Ing. František Pártl
diplomová práce

Medicínská informatika
2019/2020

Vedoucí práce:
Ing. Kamil Ekštein, Ph.D.

FAKULTA
APLIKOVANÝCH VĚD
ZÁPADOČESKÉ
UNIVERZITY
V PLZNI

<KIV> KATEDRA INFORMATIKY
A VÝPOČETNÍ TECHNIKY

Návrh hashovacího algoritmu pro biometrický podpis uživatele

Abstrakt

Jeden z vědeckovýzkumných projektů řešených na KIV je zaměřen na vývoj systému verifikace identity osob na základě množiny charakteristik (kryptografického otisku osoby) extrahovaných z krátké audiovizuální nahrávky jejich obličeje. Cílem této diplomové práce je návrh, implementace a důkladné testování techniky extrakce otisku operující nad zvukovou stopou této nahrávky.

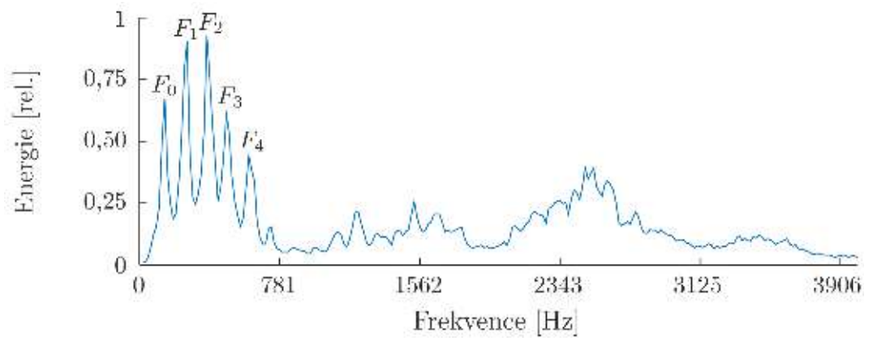
Úvod

Pořízená audiovizuální nahrávka obsahuje detail obličeje osoby, která vysloví větu „Jmenuji se <*jméno*> <*příjmení*> a přijímám tento balík.“, kde údaje v lomených závorkách jsou závislé na dotyčné osobě. Účelem diplomové práce je analýza technik získání kryptografických otisků a návrh, implementace a testování algoritmů vytvářejících otisky alespoň ze zvukové stopy video záznamu.

Východiska, analytická část

Obličej je nejznámějším biometrickým rysem, který lidé využívají každý den. Pro rozpoznávání obličejů se využívají jejich 2D, 3D nebo termografické snímky. Běžné mobilní telefony, které se v projektu mají používat, obvykle disponují pouze fotoaparáty, takže text práce popisuje pouze metody rozpoznávání obličejů na základě jejich 2D snímků. Po nalezení a segmentaci obličeje ze snímku je obličej rozpoznáván buď na základě obrazových dat pomocí statistických metod PCA (*Principal Component Analysis*) či LDA (*Linear Diskriminant Analysis*), nebo podle topologických vlastností obličeje, tj. pozice nosu, očí, úst apod., technikami ASM (*Active Shape Model*) či AAM (*Active Appearance Model*).

Řeč je u člověka přirozenou a také nejčastěji využívanou formou komunikace. Za účelem tvorby řeči existuje v lidském těle tzv. *hlasový trakt*, který je rozdělen na dýchací, hlasové a artikulační ústrojí. Dýchací ústrojí tvoří fundamentální zdroj energie, ale hlas samotný vzniká až v ústrojí hlasovém, kde kmitáním hlasivek vzniká základ tzv. *znělých zvuků*, a artikulačním ústrojí, které různým postavením artikulátorů (jazyka, zubů apod.) tvoří tzv. *neznělé zvuky*. Zatímco frekvenční spektrum těchto neznělých zvuků má spíše povahu šumu, u znělých zvuků se objevují extrémy v oblasti rezonančních frekvencí dutin artikulačního ústrojí. Tyto frekvence se nazývají *formanty* a využívají se hlavně k rozpoznávání obsahu řeči. Poloha formantů, která je ukázána na obrázku 1, je těsně svázána s volumetrickými vlastnostmi dutin osoby a dává hlasu jeho barvu.



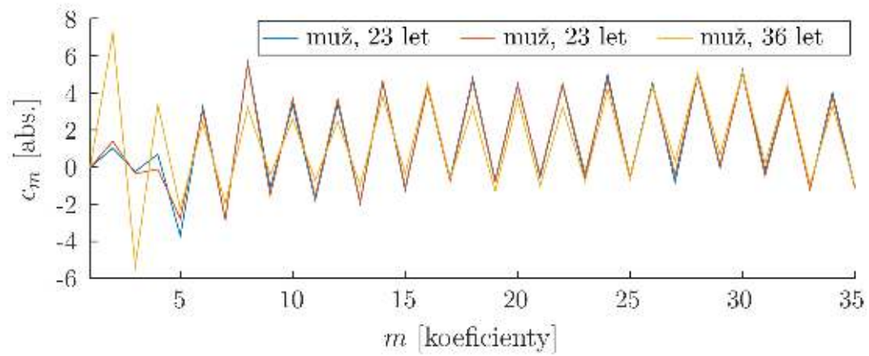
Obrázek 1: Odhad výkonové spektrální hustoty segmentu akustického signálu znělé hlásky s vyznačenými formantovými frekvencemi.

Hlavní aspekty realizace

Vzorky řečového signálu byly analyzovány jak v časové, tak ve frekvenční oblasti. Pro parametrizaci vypočítaných odhadů výkonových spektrálních hustot krátkodobých segmentů signálu byla použita technika MFCC.

Pro účely testování navržených algoritmů byla za pomoci vytvořené mobilní aplikace a pěti přispěvatelů vytvořena množina multimodálních podpisů čítající celkově 153 video záznamů od 34 dobrovolníků (15 žen a 19 mužů). Pro zvýšení robustnosti navržených algoritmů byly video záznamy pořizovány v různých prostředích a s různou úpravou vzhledu dobrovolníka.

Návrh samotných algoritmů započal implementací aplikace s vizualizačními experimenty prováděnými přímo nad vzorky pořízeného datasetu. Pomocí této aplikace vznikl i obrázek 2, kde je vidět minimální, resp. znatelný rozdíl mezi koeficienty téhož, resp. různých dobrovolníků.



Obrázek 2: MFCC koeficienty vypočítané z řečových signálů dvou dobrovolníků z datasetu.

Všechny techniky byly implementovány v rozsáhlé knihovně *libpe* napsané v jazyce C++.

Na základě vizualizačních experimentů byly navrženy například algoritmy využívající techniky DTW, pomocí které je určována podobnost dvou průměrných vektorů MFCC koeficientů. Pokud je tato podobnost menší než konstantní práh, otisky jsou považovány za ekvivalentní. Jiné algoritmy chápou vektory MFCC koeficientů jako body n-rozměrného prostoru, ve kterých je technikou K-means++ určován konstantní počet shluků. Dva otisky jsou poté označeny za ekvivalentní, pokud střední kvadratická chyba počítaná vždy ze dvou nejbližších centroidů je menší než práh.

Nakonec byly navržené algoritmy ve svých stovkách tisíc konfiguracích testovány pomocí silně paralelizované konzolové aplikace, a to přímo nad vzorky nashromážděného datasetu.

Dosažené výsledky

Při stanovení prahové hodnoty tak, aby otisky téže osoby byly ekvivalentní, bylo dosaženo až 18% poměru nerozlišitelných otisků. Tento údaj platí pro algoritmy založené na technikách DTW i K-means++, ale pro svůj determinismus byla nakonec zvolena technika DTW.

Závěr

Pomocí implementovaných algoritmů je možné z řečových signálů extrahovat takové charakteristiky, které jsou pro všechny podpisy téže osoby ekvivalentní. Zároveň algoritmy správně verifikují až 82 % ze všech kombinací akustických podpisů nashromážděného datasetu. Společně s charakteristikami z vizuální složky toto procento ještě stoupne.