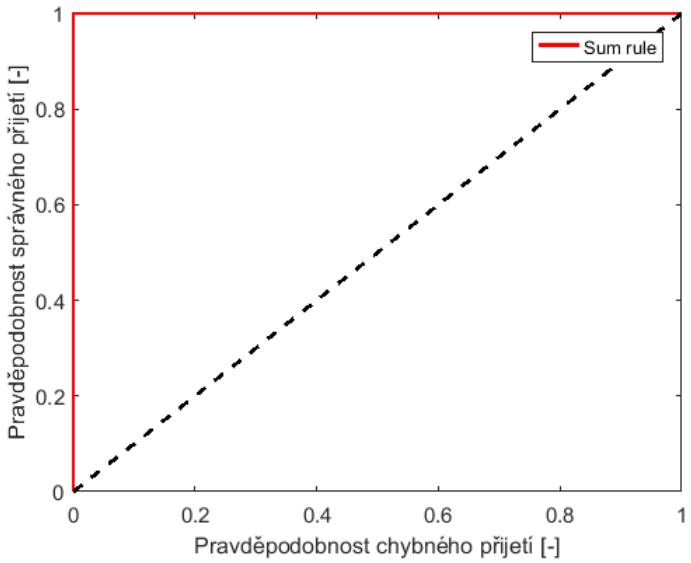


Obr. 6.5: Histogram rozdělení skóre oprávněných a neoprávněných uživatelů s vyznačenými hodnotami rozhodovacího prahu - Sum rule.

důraz na bezpečnost (pokud by byl přístup povolen podvodníkovi, mohlo by to mít fatální následky), například pro přístup k bankovnímu účtu. V případě použití OR pravidla je naopak dosaženo nejvyšší hodnoty FAR (6.2%). Hodnota FRR je v tomto případě 0.0%. Tento typ fúze je vhodné využít v případě, kdy je kladen důraz na komfort referenčních uživatelů. Referenční uživatelé nejsou obtěžování opětovnými pokusy o autentizaci,

Tab. 6.4: Kontingenční tabulka - fúze pomocí Sum rule, rozhodovací práh 50%.

Skutečný výstup	*Oprávněný uživatel*	180 50.0%	1 0.3%	99.4%
	Neoprávněný uživatel	0 0.0%	179 49.7%	100.0%
		100.0%	99.4%	**99.7%**
		Oprávněný uživatel	*Neoprávněný uživatel*	
		Požadovaný výstup		



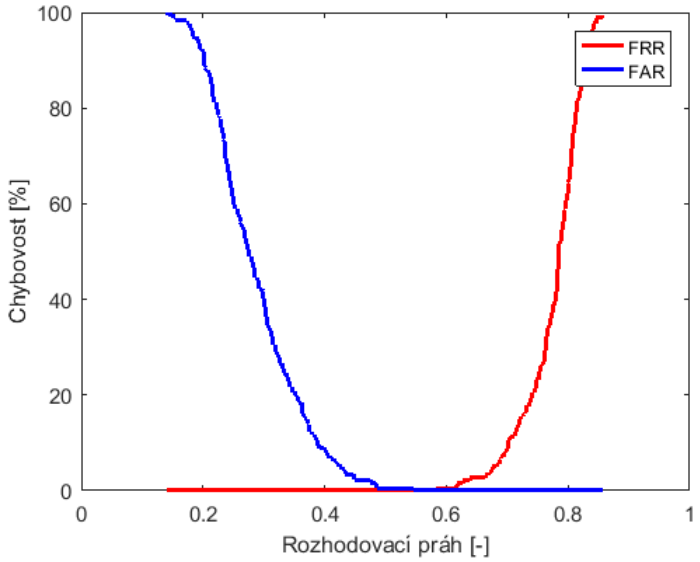
Obr. 6.6: ROC křivka - Sum rule.

Tab. 6.5: Tabulka hodnot FAR, FRR a EER pro jednotlivé typy fúzí pro rozhodovací práh 50%

Typ fúze	FAR [%]	FRR [%]	EER [%]
AND rule	0.0	3.3	-
OR rule	6.2	0.0	-
Max rule	0.6	0.0	0.55
Sum rule	0.6	0.0	0.00

na druhou stranu se může stát, že některý z podvodníků bude označen jako referenční uživatel.

Zatímco dvě předchozí metody přináší zlepšení systému pouze v závislosti na oblasti jeho použití (snížení hodnoty FAR nebo FRR), tak fúze typu pravidlo o maximální pravděpodobnosti a pravidlo o sčítání pravděpodobností přináší celkové zvýšení přesnosti bez



Obr. 6.7: Závislost hodnot FAR a FRR na velikosti rozhodovacího prahu - Sum rule.

ohledu na oblast použití. V obou případech propojení na úrovni rozhodnutí o verifikaci bylo dosaženo maximální přesnosti systému a to 99.7%. Tato přesnost odpovídá skutečnosti, že systém udělal jednu chybu při evaluaci testovací sady (360 vzorků). Přestože oba systémy dosáhly stejné přesnosti, tak z pohledu EER se jejich výsledky liší. Odlišnost hodnot EER je způsobena principem jednotlivých fúzí. V případě použití fúze Max rule systém dosáhl hodnoty EER 0.55%, kdežto pomocí Sum rule bylo dosaženo nulové hodnoty EER. To v praxi znamená, že při vhodně nastaveném rozhodovacím prahu systém založený na pravidle o sčítání pravděpodobností neprodukuje žádný typ chyby na příslušné testovací sadě. Hodnota verifikačního prahu pro dosažení nulové hodnoty EER je 0.55 (55%). Z tohoto pohledu lze označit vícenásobný systém založený na pravidle o sčítání pravděpodobností za nejlepší variantu.

7 EXPERIMENTÁLNÍ OVĚŘENÍ FUNKČNOSTI NAVRŽENÉHO VÍCENÁSOBNÉHO AUTENTIZAČNÍHO SYSTÉMU A POROVNÁNÍ PŘESNOSTI S AKTUÁLNĚ POUŽÍVANÝMI SYSTÉMY BIOMETRICKÉ AUTENTIZACE

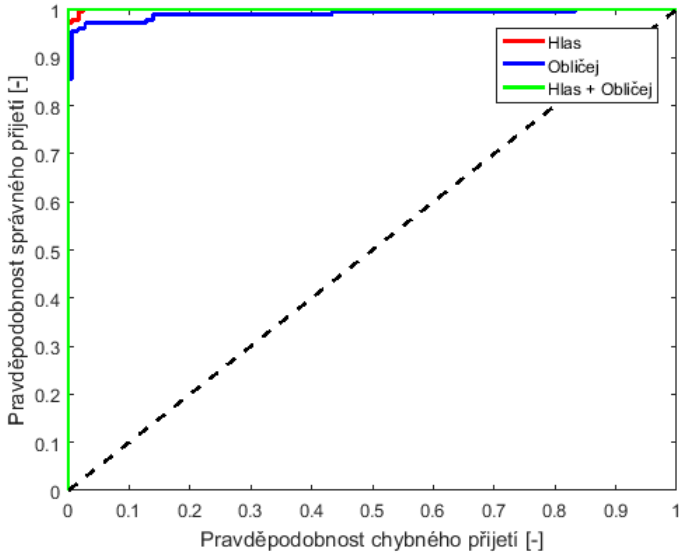
V první části této kapitoly je provedeno porovnání unimodálních biometrických autentizačních systému navržených v kapitolách 4 a 5 s vícenásobným biometrickým autentizačním systémem navrženým v kapitole 6. Druhá část kapitoly je zaměřena na porovnání navrženého vícenásobného systému s již existujícími obdobnými systémy.

7.1 Porovnání navržených biometrických systémů

Porovnání navržených unimodálních biometrických systémů s vícenásobným biometrickým systémem je provedeno jak z pohledu chybovosti a přesnosti, tak z pohledu bezpečnosti. V tabulce 7.1 jsou uvedeny výsledky hodnot FAR, FRR, EER a přesnost pro tři typy biometrických systémů. První je biometrický systém založený na ověření totožnosti pomocí hlasu. Tento systém využívá příznakový vektor skládající se z parametrů MFCC a delta MFCC, pro klasifikaci je použit SVM klasifikátor. Druhý biometrický systém využívá pro verifikaci informací o geometrii obličeje. Systém opět používá pro klasifikaci SVM klasifikátor a jako příznakový vektor parametry HOG. Vícenásobný biometrický systém je založen na fúzi předchozích dvou systémů. Fúze je provedena pomocí Sum rule. Z tabulky je patrné, že fúzí unimodálních systémů je dosaženo snížení FAR, FRR, EER a zvýšení celkové přesnosti systému. Navíc v případě vícenásobného systému jsme schopni nastavením vhodného rozhodovacího prahu dosáhnout nulové hodnoty EER na testovací sadě, což v oblasti biometrických systémů znamená ideální případ (systém správně rozpozná všechny referenční uživatele a správně odmítne všechny podvodníky). Použitím vícenásobného biometrického je také dosaženo vyšší bezpečnosti. U vícenásobných systémů je mnohem obtížnější provést podvržení identity referenčního uživatele. Na obrázku 7.1 jsou zobrazeny ROC křivky jednotlivých biometrických systémů.

Tab. 7.1: Tabulka hodnot FAR, FRR, EER a přesnosti pro jednotlivé typy biometrických systémů - rozhodovací práh 50%

Systém	FAR [%]	FRR [%]	EER [%]	Přesnost [%]
Hlas	3.3	0.0	1.7	98.3
Obličej	2.7	3.3	2.8	96.9
Hlas + Obličej	**0.6**	**0.0**	**0.0**	**99.7**



Obr. 7.1: ROC křivky - biometrické systémy.

7.2 Porovnání navrženého vícenásobného biometrického systému s obdobnými existujícími systémy

V této podkapitole je uveden přehled nejvýznamnějších publikací z oblasti audio-vizuální verifikace/identifikace. Vzhledem k tomu, že téměř v každé práci je použita jiná databáze nebo jiné podmínky experimentu, tak není uvedeno přímé srovnání číselných hodnot, které by bylo v tomto případě bezvýznamné.

První z významných publikací této problematiky je [55] z roku 1993, kde autoři využívají pro kombinaci podsystému textově závislé identifikace hlasu a podsystému identifikace obličeje pravidlo o sčítání pravděpodobností. V experimentu byly použity parametry MFCC a jako obrazové parametry byly použity vzdálenosti mezi markantními body na tváři. Pro klasifikaci byla v obou podsystémech použita ANN. Rozhodnutí o identifikaci bylo provedeno na základě porovnání rozhodovacího prahu a výsledného skóre po fúzi. Nastavením vhodného prahu autoři dosáhli hodnoty EER 1.5%. Těchto výsledků bylo dosaženo na vlastní databázi tvořené 10 uživateli.

V literatuře [56] autoři využívají pro kombinaci biometrických charakteristik pravidlo o násobení pravděpodobností. V experimentu použili vlastní databázi, která obsahovala vzorky od 33 osob. Podsystém založený na identifikaci hlasem využíval parametry MFCC + delta MFCC, které byly klasifikovány pomocí vektorové kvantizace. Obličejový podsystém pracuje s geometrickými informacemi o markatních bodech (pozice a šířka nosu, očí atd.). Tyto parametry byly klasifikovány pomocí bayesova klasifikátoru. Celková přesnost navrženého systému byla 95%.

Publikace [57] je rozšířením předchozí práce [56] se snahou zvýšení identifikační přes-

nosti. Autoři v tomto experimentu provedli rozdělení parametrů do více kategorií, kde každá kategorie byla klasifikována zvlášť (MFCC, delta MFCC, geometrické vlastnosti nosu, geometrické vlastnosti očí atd.) a poté provedli fúzi jednotlivých pravděpodobností pomocí pravidla o násobení pravděpodobností. Bylo tak dosaženo zvýšení přesnosti o 3% na celkových 98%.

Autoři [58] navrhli biometrický autentizační systém, který využívá tři biometrické charakteristiky (hlas, obličej, pohyb rtů). Rozhodnutí o verifikaci je provedeno na základě kombinace dvou ze tří podsystémů. V první fázi jsou vybrány dva ze tří podsystémů (klasifikátorů), které poskytují nejlepší výsledky. Následně je provedena kombinace skóre vybraných podsystémů a porovnání tohoto skóre s přednastaveným prahem. Využitím takového typu fúze dojde ke zvýšení robustnosti systému proti rušení.

V literatuře [22] je představen vícenásobný biometrický autentizační systém založený na verifikaci hlasem a verifikaci pomocí obličeje. Pro experiment byla použita databáze XM2VTS. Autoři pro podsystém verifikace hlasem využili parametry LPC, které klasifikovali pomocí skrytých markovových modelů. Pro reprezentaci obličeje byly použity deformační modely EGM. Fúze je v tomto případě prováděna na úrovni verifikační míry a je na ní pohlíženo jako na klasifikační problém. Z toho důvodu byly pro kombinaci jednotlivých podsystému použity klasifikátory SVM a bayesův. Použitím SVM klasifikátoru bylo dosaženo hodnoty EER 1.2%, v případě druhého klasifikátoru 0.6%.

Publikace [59] popisuje systém, který využívá pro kombinaci podsystémů AND pravidlo. Jedná se tedy o typ fúze na úrovní rozhodnutí o verifikaci. V rámci experimentu byla použita vlastní databáze, která byla tvořena 30 účastníky. Autoři použitím AND pravidla dokázali minimalizovat hodnotu FAR. Jako parametry popisující hlas autoři použili koeficienty vycházející z vlnkové transformace. Obrazové parametry představovaly vzhled očí. Oba typy parametrů byly klasifikovány pomocí vícevrstvé neuronové sítě.

Autoři v literatuře [29] využívají stejných metod fúze (klasifikátor SVM a bayesův klasifikátor) jako autoři v literatuře [22] s tím rozdílem, že v rámci jednotlivých podsystémů využívají jiné parametry a klasifikátory. Pro reprezentaci hlasu jsou použity parametry MFCC a jejich první a druhá derivace. Obličej je reprezentován parametry získanými pomocí metody hlavních komponent. V obou případě je pro klasifikaci využit klasifikátor GMM. Vzorky pro experiment byly získány z databáze VidTIMIT.

Literatura [60] popisuje biometrický identifikační systém, který je založen na fúzi pomocí pravidla o sčítání pravděpodobností. Hlasový podsystém využívá pro reprezentaci hlasu prozodických parametrů (základní frekvence, počet průchodů nulou, formanty atd.). Podsystém identifikace obličeje využívá parametrů popisujících geometrické vlastnosti markantních bodů. Klasifikace jednotlivých parametrů je provedena pomocí umělé neuronové sítě. Autoři použili v rámci experimentu vlastní databázi, která byla tvořena 20 respondenty. Na této databázi se podařilo dosáhnout identifikační přesnosti 97.5%.

V rámci publikace [63] byl popsán vícenásobný verifikační systém využívající pravidlo o sčítání pravděpodobností. Autoři v tomto experimentu použili databázi VidTIMIT.

Podsystém verifikace hlasem byl založen na použití LPC koeficientů, které byly klasifikovány GMM klasifikátorem. Podsystém verifikace obličeje využíval koeficienty získané pomocí 2DLDA, tyto koeficienty byly následně klasifikovány pomocí K-NN klasifikátoru.

Výzkum v práci [61] je zaměřen na porovnání různých typů fúzí a s tím spojených normalizačních technik. Data pro experiment pochází ze dvou databází XM2VTS (obrazová data) a TIMIT (hlasová data). Autoři dosáhli nejlepších výsledků s využitím pravidla o sčítání pravděpodobností a pravidla o násobení pravděpodobností.

Další varianty vícenásobných biometrických systémů jsou uvedeny v [62], [64] a [65].

Na základě rešerše existujících systémů lze tvrdit, že navržený vícenásobný biometrický systém v rámci dizertační práce je jedinečný z pohledu použitých metod (žádný z výše uvedených systémů není založen na stejných technikách a metodách jako navržený systém). Zároveň navržený systém dosahuje vysoké přesnosti a nízké chybovosti v porovnání s existujícími systémy, což ho předurčuje k možnosti reálného nasazení.

8 ZÁVĚR A PŘÍNOS PRÁCE

Biometrické autentizační systémy se v současné době stávají synonymem samotné autentizace. S rostoucím využívání biometrických metod rostou i požadavky na systémy, které tyto metody využívají. Pro většinu biometrických metod již není limitující pouze jejich přesnost, ale především bezpečnost. Existuje již celá řada přístupů, jak je možné napodobit různé biometrické charakteristiky. Z tohoto důvodu se hledají nové cesty, jak vhodně zamezit podvržení identity. Jednou z těchto cest je využití vícenásobné biometrie. Díky vícenásobné biometrii je mnohem těžší nebo dokonce nemožné provést podvržení identity a zároveň je dosaženo zvýšení přesnosti celého systému. Tato skutečnost byla hlavním podnětem pro vypracování této dizertační práce. Dizertační práce byla zaměřena na návrh vícenásobného biometrického autentizačního systému, který vhodně kombinuje dvě biometrické charakteristiky a to hlas a tvář. Tyto dvě biometrické charakteristiky byly zvoleny s ohledem na jejich vlastnosti (přijatelnost pro uživatele, jednoduché bezkontaktní snímání, přesnost, rychlost, spolehlivost). Samotný návrh vícenásobného systému byl rozdělen do tří částí. V první části byl proveden návrh podsystému zajišťujícího autentizaci pomocí řečového signálu. V druhé části byl navržen druhý autentizační podsystém založený na ověření identity pomocí geometrie obličeje. V poslední části byly ověřovány různé fúzní strategie navržených podsystémů za účelem nalezení nejvhodnějšího architektury vícenásobného biometrického autentizačního systému.

Návrh hlasového autentizačního systému probíhal jak z pohledu hledání vhodných parametrů, tak z pohledu nalezení dostatečně přesného klasifikátoru. Při návrhu byly porovnávány parametry MFCC, delta MFCC, delta-delta MFCC, LPC a jejich kombinace. Klasifikace parametrů probíhala pomocí SVM klasifikátoru a MLNN klasifikátoru. Volba parametrů a klasifikátorů byla provedena na základě rešerše a na základě konkrétních požadavků kladených na systém. Jednotlivé kombinace použitých parametrů a příslušného klasifikátoru byly hodnoceny pomocí hodnot FAR, FRR, EER a přesnosti. Ověřování výsledků bylo prováděno vzhledem k databázi Comtech, která byla v rámci dizertační práce navržena a vytvořena. Nejlepších výsledků bylo dosaženo pro příznakový vektor složený z parametrů MFCC a delta MFCC při použití SVM klasifikátoru. Pro nastavený výchozí rozhodovací práh 50% dosáhly hodnoty FAR 2.3% a FRR 0.27%. Tyto hodnoty odpovídají přesnosti systému 98.7%. Hodnota EER pro tento příznakový vektor byla 0.85%, což odpovídá nastavenému rozhodovacímu prahu na 55%. Zvýšením rozhodovacího prahu na 65% bylo dosaženo nulové hodnoty FAR, což je z pohledu autentizace ideální případ (žádný z podvodníků není označen jako referenční uživatel). Tento hlasový autentizační systém je reálně nasazen jako součást komplexního systému pro zabezpečenou komunikaci v rámci projektu TA ČR TF01000091, kde slouží jako doplňková ochrana. Princip nasazení systému do komplexního řešení je uveden v práci [Tov08]. Teoretickým přínosem této části dizertační práce je skutečnost, že klasifikátor, který dodržuje časovou sekvenci zpracovávaných segmentů vyžaduje pro dosažení maximální přesnosti pouze minimální

počet signifikantních parametrů, kdežto klasifikátor, který tuto sekvenci poruší potřebuje pro dosažení maximální přesnosti maximální počet parametrů. Dalším přínosem je vznik české databáze řečových vzorků, která může být využita, jak pro textově nezávislé rozpoznávání řečníka, tak pro textově závislé.

Návrh autentizačního systému založeného na ověření identity pomocí geometrie obličeje byl obdobný jako návrh hlasového autentizačního systému. Opět bylo úkolem nalézt vhodné parametry a klasifikátor produkující nejnižší počet chyb. Analyzovanými obrazovými příznaky byly LBP, HOG a jejich kombinace. Klasifikace těchto parametrů byla provedena pomocí SVM klasifikátoru a MLNN klasifikátoru. Volba stejných klasifikátorů jako v případě hlasové autentizace je založena na jejich vhodnosti pro řešení binární klasifikace a také na snaze o zjednodušení výsledného vícenásobného systému. Jednotlivé varianty realizace systému byly mezi sebou porovnávány opět pomocí hodnot FAR, FRR, EER a přesnosti. Obrazová data pro návrh systému pocházela z databáze AR Face Database. Výsledkem návrhu byl systém založený na HOG parametrech s využitím SVM klasifikátoru. Takto navržený systém dosáhl hodnot FAR 2.7% a FRR 3.3% pro rozhodovací práh 50%, což odpovídá celkové přesnosti systému 96.9%. Hodnota EER byla pro tento systém 2.8%. Stejně jako u předchozího systému lze dosáhnout nižší hodnoty FAR pomocí zvýšení rozhodovacího prahu. Experimentální výsledky návrhu autentizačního systému založeného na ověření identity pomocí geometrie obličeje byly publikovány v práci [Tov09]. Teoretickým přínosem je v tomto případě poznatek o volbě vhodného klasifikátoru s ohledem na velikost trénovací sady.

Výsledný vícenásobný biometrický autentizační systém je založen na vhodné kombinaci navrženého hlasového autentizačního systému a autentizačního systému využívajícího geometrii obličeje pro ověření identity. V dizertační práci byly porovnávány dva typy úrovní fúze (propojení na úrovni rozhodnutí o verifikaci a propojení na úrovni verifikační míry). Pro každou z úrovní byly analyzovány dvě strategie, které reprezentující danou úroveň. V rámci propojení na úrovni rozhodnutí o verifikaci byly experimentálně ověřeny strategie AND a OR. V případě propojení na úrovni verifikační míry byly analyzovány dvě strategie pravidlo o maximální pravděpodobnosti a pravidlo o sčítání pravděpodobností. Z pohledu velikosti hodnoty FAR bylo dosaženo nejlepšího výsledku použitím fúze pomocí AND pravidla, kde hodnota FAR dosáhla nulové hodnoty. To v praxi znamená, že systém neoznačí za oprávněného uživatele ani jednoho podvodníka. Na druhou stranu použitím AND pravidla dosáhneme nejvyšší hodnoty FRR 3.3% ze všech testovaných strategií. Z tohoto pohledu se jeví jako ideální řešení využití fúze založené na pravidle o sčítání pravděpodobností, kde bylo dosaženo hodnot FAR 0.6%, FRR 0.0% a přesnosti 99.7% pro rozhodovací práh 50%. Současně lze dosáhnout nulové hodnoty EER při nastavení rozhodovacího prahu na 55%. Z pohledu činnosti systému to znamená, že pro takto nastavený rozhodovací práh systém neprodukuje žádné chyby na příslušné testovací sadě. Při porovnání navrženého vícenásobného autentizačního systému s unimodálními systémy je patrné snížení chybovosti a zvýšení přesnosti v případě vícenásobného systému. Současně

je dosaženo vysokého stupně bezpečnosti, díky kterému je prakticky nemožné provést podvržení identity (podvodník by musel věrohodně napodobit hlas pronášející tajné přístupové heslo, který náleží referenčnímu uživatel a musel by současně dokonale napodobit obličej téže osoby). Z porovnání s existujícími obdobnými systémy je patrné, že navržený systém v rámci dizertační práce je jedinečný jak z pohledu použitých metod, tak z pohledu přesnosti a bezpečnosti. Dosažením nulové hodnoty EER je možné navržený systém porovnávat i s jinými typy vícenásobných biometrických systémů dosahujících vyšší přesnosti. Z tohoto pohledu se následně projeví výhody použitých biometrických charakteristik jako jsou přijatelnost pro uživatele, jednoduché bezkontaktní snímání a rychlost.

Navržený vícenásobný biometrický autentizační systém představuje unikátní přístup pro dosažení vysoce přesné a bezpečné autentizace, který může být využit téměř ve všech oblastech ověřování identity.

LITERATURA

[1] RAK, Roman, Vašek MATYÁŠ a Zdeněk ŘÍHA. *Biometrie a identita člověka: ve forenzních a komerčních aplikacích*. 1. vyd. Praha: Grada, 2008. ISBN 9788024723655.

[2] ALBESHER, Badr, Fatih KURUGOLLU, Ahmed BOURIDANE a Asim BAIG. Cascaded multimodal biometric recognition framework. *IET Biometrics* [online]. 2014, **3**(1), 16-28 [cit. 2016-03-07]. DOI: 10.1049/iet-bmt.2012.0043. ISSN 20474938. Dostupné z: http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2012.0043

[3] POH, Norman a Michael SCHUCKERS. Biometrics statistics: a foreword and introduction to the special issue. *IET Biometrics* [online]. 2015, **4**(4), 206-208 [cit. 2016-03-07]. DOI: 10.1049/iet-bmt.2015.0100. ISSN 20474938. Dostupné z: http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2015.0100

[4] NGO, David Chek Ling, Andrew Beng Jin TEOH a Jiankun HU. *Biometric security / edited by David Chek Ling Ngo, Andrew Beng Jin Teoh and Jiankun Hu*. New Castle upon Tyne, UK: Cambridge Scholars Publishing, 2015. ISBN 9781443871839.

[5] PATO, Joseph N a Lynette I MILLETT. *Biometric recognition: challenges and opportunities*. Washington, D.C.: National Academies Press, c2010.

[6] DAQROUQ, Khaled a Tarek A. TUTUNJI. Speaker identification using vowels features through a combined method of formants, wavelets, and neural network classifiers. *Applied Soft Computing* [online]. 2015, **27**, 231-239 [cit. 2016-03-07]. DOI: 10.1016/j.asoc.2014.11.016. ISSN 15684946. Dostupné z: http://linkinghub.elsevier.com/retrieve/pii/S1568494614005778

[7] BUSCH, Christoph, Herbert REININGER, Klaus KASPER, Stefan BILLEB a Christian RATHGEB. Biometric template protection for speaker recognition based on universal background models. *IET Biometrics* [online]. 2015, **4**(2), 116-126 [cit. 2016-03-07]. DOI: 10.1049/iet-bmt.2014.0031. ISSN 20474938. Dostupné z: http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2014.0031

[8] PSUTKA, Josef. *Mluvíme s počítačem česky*. Vyd. 1. Praha: Academia, 2006. Česká matice technická (Academia). ISBN 8020013091.

[9] XU, Yunfei, Yonghong YAN, Houjun HUANG, Hai YANG a Ruohua ZHOU. Voice biometrics using linear Gaussian model. *IET Biometrics* [online]. 2014, **3**(1), 9-15 [cit. 2016-03-07]. DOI: 10.1049/iet-bmt.2013.0027. ISSN 20474938. Dostupné z: http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2013.0027

[10] WANG, Jia-Ching, Chang-Hong LIN, En-Ting CHEN a Pao-Chi CHANG. Spectral-temporal receptive fields and MFCC balanced feature extraction for noisy speech recognition. In: *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific* [online]. IEEE, 2014, s. 1-4 [cit.

2016-03-07]. DOI: 10.1109/APSIPA.2014.7041624. ISBN 9786163618238. Dostupné z: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7041624

[11] RICHARDSON, Fred, Douglas REYNOLDS a Najim DEHAK. Deep Neural Network Approaches to Speaker and Language Recognition. *IEEE Signal Processing Letters* [online]. 2015, **22**(10), 1671-1675 [cit. 2016-03-07]. DOI: 10.1109/LSP.2015.2420092. ISSN 10709908. Dostupné z: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7080838

[12] MCLAREN, Mitchell, Yun LEI a Luciana FERRER. Advances in deep neural network approaches to speaker recognition. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* [online]. IEEE, 2015, s. 4814-4818 [cit. 2016-03-07]. DOI: 10.1109/ICASSP.2015.7178885. ISBN 9781467369978. Dostupné z: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7178885

[13] MCLAREN, Mitchell a Yun LEI. Improved speaker recognition using DCT coefficients as features. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* [online]. IEEE, 2015, s. 4430-4434 [cit. 2016-03-07]. DOI: 10.1109/ICASSP.2015.7178808. ISBN 9781467369978. Dostupné z: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7178808

[14] KAUR, Kirandeep a Neelu JAIN. Performance analysis of text-dependent speaker recognition system based on template model based classifiers. In: *2015 International Conference on Signal Processing, Computing and Control (ISPCC)* [online]. IEEE, 2015, s. 36-39 [cit. 2016-03-07]. DOI: 10.1109/ISPCC.2015.7374994. ISBN 9781479984367. Dostupné z: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7374994

[15] TAN, Hengliang, Zhengming MA a Bing YANG. Face recognition based on the fusion of global and local HOG features of face images. *IET Computer Vision* [online]. 2014, **8**(3), 224-234 [cit. 2016-03-07]. DOI: 10.1049/iet-cvi.2012.0302. ISSN 17519632. Dostupné z: http://digital-library.theiet.org/content/journals/10.1049/iet-cvi.2012.0302

[16] BOURLAI, Thirimachos, Ayman ABAZA, Arun ROSS a Mary Ann HARRISON. Design and evaluation of photometric image quality measures for effective face recognition. *IET Biometrics* [online]. 2014, **3**(4), 314-324 [cit. 2016-03-07]. DOI: 10.1049/iet-bmt.2014.0022. ISSN 20474938. Dostupné z: http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2014.0022

[17] SU, Ching-Yao a Jar-Ferr YANG. Histogram of gradient phases: a new local descriptor for face recognition. *IET Computer Vision* [online]. 2014, **8**(6), 556-567 [cit. 2016-03-07]. DOI: 10.1049/iet-cvi.2013.0208. ISSN 17519632. Dostupné z: http://digital-library.theiet.org/content/journals/10.1049/iet-cvi.2013.0208

[18] BOUGHRARA, Hayet, Chokri BEN AMAR, Mohamed CHTOUROU a Liming CHEN. Face recognition based on perceived facial images and multilayer perceptron neural network using constructive training algorithm. *IET Computer Vision* [online]. 2014, **8**(6), 729-739 [cit. 2016-03-07]. DOI: 10.1049/iet-cvi.2013.0294. ISSN 17519632. Dostupné z: http://digital-library.theiet.org/content/journals/10.1049/iet-cvi.2013.0294

[19] YONG XU, XIAOZHAO FANG, XUELONG LI, JIANG YANG, Jane YOU, HONG LIU a SHAOHUA TENG. Data Uncertainty in Face Recognition. *IEEE Transactions on Cybernetics* [online]. 2014, **44**(10), 1950-1961 [cit. 2016-03-07]. DOI: 10.1109/TCYB.2014.2300175. ISSN 21682267. Dostupné z: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6729058

[20] ROSS, Arun, Anil K. JAIN. Multimodal biometrics: An overview. *Signal Processing Conference, 2004 12th European* [online]. 2004, 1221-1224 [cit. 2016-09-22]. Dostupné z: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7080214&isnumber=7079654

[21] DOUGMAN, John. Combining multiple biometrics. Dostupné z: https://www.cl.cam.ac.uk/j̃gd1000/combine/

[22] BEN-YACOUB, Souheil, Yousri ABDELJAOUED a Eddy MAYORAZ. Fusion of face and speech data for person identity verification. *IEEE Transactions on Neural Networks* [online]. 1999, **10**(5), 1065-1074 [cit. 2016-03-07]. DOI: 10.1109/72.788647. ISSN 1045-9227. Dostupné z: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=788647&isnumber=17091

[23] ROSS, Arun, Anil K. JAIN. Information fusion in biometrics. *Pattern Recognition Letters*. 2003, **24**(14), 2115-2125. Dostupné z: http://www.sciencedirect.com/science/article/pii/S0167865503000795

[24] CHAW, Chia, Nasser SHERKAT a Lars NOLLE. Confidence Partition and Hybrid Fusion in Multimodal Biometric Verification System. *Springer Berlin Heidelberg*. 2009, 212-219. DOI: 10.1007/978-3-642-04391-8_28. ISBN 978-3-642-04391-8. Dostupné z: http://dx.doi.org/10.1007/978-3-642-04391-8_28

[25] SCHEIDAT, Tobias, Michael BIERMANN, Jana DITTMANN, Claus VIELHAUSER a Karl KÜMMEL. Multi-biometric Fusion for Driver Authentication on the Example of Speech and Face. *Springer Berlin Heidelberg*. 2009, 220-227. ISBN 978-3-642-04390-1. Dostupné z: http://dl.acm.org/citation.cfm?id=1812740.1812778

[26] KITTLER, Josef. Combining classifiers: A theoretical framework. *Pattern Analysis and Applications*. 1998, **1**(1), 18-27. DOI: 10.1007/BF01238023. ISSN 1433-755X. Dostupné z: http://dx.doi.org/10.1007/BF01238023

[27] KITTLER, Josef, Mohamed Hatef, Robert P.W. DUIN a Jiri MATAS. On Combining Classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.*. 1998, **20**(3), 226-239. DOI: 10.1109/34.667881. ISSN 0162-8828. Dostupné z: http://dx.doi.org/10.1109/34.667881

[28] JAIN, Anil, Karthik NANDAKUMAR a Arun ROSS. Score normalization in multimodal biometric systems. *Pattern Recognition*. 2005, **38**(12), 2270-2285. DOI: 10.1016/j.patcog.2005.01.012. ISSN 0031-3203. Dostupné z: http://dx.doi.org/10.1016/j.patcog.2005.01.012

[29] SANDERSON, Conrad a Kuldip K. PALIWAL. Identity verification using speech and face information. *Digital Signal Processing*. 2004, **14**(5), 449-480. DOI: 10.1016/j.dsp.2004.05.001. Dostupné z: http://dx.doi.org/10.1016/j.dsp.2004.05.001

[30] SASWATI, Debnath, B. SONI, U. BARUAH a D. K. SAH. Text-dependent speaker verification system: A review. *2015 IEEE 9th International Conference on*. 2015, 1-7. DOI: 10.1109/ISCO.2015.7282386. ISBN 978-1-4799-6480-2. Dostupné z: http://ieeexplore.ieee.org/document/7282386/

[31] MARKOWITZ, Judith A. Voice Biometrics. *Communications of the ACM*. 2000, **43**(9), 66-73. DOI: 10.1145/348941.348995. ISSN 00010782. Dostupné z: http://portal.acm.org/citation.cfm?doid=348941.348995

[32] LARCHER, Anthony, Kong Aik LEE, Bin MA a Haizhou LI. Text-dependent speaker verification: Classifiers, databases and RSR2015. *Speech Communication*. 2014, **60**, 56-77. DOI: 10.1016/j.specom.2014.03.001. ISSN 0167-6393. Dostupné z: http://www.sciencedirect.com/science/article/pii/S0167639314000156

[33] KINNUNEN, Tomi a Haizhou LI. An overview of text-independent speaker recognition: From features to supervectors. *Speech Communication*. 2010, **52**(1), 12-40. DOI: 10.1016/j.specom.2009.08.009. ISSN 0167-6393. Dostupné z: http://www.sciencedirect.com/science/article/pii/S0167639309001289

[34] REYNOLDS, Douglas A., Thomas F. QUANTIERI a Robert B. BUNN. Speaker Verification Using Adapted Gaussian Mixture Models. *Digital Signal Processing*. 2000, **10**(1), 19-41. DOI: 10.1006/dspr.1999.0361. ISSN 1051-2004. Dostupné z: http://www.sciencedirect.com/science/article/pii/S1051200499903615

[35] S. TOLBA, Ahmad, A. H. EL-BAZ a A. A. EL-HARBY. Face recognition: A literature review. *International Journal of Signal Processing*. 2006, **2**(2), 88-103. Dostupné z: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.179.2182&rep=rep1&type=pdf

[36] ZHAO, W., R. CHELLAPPA, P. J. PHILLIPS a A. ROSENFELD. Face recognition. *ACM Computing Surveys*. 2003, **35**(4), 399-458. DOI: 10.1145/954339.954342. ISSN 03600300. Dostupné z: http://portal.acm.org/citation.cfm?doid=954339.954342

[37] YANG, Ming-Hsuan, David J. KRIEGMAN a Narendra AHUJA. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002, **24**(1), 34-58. DOI: 10.1109/34.982883. Dostupné z: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=982883&isnumber=21178

[38] WANG, Yi-Qing. An analysis of the Viola-Jones face detection Algorithm. *Image Processing On Line*. 2014, **4**, 128-148. DOI: 10.5201/ipol.2014.104. Dostupné z: https://doi.org/10.5201/ipol.2014.104

[39] LI, Haoxiang, Zhe LIN, Xiaohui SHEN, Jonathan BRANDT a Gang HUA. A convolutional neural network cascade for face detection. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVRP)*. 2015, 5325-5334. DOI: 10.1109/CVPR.2015.7299170. ISBN 978-1-4673-6964-0. Dostupné z: http://ieeexplore.ieee.org/document/7299170/

[40] SUN, Yi, Xiaogang WANG a Xiaoou TANG. Deep Convolutional Network Cascade for Facial Point Detection. *2013 IEEE Conference on Computer Vision and Pattern Recognition*. 2013, 3476-3483. DOI: 10.1109/CVPR.2013.446. ISBN 978-0-7695-4989-7. Dostupné z: http://ieeexplore.ieee.org/document/6619290/

[41] VIOLA, Paul, Michael J. JONES a Xiaoou TANG. Robust Real-Time Face Detection. *2013 IEEE Conference on Computer Vision and Pattern Recognition*. 2013, 3476-3483. DOI: 10.1023/B:VISI.0000013087.49260.fb. ISBN 978-0-7695-4989-7. Dostupné z: http://link.springer.com/10.1023/B:VISI.0000013087.49260.fb

[42] DALAL, Navneet a Bill TRIGGS. Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. 2005, 886-893. DOI: 10.1109/CVPR.2005.177. ISBN 0-7695-2372-2. Dostupné z: http://dx.doi.org/10.1109/CVPR.2005.177

[43] LI, Wenhui, Yifeng LIN a Bo FU. Fast Human Detection Using a Cascade of United Hogs. *International Conference in Swarm Intelligence*. 2011, 327-332. DOI: 10.1007/978-3-642-21524-7_39. ISBN 978-3-642-21523-0. Dostupné z: http://link.springer.com/10.1007/978-3-642-21524-7_39

[44] VIOLA, Paul a Michael JONES. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. 2001, I-511-I-518. DOI: 10.1109/CVPR.2001.990517. ISBN 10.1109/CVPR.2001.990517. Dostupné z: http://ieeexplore.ieee.org/document/990517/

[45] JIA, Hui-Xing a Yu-Jin ZHANG. Fast Human Detection by Boosting Histograms of Oriented Gradients. *Fourth International Conference on Image and Graphics (ICIG 2007)*. 2007, 683-688. DOI: 10.1109/ICIG.2007.53. ISBN 978-0-7695-2929-5. Dostupné z: http://ieeexplore.ieee.org/document/4297169/

[46] ZHU, Qiang, Mei-Chen YEH, Kwang-Ting CHENG a Shai AVIDAN. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. CVPR 2006*. 2006, 1491-1498. DOI: 10.1109/CVPR.2006.119. ISBN 0-7695-2597-0. Dostupné z: http://dx.doi.org/10.1109/CVPR.2006.119

[47] SUN, Yi, Xiaogang WANG a Xiaoou TANG. Hybrid Deep Learning for Face Verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2016, **38**(10), 1997-2009. DOI: 10.1109/TPAMI.2015.2505293. ISSN 0162-8828. Dostupné z: http://ieeexplore.ieee.org/document/7346495/

[48] HU, Guosheng, Yongxin YANG, Dong YI, Josef KITTLER, William CHRISTMAS, Stan Z. LI a Timothy HOSPEDALES. When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition. *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*. 2015, 384-392. DOI: 10.1109/ICCVW.2015.58. ISBN 978-1-4673-9711-7. Dostupné z: http://ieeexplore.ieee.org/document/7406407/

[49] KRIZHEVSKY, Alex, Ilya SUTSKEVER a Geoffrey E. HINTON. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25*. 2012, 1097-1105. Dostupné z: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[50] CAIFENG, Shan. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*. 2012, **33**(4), 431-437. DOI: http://dx.doi.org/10.1016/j.patrec.2011.05.016. ISSN 0167-8655. Dostupné z: http://www.sciencedirect.com/science/article/pii/S0167865511001607

[51] HEISELE, Bernd, Purdy HO, Jane WU a Tomaso POGGIO. Face recognition: component-based versus global approaches. *Computer Vision and Image Understanding*. 2003, **91**(1-2)91, 6-21. DOI: 10.1016/S1077-3142(03)00073-0. ISSN 10773142. Dostupné z: http://linkinghub.elsevier.com/retrieve/pii/S1077314203000730

[52] AGADA, Ruth a Jie YAN. Edge based mean LBP for valence facial expression detection. *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. 2015, 1-7. DOI: 10.1109/ICECCT.2015.7226014. ISBN 978-1-4799-6084-2. Dostupné z: http://ieeexplore.ieee.org/document/7226014/

[53] BHANDARI, Smriti H. a Amruta G. YADRAVE. Local binary pattern approach for rotation invariant texture classification. *2015 International Conference on Circuits, Power and Computing Technologies*. 2015, 1-4. DOI: 10.1109/ICCPCT.2015.7159346. ISBN 978-1-4799-7075-9. Dostupné z: http://ieeexplore.ieee.org/document/7159346/

[54] MARTINEZ, A. a R. BENAVENTE. The AR Face Database. *CVC Technical Report #24*. 1998.

[55] CHIBELUSHI, Claude C., Farzin DERAVI a John MASON. Voice and facial image integration for speaker recognition. *IEEE International Symposium and Multimedia Technologies and Future Applications*. 1993

[56] BRUNELLI, Roberto, Daniele FALAVIGNA, Tomaso POGGIO a Luigi STRINGA. Automatic person recognition by acoustic and geometric features. *Machine Vision and Applications*. 1995, **8**(5) 317-325. DOI: 10.1007/BF01211493. ISSN 1432-1769. Dostupné z: https://doi.org/10.1007/BF01211493

[57] BRUNELLI, Roberto a Daniele FALAVIGNA. Person Identification Using Multiple Cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1995, **17**(10), 955-965. DOI: 10.1109/34.464560. ISSN 0162-8828. Dostupné z: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=464560

[58] DIECKMANN, Ulf, Peter PLANKENSTEINER a Thomas WAGNER. SESAM: A biometric person identification system using sensor fusion. *Pattern Recognition Letters*. 1997, **18**(9), 827-833. DOI: 10.1007/BFb0016009. ISSN 0167-8655. Dostupné z: https://doi.org/10.1007/BFb0016009

[59] POH, Norman a Jerzy KORCZAK. Hybrid Biometric Person Authentication Using Face and Voice Features. *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication*. 2001, 348-353. DOI: 10.1007/3-540-45344-X_51. ISBN 978-3-540-45344-4. Dostupné z: https://link.springer.com/content/pdf/10.1007/3-540-45344-X_51.pdf

[60] KALA, Rahul, Harsh VAZIRANI, Anupam SHUKLA, Ritu TIWARI. Fusion of Speech and Face by Enhanced Modular Neural Network. *Proceedings of the Fourth International Conference on Information Systems, Technology and Management*. 2010, 363-372. DOI: 10.1007/978-3-642-12035-0_37. ISBN 978-3-642-12035-0. Dostupné z: https://link.springer.com/chapter/10.1007/978-3-642-12035-0_37

[61] FOUDA, Yasser M. Fusion of Face and Voice: An improvement. *International Journal of Computer Science and Network Security*. 2012, **12**(4), 37-43. Dostupné z: http://paper.ijcsns.org/07_book/201204/20120406.pdf

[62] GHAYOUMI, Mehdi. A review of multimodal biometric systems: Fusion methods and their applications. *2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS)*. 2015, 131-136. DOI: 10.1109/ICIS.2015.7166582. Dostupné z: http://ieeexplore.ieee.org/document/7166582/

[63] RAGHAVENDRA, Ramachandra, Ashok RAO a Hemantha G. KUMAR. Multimodal person verification system using face and speech. *Proceedings of the International Conference and Exhibition on Biometrics Technology*. 2010, 181-187. DOI: 10.1016/j.procs.2010.11.023. ISSN 1877-0509. Dostupné z: http://www.sciencedirect.com/science/article/pii/S1877050910003534

[64] SOLTANE, Mohamed a Mimen BAKHTI. Multi-modal biometric authentications: concept issues and applications strategies. *International Journal of Advanced Science and Technology*. 2012, **48**. Dostupné z: https://pdfs.semanticscholar.org/8e60/8ebd80c59c4a5231aeba5bcb9d60c82e0b7e.pdf

[65] ALEKSIC, Petar S. a Aggelos K. KATSAGGELOS. Audio-Visual Biometrics. *Proceedings of the IEEE*. 2006, **94**(11), 2025-2044. DOI: 10.1109/JPROC.2006.886017. ISSN 0018-9219. Dostupné z: http://ieeexplore.ieee.org/abstract/document/4052464/

CITOVANÉ PŘÍSPĚVKY AUTORA V PRÁCI

[Tov01] TOVAREK, Jaromir, Pavol PARTILA, Jan ROZHON, Miroslav VOZNAK, Jan SKAPA, Dominik UHRIN a Zdenka CHMELIKOVA. Optimization of multilayer neural network parameters for speaker recognition. *Proceedings of SPIE - The International Society for Optical Engineering*. 2016. DOI: 10.1117/12.2223545. Dostupné z: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9850/1/Optimization-of-multilayer-neural-network-parameters-for-speaker-recognition/10.1117/12.2223545.full. Konferenční článek (SJR=0.228)

[Tov02] TOVAREK, Jaromir, Pavol PARTILA, Miroslav VOZNAK, Martin MIKULEC and Miralem MEHIC. Detection of cardiac activity changes from human speech. *Proceedings of SPIE - The International Society for Optical Engineering*. 2015. DOI: 10.1117/12.2177282. Dostupné z: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9496/94960V/Detection-of-cardiac-activity-changes-from-human-speech/10.1117/12.2177282.full. Konferenční článek (SJR=0.23)

[Tov03] PARTILA, Pavol, Jaromir TOVAREK a Miroslav VOZNAK. Self-organizing map classifier for stressed speech recognition. *Proceedings of SPIE - The International Society for Optical Engineering*. 2016. DOI: 10.1117/12.2224253. Dostupné z: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9850/1/Self-organizing-map-classifier-for-stressed-speech-recognition/10.1117/12.2224253.full. Konferenční článek (SJR=0.228)

[Tov04] PARTILA, Pavol, Miroslav VOZNAK a Jaromir TOVAREK. Pattern recognition methods and features slection for speech emotion recognition system. *The Scientific World Journal*. 2015. DOI: 10.1155/2015/573068. ISSN 2356-6140. Dostupné z: http://www.hindawi.com/journals/tswj/2015/573068/. Článek v časopise (SJR=0.32)

[Tov05] PARTILA, Pavol, Jaromir TOVAREK, Jaroslav FRNDA, Miroslav VOZNAK, Marek PENHAKER a Tomáš PETEREK. Emotional Impact on Neurological Characteristics and Human Speech. *Advances in Intelligent Systems and Computing*. DOI: 10.1007/978-3-319-07773-4_52. Dostupné z: http://link.springer.com/10.1007/978-3-319-07773-4_52. Konferenční článek (SJR=0.149)

[Tov06] PARTILA, Pavol, Miroslav VOZNAK, Tomas PETEREK, Marek PENHAKER, Vilem NOVAK, Jaromir TOVAREK, Miralem MEHIC a Lukas VOJTECH. Impact of human emotions on physiological characteristics. *Proceedings of SPIE - The International Society for Optical Engineering*. 2014. DOI: 10.1117/12.2050679. Dostupné z: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9118/1/Impact-of-human-emotions-on-physiological-characteristics/10.1117/12.2050679.full. Konferenční článek (SJR=0.237)

[Tov07] PARTILA, Pavol, Jaromir TOVAREK, Miroslav VOZNAK a Jakub SAFARIK. Classification methods accuracy for speech emotion recognition system. *Advances in Intelligent Systems and Computing*. 2014. DOI: 10.1007/978-3-319-07401-6_44. Dostupné z: http://link.springer.com/10.1007/978-3-319-07401-6_44. Konferenční článek (SJR=0.149)

[Tov08] TOVAREK, Jaromir a Pavol PARTILA. Speaker identification for the improvement of the security communication between law enforcement units. *Proceedings of SPIE - The International Society for Optical Engineering*. 2017. DOI: 10.1117/12.2261796. Dostupné z: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10200/1/Speaker-identification-for-the-improvement-of-the-security-communication-between/10.1117/12.2261796.full. Konferenční článek (SJR=0.228)

[Tov09] TOVAREK, Jaromir, Miroslav VOZNAK, Jan ROZHON, Filip REZAC, Jakub SAFARIK a Pavol PARTILA. Different approaches for face authentication as part of a multimodal biometrics system. *Advances in Electrical and Electronic Engineering*. 2017. DOI: 10.15598/aeee.v16i1.2547. Článek v časopise (přijat k vydání). (SJR=0.247)

PUBLIKAČNÍ ČINNOST AUTORA

K doložení svých vědecko–výzkumných aktivit přikládám i aktuální stav záznamů v relevantních vědeckých databázích ke dni odevzdání tohoto dokumentu.

- Publikace v bibliografické databázi ISI - Web of Knowledge: **8**
- Publikace v bibliografické databázi SCOPUS: **12**
- h-index podle ISI - Web of Knowledge: **1** (2 citace/2 bez autocitací)
- h-index podle SCOPUS: **2** (11 citací/7 bez autocitací)

Celkový počet indexovaných výstupů v Rejstříku výsledků vědy a výzkumu RIV, viz https://www.rvvi.cz : 14

- Publikační výstupy: **11**
- Aplikované výsledky : **3**

Participace na řešení projektů během studia

- **TA ČR DELTA TF01000091** – Aplikovaný výzkum: Bezpečnost mobilních zařízení a komunikace (2015-2017).
- **SP2017/174** – Specifický výzkum: Sítě a jejich bezpečnost, modelování, simulace, vytěžování znalostí a komunikační technologie pro chytrá města.
- **SP2016/170** – Specifický výzkum: Vytěžování informací z komunikačních sítí, jejich modelování a simulace II.
- **SP2015/82** – Specifický výzkum: Vytěžování informací z komunikačních sítí, jejich modelování a simulace I.

A STRUKTURA DATABÁZE COMTECH

Všechny nahrávky databáze jsou zaznamenány ve WAV formátu se vzorkovací frekvencí 8kHZ a 32 bitovým rozlišením. Celková velikost databáze je 250MB (210MB - textově nezávislá část, 40MB - textově závislá část). Struktura databáze je znázorněna tabulkou A.1.

Tab. A.1: Struktura databáze Comtech

	Databáze COMTECH							
	Textově nezávislá				Textově závislá			
	Sekce 1	...	Sekce 6	Sekce 7 (podvodníci)	Sekce 1	...	Sekce 6	Sekce 7 (podvodníci)
Řečníci	00	00	00	YY	00	00	00	YY
	...	...	...	...	...	...	...	...
	XX	XX	XX	ZZ	XX	XX	XX	ZZ

Struktura jednotlivých nahrávek je následující: **UUUUUUUU-VVVV-WW-sekceX-YY-ZZ.wav**.

- UUUUUUUU - reprezentuje datum (rok, měsíc, den) pořízení nahrávky, 20170310.
- VVVV - je čas (hodina, minuta) pořízení nahrávky, 0831.
- WW - představuje číselný identifikátor řečníka, 00.
- sekceX - udává číslo dané sekce, sekce1.
- YY - označuje, zda se jedná o textově závislou nebo textově nezávislou promluvu (TI,TD).
- ZZ - udává číslo promluvy v dané sekci (01-10).

Výsledná nahrávka poté může vypadat následovně: **20170310-0831-00-sekce1-TI-01.wav**.

B TEXTY PROMLUV PRO TEXTOVĚ NEZÁVISLOU ČÁST DATABÁZE COMTECH

B.1 Sekce 1

- Lidé již počítali s tím, že banky a finanční instituce promítnou do úrokových sazeb hypoték vyšší náklady související se změnou zákona o spotřebitelském úvěru.
- Ve francouzských Alpách včera zasypala lavina hojně využívanou sjezdovku. Původní obavy, že sníh strhl nějaké lyžaře, se nepotvrdily.
- S největší pravděpodobností v jeho podání dojde spíš na předávání pomyslné pochodně mladšímu následovníkovi než na akční scény.
- Když se ale rodina rozroste, stěhují se lidé zpět do města, protože potřebují pro děti občanskou vybavenost.
- Nejdražší objekty, které se za poslední rok prodaly, byly roubenka v Harrachově a rekreační objekt v Jihočeském kraji za téměř 8 milionů korun.
- Poslední dobou se o národních parcích mluví v souvislosti s kontroverzní novelou zákona o ochraně přírody, která by sice měla umožnit větší otevření parků, ale ruku v ruce s tím jde problém, že nejvíce chráněných míst by bylo méně.
- Hned čtyři z deseti lidí museli v posledních pěti letech práci řemeslníka reklamovat. Nejčastěji to bylo kvůli skryté vadě, nedodržení dohodnutých termínů a evidentním nedostatkům v kvalitě.
- Společně se Slováky a Poláky mají totiž nejkratší dovolenou v Evropské unii, a to i přesto, že firmám za poslední roky rostou zisky.
- Není-li vymezení prostoru nástupiště zřejmé nebo není-li viditelným značením takový prostor jinak vymezen, pokládá se pro účely tohoto zákona za nástupiště veřejně přístupný prostor o šířce pěti metrů a délce 30 metrů před a pěti metrů za označníkem zastávky ve směru jízdy dopravního prostředku.
- Přijímací řízení na vysoké školy je pro mladé lidi často náročné nejen kvůli těžkým testům a pohovorům. Zájemci o studium musí také mnohdy kvůli zkouškám jezdit přes celou republiku, a to je navíc stojí nemalé peníze.

B.2 Sekce 2

- Projekt je řešen velmi vzdušně kombinací nízkopodlažní části domů se sedmipatrovými věžemi.
- Vrcholný manažer si podle nich o peníze říkal na schůzkách s manažery společností, které se ucházely o zakázky, na něž Evropská unie chtěla poskytnout dotaci 708 milionů korun.
- Současná politická reprezentace od začátku svého působení nikdy nerozhodla o vyvěšení jiné vlajky, než jakými jsou vlajka České republiky, Evropské unie a Prahy.

- Především proto, že by umožnila Vojenskému zpravodajství k téměř libovolnému provozovateli připojení umístit odposlouchávací zařízení, které by získalo přístup ke všemu, co sítí prochází.
- Jednotlivé státy se pod záminkou budování obrany snaží uchvátit část kyberprostoru pod svou kontrolu.
- Kromě potravin šly nahoru také pohonné hmoty, ceny rostly také v pohostinství v důsledku zavedení elektronické evidence tržeb.
- Zvířetem, které je v současnosti nejvíce dotčeno nelegálním obchodem je luskoun krátkoocasý.
- Oficiální tiskovou konferenci o své případné kandidatuře na post prezidenta sice Miloš Zeman chystá až dnes, už včera ale své rozhodnutí na recepci ve Španělském sále oznámil tisícovce svých podporovatelů.
- Podporovatelé nezávislých kandidátů na prezidenta budou muset na nominační petici připsat ke svému jménu číslo občanského průkazu či pasu.
- Ač se snažil být srdečný a odsoudil skutečnost, že někde se se ženami zachází jako s občany druhé kategorie, vysloužil si kritiku na sociálních sítích.

B.3 Sekce 3

- Olej jednoho z nejvýznamnějších malířů současnosti, jehož obrazy letos představí Národní galerie v Praze, se tak stal dosud nejdražší vydraženou krajinomalbou na světě.
- Po radikální změně prostředí a nástupu na nejistou malířskou dráhu se ocitá v bídě a často na pokraji smrti.
- Nejlepší dny pro nákupy jsou pátek a sobota, protože ve čtvrtek přilétá na letiště kontejner s čerstvým ovocem, zeleninou, rybami a bylinkami.
- Je to takový festival irské kultury, který se koná ve všech městech i na ostrovech.
- Všichni moji přátelé tady se podobají mým kamarádům v Irsku.
- Na začátku nás spoutali a měli jsme hodinu na to se z místnosti dostat, a vyhnout se tak tvrdému výslechu.
- Ve sklepě našli roztrhané občanky, poházené oblečení a zřejmě i skvrny od krve.
- Zkušební provoz potvrdil, že se v době akutního nedostatku řidičů, zvyšujících se cen pohonných hmot a nedostatečného místa v obratišti Nádraží Veleslavín vyplatí provozovat kapacitnější vozidla namísto zdvojených spojů.
- Jestli budeme v realizaci našich záměrů úspěšní, závisí na získání finančních prostředků z dotací, příspěvků a zápůjček.
- Novinkou je také, že se přírodní dobroty zdarma rozšíří i na druhý stupeň základních škol, naopak z mléčného programu se vyjmou mateřské školy.

B.4 Sekce 4

- Podle ministerstva zemědělství nové nastavení programů umožní větší zapojení lokálních producentů zeleniny i mlékáren.
- Tuto změnu přitom banky a finanční instituce avizovaly dopředu, proto se spotřebitelé snažili historicky nejnižších sazeb využít do poslední chvíle.
- Jako pozitivní stránku nové úpravy vnímáme fakt, že díky licenčnímu řízení z trhu ustoupí řada neseriózních podnikatelů, kteří svým jednáním často vrhali stín na slušné poskytovatele úvěrů.
- Češi mají pochopitelně o víc dnů volna zájem a někteří zaměstnavatelé na extra volno lákají nové zaměstnance.
- To nabídne přibližně pětatřicet parkovacích míst, každé z nich bude monitorováno senzorem a řidiči se tak prostřednictvím aplikace v telefonu včas dozvědí, které místo je volné.
- Dokáže například upozornit na zapomenuté osvětlení učeben o víkendu či přetopené místnosti, ale i automaticky zjišťovat a hlásit poruchy.
- Ti upozorňují na to, že chytrá čidla dají možnost získávat přesná data o pohybu lidí, kteří se tak mohou ocitnout pod neustálým dohledem.
- Kromě toho, že se sníží spotřeba energie, umí světla detekovat pohyb a sama se zapnou nebo vypnou.
- Kvůli omezenému rozpočtu na rozšiřování a zkvalitňování sociálních služeb pro seniory nasadili inteligentní řešení.
- Nezaznamenali jsme v souvislosti se spuštěním nové generace vysílání žádné problémy ze strany diváků s přijímáním našich kanálů.

B.5 Sekce 5

- Někteří obchodníci na účtenky tisknou i nepovinné údaje, jako například reklamní sdělení, logo, soutěže, informace k věrnostnímu programu a podobně, účtenku využívají jako nástroj ke komunikaci a záměrně ji prodlužují.
- Většina obchodníků chce za papír spíše ušetřit, a proto hledá, jak rozměry papírového paragonu zmenšit.
- Ačkoliv existují možnosti, jak na účtenkách ušetřit, pro řadu podniků a obchodů je nová povinnost dalším výdajem.
- K pivu se bude na horní palubě podávat převážně staročeská kuchyně, v podpalubí čeká pivnice se studenou kuchyní s nabídkou místní udírny a netradiční specialitou, nakládaným karbanátkem.
- Máme tu projekt senior akademie, vzdělávání pro domovníky, dopravní výchovu dětí a mládeže včetně námi spravovaného dopravního hřiště, na které přijde přes třicet tisíc lidí ročně.

- Unést se dal také šéfredaktor místního zpravodaje, který v jeho březnovém vydání použil při tvorbě ankety s obyvateli této městské části o problémových divočácích smyšlená jména a fotografie zpovídaných.
- Podle ministerstva zdravotnictví postačí, že hospodský nevyzve kuřáka, aby přestal kouřit, a hned by mohl dostat pokutu.
- S přihlédnutím k účelu navrhovaného opatření a k zákonu lze za dětské hřiště považovat venkovní hrací plochu určenou pro hry dětí.
- Povinnost ji měl poskytovat každý železniční dopravce, který následně obdržel kompenzaci za toto zlevněné jízdné od státu.
- Nízká nezaměstnanost je v globálu pěkná věc, ale najít nové kvalitní zaměstnance je pak o to složitější.

B.6 Sekce 6

- Ve chvíli, kdy senioři třeba upadnou, zmáčknou tlačítko, my jim následně voláme zpět, abychom ověřili, zda je nestiskli třeba omylem a jestli máme na místo vyslat i záchranáře.
- Dvojice výzkumníků také připomněla, že dnešní mladá generace dostává často nálepku, že je líná a že do třiceti žije u rodičů.
- Málo se o tom ví, ale od prosince bude teoreticky možné, aby soud, který rozhoduje o určení výživného, navíc stanovil povinnost platit úroky z prodlení.
- Čeští zaměstnanci většinou cestují osobním autem obzvláště ve chvíli, kdy zaměstnavatel sídlí v menší obci nebo na kraji většího města.
- Většímu dojíždění bráníme spíše my, protože si myslíme, že je zaměstnanec tímto dojížděním v práci limitován a pracovní povinnosti tomu přizpůsobuje.
- Aby mohli chovatelé odchov nočních primátů pozorovat a zvířata přitom nebyla nijak rušena, byly do budek nainstalovány kamery, které vnitřní prostor nepřetržitě monitorují.
- Složitý systém a nejednoznačný popis přispívají k dezorientaci a nechuti pokusit se v dané problematice dále vzdělávat.
- V případě žen v lékařství bude navíc nutné vzdělávací program přizpůsobit tak, aby jim nebyl překážkou při zakládání rodiny.
- V poslanecké sněmovně je hodně různých zákonů, a aby se na ten náš dostala řada, je třeba ho zařadit jako pevný bod na program.
- A protože se záměrem souhlasí i vedení Středočeského kraje, které chce ve spolupráci s Českými drahami dál pokračovat, udělila dozorčí rada akciové společnosti předchozí souhlas.

B.7 Sekce 7

- Společnost chce upravit rezervační systém tak, aby si zákazníci mohli ověřit přímo v reálném čase dostupnou kapacitu v ubytovacím zařízení.
- Samostatný pohyb na neznámé dlouhé hotelové chodbě, která je zakouřená, i když jenom dýmovnicemi, je krajně nepříjemný.
- Momentálně se dokončuje proces posuzování vlivů stavby na životní prostředí, veřejnost má prostor se vyjádřit a ministerstvo životního prostředí následně připomínky vyhodnotí a vydá závěrečné stanovisko.
- Příběh je zasazen do českého prostředí, hrdinkami jsou tři generace žen, matka, dcera a vnučka, které prožívají svůj život nejprve v okupovaném a později socialistickém Československu a všechny vyrůstají bez otce.
- Vzhledem k tomu, že výzva byla fakticky založena na principu dobrovolnosti, nebyla vyslyšena zdaleka všemi místními restauracemi.
- Údaje o zákaznících nikdy nepředáváme třetím stranám a pracujeme s nimi v anonymizované podobě.
- Magistrát zrušil rozhodnutí o výběru nejvhodnější nabídky a provede nové posouzení a vyhodnocení nabídek.
- V šedesátých letech se ještě plánoval přesun tramvají pod zem, pak tehdejší vládní a stranické orgány rozhodly, že se v podzemí metropole objeví sovětské metro.
- Na některých místech kanalizace dokonce vinou radioaktivity vyrostli natolik, že je musí deratizační týmy hubit plamenomety.
- Jednou z nejvýraznějších složek návštěvnosti budou ve velikonočním období letos určitě Rusové, kteří se začali do Prahy navracet již na podzim a zejména pak na přelomu roku.

C UKÁZKA VZORKŮ AR FACE DATABASE

Obr. C.1: Ukázka vzorků pro jednoho uživatele.

D OBSAH PŘILOŽENÉ SD KARTY

Přehled adresářů a jejich souborů:

- *Databáze Comtech* - adresář obsahuje veškeré nahrávky tvořící databázi. Struktura databáze je popsaná v příloze A.
- *Zdrojové kódy* - složka obsahuje zdrojové kódy, které byly implementovány pro jednotlivé části dizertační práce. V rámci jednotlivých částí jsou k dispozici zdrojové kódy zajišťující: extrakci parametrů, rozdělení dat, trénování klasifikátorů a vyhodnocení.
- *Dizertační práce* - elektronická verze dizertační práce.