VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV AUTOMATIZACE A MĚŘICÍ TECHNIKY

DEPARTMENT OF CONTROL AND INSTRUMENTATION

SYSTÉM POČÍTAČOVÉHO VIDĚNÍ PRO ROZPOZNÁVÁNÍ EMOCÍ

A COMPUTER VISION SYSTEM FOR EMOTION RECOGNITION

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE — Bc. Jan Wójcik

AUTHOR

VEDOUCÍ PRÁCE — Ing. Ilona Janáková, Ph.D.

SUPERVISOR

BRNO 2023

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Diplomová práce

magisterský navazující studijní program **Kybernetika, automatizace a měření**

Ústav automatizace a měřicí techniky

Student: Bc. Jan Wójcik ***ID:*** 211192

Ročník: 2 ***Akademický rok:*** 2022/23

NÁZEV TÉMATU:

Systém počítačového vidění pro rozpoznávání emocí

POKYNY PRO VYPRACOVÁNÍ:

Úkolem studenta je pomocí vhodných metod zpracovávat obraz/sekvenci obličeje osoby za účelem získání relevantních dat. Cílem je rozpoznávat jakou emoci aktuálně prožívá (smutný, veselý, překvapený atd.). Předpokládá se využití výsledného zařízení u lidí, kteří s rozpoznáváním emocí mají problém (např. autisté).

1. Proveďte rešerši v oblasti detekce obličeje a příznaků vhodných pro rozpoznávání emocí.
2. Na základě rešerše vyberte vhodnou metodu. Při volbě metody zvažte/berte v úvahu možnosti vyhodnocení v reálném čase a adaptace systému na konkrétní, novou osobu.
3. Navrhněte vhodné snímací (kamera, optika) a vyhodnocovací (PC, případně minipočítač) hardwarové prostředky a vhodné opticko-mechanické uspořádání.
4. Pořiďte dostatečně rozsáhlou a pestrou databázi reálných snímků/sekvencí.
5. Zvolené algoritmy implementujte. Navrhněte vhodné uživatelské rozhraní.
6. Dosažené výsledky zhodnoťte. Definujte omezující podmínky.

DOPORUČENÁ LITERATURA:

BOTA, Patricia J., Chen WANG, Ana L. N. FRED a Hugo PLACIDO DA SILVA. A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals. IEEE Access [online]. 2019, 7, 140990-141020 [cit. 2022-09-08]. ISSN 2169-3536. Dostupné z: doi:10.1109/ACCESS.2019.2944001

Termín zadání: 6.2.2023 ***Termín odevzdání:*** 17.5.2023

Vedoucí práce: Ing. Ilona Janáková, Ph.D.

doc. Ing. Petr Fiedler, Ph.D.
předseda rady studijního programu

Abstrakt

Diplomová práce se zabývá návrhem systému pro rozpoznávání emocí, který by mohl být využit jako nástroj pro zlepšení komunikace s osobami s poruchou autistického spektra. Pro rozpoznávání emoce budou využívány data z kamery, jedná se tedy o aplikaci počítačového vidění. Práce se zabývá oblastmi jako je detekce obličeje, extrakce relevantních příznaků, hledání vhodného datasetu nebo návrh klasifikátoru. V rámci práce se uvažuje i možnost adaptace systému pro rozpoznávání emocí konkrétní osoby.

Klíčová slova

Emoce, rozpoznávání, počítačové vidění, detekce obličeje, rozhodovací strom

Abstract

The master's thesis deals with the design of an emotion recognition system, which could be used as a communication tool for people with autism spectrum disorder. Camera data will be used for emotion recognition, so it will be a computer vision application. The work deals with areas such as face detection, extraction of relevant features, finding a suitable dataset or designing a classifier. The work also considers the possibility of adapting the system for recognizing the emotions of a particular person.

Keywords

Emotion, recognition, computer vision, face detection, decision tree

Bibliografická citace

WÓJCIK, Jan. *Systém počítačového vidění pro rozpoznávání emocí.* Brno, 2023. Dostupné také z: https://www.vut.cz/studenti/zav-prace/detail/151683. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav automatizace a měřicí techniky. Vedoucí práce Ilona Janáková.

Prohlášení autora o původnosti díla

Jméno a příjmení studenta:	*Jan Wójcik*
VUT ID studenta:	*211192*
Typ práce:	*Diplomová práce*
Akademický rok:	*2022/23*
Téma závěrečné práce:	*Systém počítačového vidění pro rozpoznávání emocí*

Prohlašuji, že svou závěrečnou práci jsem vypracoval samostatně pod vedením vedoucí/ho závěrečné 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 závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

.

V Brně dne: 13. května 2023 ----------

podpis autora

Poděkování

Děkuji vedoucí diplomové práce Ing. Iloně Janákové, Ph.D. za její podporu a vedení během psaní mé diplomové práce. Její odborné rady, připomínky a zpětná vazba mi velmi pomohly v dosažení co nejlepšího výsledku.

V Brně dne: 13.05.2023 ----------

podpis autora

Obsah

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

Úvod

Systém pro rozpoznávání emocí má sloužit jako nástroj pro analýzu dat v reálném čase z kamery a vyhodnocení nejpravděpodobnější emoce, kterou snímaná osoba projevuje. Detekovaná emoce může nabývat různých hodnot, jedná se tedy o více třídovou klasifikaci. Systém má rozpoznávat emoce jako je například, štěstí, zlost, smutek, překvapení nebo znechucení, přičemž výsledná emoce má být uživateli vhodným způsobem zobrazována. Systém má být primárně používán při komunikaci s osobami s poruchou autistického spektra. Tyto osoby mají běžně s rozpoznáním emocí lidí problém a systém by jim v komunikaci mohl pomoci. Ačkoliv je primární účel systému již definován, neměl by být problém systém použít i pro rozpoznávání emocí k jiným účelům.

Systém by mohl najít využití v mateřské škole dle §16 odstavce 9 školského zákona u dětí s poruchou autistického spektra. V praxi by mohl být systém využit způsobem, že pedagog povede dialog s dítětem, přičemž sám bude snímán kamerou. Systém pak bude vyhodnocovat emoci pedagoga, která bude zobrazována vhodným způsobem dítěti. Dítě tedy bude moci zároveň vnímat reálnou emoci pedagoga i emoci predikovanou systémem. Tímto způsobem bude moci porovnávat oba výstupy a reálný výraz obličeje si bude moct spojit se správně klasifikovanou emocí.

Práce se zabývá rešerší v oblasti rozpoznávání emocí, převážně v oblasti detekce a zarovnání obličeje, hledání vhodných příznaků, jejich reprezentace a klasifikace. V rámci práce jsou porovnávány různé přístupy a metody, přičemž je snaha implementovat již vyvinuté algoritmy a řešení a případně je vhodným způsobem modifikovat. Mimo analýzy obličeje se práce zabývá i využitím gest a postojů pro emoční rozpoznávání. Předmětem rešerše je také hledání vhodné databáze obrázků, která je pro návrh systému nezbytná. Kvalita fungování metod je ověřována na těchto databázích a výsledky testování jsou vhodným způsobem interpretovány. Cílem práce je vytvořit systém s jednoduchým uživatelským rozhraním, který bude snadno použitelný na osobním počítači nebo notebooku s webkamerou. Vzhledem k rozmanitosti způsobů vyjádření emocí se uvažuje i možnost doučení systému na emoci, kterou nemusí ve výchozím nastavení správně rozpoznat.

1. REŠERŠE

Kapitola obsahuje informace o možném využití rozpoznávání emocí člověka v praktických úlohách. Na vhodných příkladech demonstruje nasazení rozpoznávání emocí a jeho využití v komerčních i nekomerčních aplikacích. Dále jsou popsány metody získávání vhodných příznaků z digitálních snímků a videí. Závěr kapitoly je věnován nalezeným řešením, které lze v práci využít.

1.1 Využití rozpoznávání emocí v praxi

Emoce jsou nedílnou součástí člověka a z velké části se objevují v jeho neverbální komunikaci. Projevením emoce může člověk vyjadřovat svou aktuální náladu, psychický stav nebo pocity, přičemž vyjádření emoce může být vědomé, ale i nevědomé. Dlouhou dobu bylo rozpoznávání emocí pouze disciplínou lidí. Emoce a výrazy obličeje člověka byly proto zkoumány pouze z hlediska oboru psychologie. V posledních letech se však objevily nápady, jak emoční rozpoznávání využít masově, a proto následoval výzkum i v oblasti techniky. S rozvojem technologií kamer a rostoucím výpočetním výkonem se vývojáři různých týmů snaží rozpoznávání emocí realizovat strojově.

Využití počítačového vidění pro rozpoznávání obličeje našlo uplatnění v mnoha komerčních i nekomerčních aplikacích. Výhoda čtení emocí strojově je především v jeho nízkých nákladech, rychlosti nebo možnosti zpracování velkého objemu dat v reálném čase. Pro vyhodnocování velkého množství dat by bylo zkoumání člověkem velice drahé, a proto se rozpoznávání emocí používá v oblastech jako je, průzkum trhu, personalizované a chytré automobily, pohovory a rozhovory nebo testování videoher [12].

Průzkum trhu

V rámci testování služeb či výrobků je nutné získat a zpracovat enormní množství statistických údajů. Pro získání dat jsou často používány dotazníky, které však v dnešní uspěchané době nenesou velkou oblibu. Proto některé společnosti začaly pro průzkum trhu používat emoční rozpoznávání pomocí počítačového vidění. Příkladem může být společnost CocaCola, která pro svou limitovanou edici „Marshmello“ získávala data prostřednictvím webové aplikace. Fanoušci CocaColy se sami natáčeli svým mobilním telefonem, zatímco ochutnávali novou příchuť světoznámé limonády. Na pozadí jejich emoce zkoumala umělá inteligence od společnosti MorphCast [11].

Personalizované a chytré automobily

V automobilech řidiči často prožívají stresové či jiné negativní situace. Rozpoznáváním emoce a přizpůsobením prostředí auta se pokoušejí vývojáři zlepšit požitek z jízdy či jeho bezpečnost. Díky informaci o emoci řidiče lze měnit teplotu, osvětlení nebo pozměnit

žánr právě poslouchané hudby. Systém pak dokáže detekovat i únavu řidiče, čímž může předejít potencionálnímu mikrospánku. [12][13]

Obrázek 1 - Signály pro emoční rozpoznávání a jejich možné umístění ve vozidle [26]

Pohovory a rozhovory

Uplatnění lze najít i v oblasti lidských zdrojů z hlediska video pohovorů a rozhovorů. Algoritmus rozpoznávání emocí může pomoci s výběrem vhodného kandidáta na pracovní pozici. Výhoda algoritmu je ta, že nebere v potaz možnou zaujatost osoby, která pohovor provádí. Tímto způsobem umělá inteligence pomáhá například firmě Unilever, která je jedním z největších výrobců potravin, výrobků pro domácnost či péči o tělo [14].

Testování videoher

Rozpoznávání emocí má své místo i v herním průmyslu. Videohry mají za cíl v hráčích evokovat různé emoce a pocity. Zpětnou vazbu ve formě pocitů, názorů a návrhů na vylepšení her poskytují testeři. Ti hledají nedostatky, chyby a podávají vývojářům zprávy o tom, jak na ně hra působí. Nicméně zpětné vybavení pocitu, jaký v testerovi videohra zanechala, nemusí být úplně přesné. Rozpoznávání emocí má proto veliký potenciál v této oblasti. Vývojáři mohou získávat zpětnou vazbu v reálném čase, zatím co tester hru hraje. [12]

1.2 Metody získávání příznaků

Příznaky v oblasti počítačového vidění lze definovat jako kusové informace ze zkoumaného obrazu. Tyto kusové informace se pak využívají pro řešení daného

problému, v tomto případě rozpoznání emoce člověka.

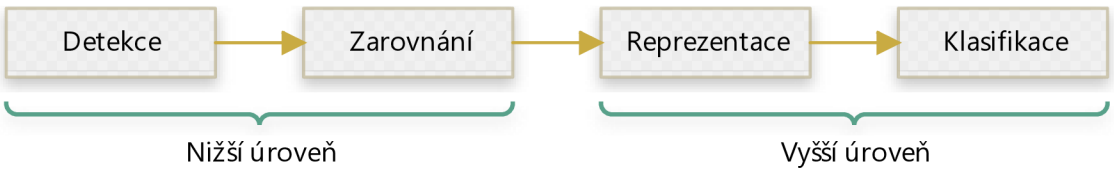
Metody získávání příznaků pro rozpoznávání emocí mohou vycházet z různých principů. První způsob vychází z extrakce příznaků ze samotného obličeje. Získávání příznaků z obličeje je přirozeně nejpoužívanější metoda. Člověk totiž pomocí mimiky obličeje vyjadřuje své emoce, aniž by si to sám uvědomoval. Hledané příznaky pak mohou být pozice a tvar úst, pozice obočí, úroveň otevření očí atd. Tento způsob využívá principů rozpoznávání obličeje, což je překlad z anglického „facial recognition“. Člověk však může svou emoci projevovat i jinými způsoby, jako je hlasitost a barva hlasu, slovní důraz či gestikulace. Zkoumání hlasitosti, barvy hlasu nebo důrazu je spíše kompetencí oblasti zpracováni řeči, a proto tato metoda nebude v práci uvažována. Zajímavou úlohou je ale zkoumání posledního zmíněného projevu, tedy gestikulace.

Tento způsob nevyužívá obličej pro extrakci příznaků vůbec. Příznaky se totiž extrahují na základě detekce postavení a gest, které snímané osoby vykazují. Tento přístup může mít své výhody, jelikož dataset může být cenzurovaný. Obrázky nebo videa v databázi mohou obsahovat obličeje, které jsou rozmazané, a tedy osoby v datasetu nelze nijak identifikovat. Tento dataset pak například může splňovat obecné nařízení o ochraně osobních údajů – GDPR.

Oba způsoby, jak získávání příznaků z obličeje, tak z postoje a gest budou detailněji popsány dále v textu.

1.2.1 Extrakce příznaků z obličeje

Metoda vychází z rozpoznávání obličeje, přičemž moderní přístup pro rozpoznávání obličejů se skládá ze čtyř hlavních částí. Těmi jsou detekce, zarovnání, reprezentace a klasifikace [6]. Detekce a zarovnání by se z hlediska řetězce zpracování obrazu daly zařadit do předzpracování a segmentace, tedy do nižší úrovně zpracování obrazu. Naopak reprezentaci lze zařadit do vyšší úrovně zpracování, tedy popisu, detekce objektů atd. Řetězec čtyř hlavních částí rozpoznávání obličeje lze vidět na Obrázek 2.

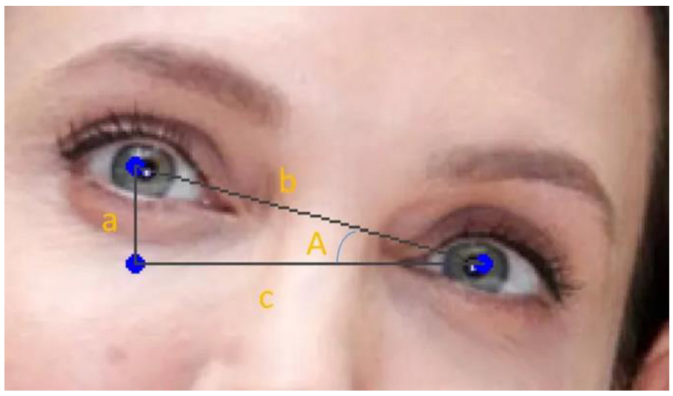


Obrázek 2 - Řetězec rozpoznávání obličeje

Detekce a zarovnání

Pro detekci lze využít opensource knihovny jako je OpenCV nebo Dlib, které jsou v oblasti počítačového vidění velice používané. Obě knihovní řešení umožňují detekovat obličej v reálném čase. Pro detekci využívají různé metody, jako například Haar cascade classifier (Haarovy kaskádní klasifikátory), Single shot MultiBox detektory, Histogram of Oriented Gradients (HOG) či Maximum margin object detector (MMOD) [8].

Knihovny mimo detekce obličeje umožňují i detekovat pozice očí, což lze využít pro další fázi řetězce – zarovnání obličeje. Pomocí detekovaných souřadnic očí lze sestrojit pomyslný trojúhelník, jako na Obrázek 3.



Obrázek 3 - Zarovnání obličeje pomocí rotace [8]

Pomocí dvou bodů lze snadno dopočítat libovolné strany trojúhelníka a pomocí goniometrických funkcí pak dopočítat úhel A. O tento úhel je pak potřeba otočit obrázek, aby došlo k jeho zarovnání. Příklad lze vidět na Obrázek 4.

Obrázek 4 - Příklad zarovnání obličeje [8]

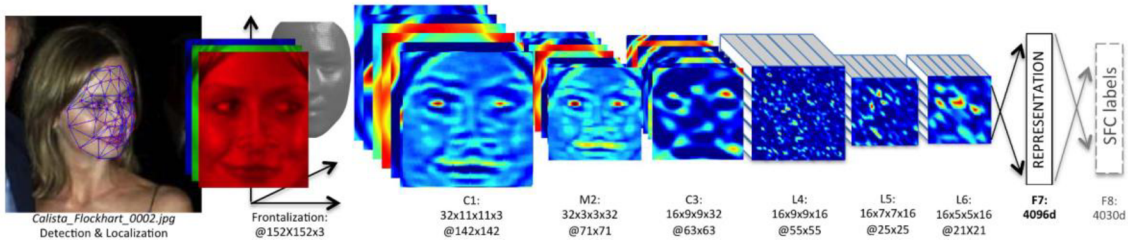
Výše zmíněná metoda zarovnání je jen jedna z mnoha a jedná se o jednu z nejjednodušších způsobů, jak zajistit zarovnání obličeje. V oblasti zarovnání obličeje byly vyvinuty i další metody, které se mohou lišit v rychlosti, přesnosti či jiných srovnávacích parametrech. Patří mezi ně například LBF (Local Binary Features), ERT (Ensemble of Regression Trees) nebo DAN (Deep Aligment Network) [10].

Dále v řetězci rozpoznávání obličeje následuje reprezentace, která bude popsána podrobněji než předchozí dva procesy. Pro rozpoznávání obličejů byla v minulosti vyvinuta řada modelů, fungujících na konvolučních neuronových sítích, dále KNN. Tyto modely jsou schopny vstupní data ve formě obrázků přepočítat na multidimenzionální vektory, které reprezentují sadu příznaků daného obličeje. V tomto případě jde tedy o algoritmické řešení daného problému [6].Vycházet lze ale i z fyziologie člověka. Existuje způsob jak pomocí systému pravidel „zakódovat“ téměř jakýkoliv výraz člověka,

a tedy i jeho emoci. Příznaky v tomto případě reprezentuje kód výrazu člověka a systém se nazývá FACS [2]. Oba způsoby reprezentace, jak použití modelů, tak použití systému kódování bude popsáno dále v textu.

Reprezentace s využitím naučených modelů

Jak již bylo v textu zmíněno, modely jsou realizovány konvolučními neuronovými sítěmi. KNN je vždy natrénována tak, aby byla schopna klasifikovat jednotlivé osoby [9]. Obecně se konvoluční neuronové sítě skládají z vrstev různých typů, jako jsou konvoluční, aktivační, redukční, plně propojena a další. Jako příklad je uveden model DeepFace, který byl vyvinut společností Facebook. Architekturu KNN modelu DeepFace lze vidět na Obrázek 5.



Obrázek 5 - DeepFace architektura citace

Vstupem KNN je barevný obrázek o velikosti 152x152x3 pixelů. Jedná se tedy o tříkanálový RGB (red, green, blue) snímek. Ten je předložen první konvoluční vrstvě, která pomocí konvolučních filtrů (11x11x3 pixelu) na výstupu generuje 32 příznakových map (features map). Ty jsou pak předloženy redukční vrstvě, která počítá maximum z oblasti 3x3 pixelu a krokem 2. Výstup redukční vrstvy je předložen druhé konvoluční vrstvě. V těchto prvních třech vrstvách se z obličeje extrahují nízko úrovňové příznakové vektory, jako je textura či hrany. Následující tři vrstvy jsou lokálně propojeny, přičemž jejich filtry nejsou stejné pro celou příznakovou mapu, ale mění se na základě vyhodnocované oblasti. Například oblast mezi očima vykazuje velice odlišné vlastnosti v porovnání s oblastí mezi nosem a ústy. Z tohoto důvodu nelze využít klasickou konvoluční vrstvu s jedním filtrem, protože by nebyla schopna detekovat efektivně hledané příznaky. U lokálně propojených vrstev DeepFace modelu se využívá toho, že jsou všechny vstupní snímky zarovnány, potom se na různé části příznakových map aplikují různé filtry. Poslední dvě vrstvy jsou plně propojeny, takže každý výstup je spojen se všemi vstupy. [9]

Výstup první plně propojené vrstvy je vektor reprezentující obličej na zkoumaném obrázku. Tento vektor má rozměr 4096, jde tedy o multidimenzionální vektor, který obsahuje příznaky konkrétního obličeje [9]. Pro člověka je nemožné si představit prostor který má více než tři dimenze, nicméně z hlediska matematiky není problém s těmito vektory počítat. Obdobným způsobem jsou sestaveny architektury dalších CNN modelů, které jsou shrnuty v Tabulka 1. Tabulka obsahuje informace o názvu modelu, velikosti vstupního obrazu, rozměru výstupního vektoru a vývojáře modelu.

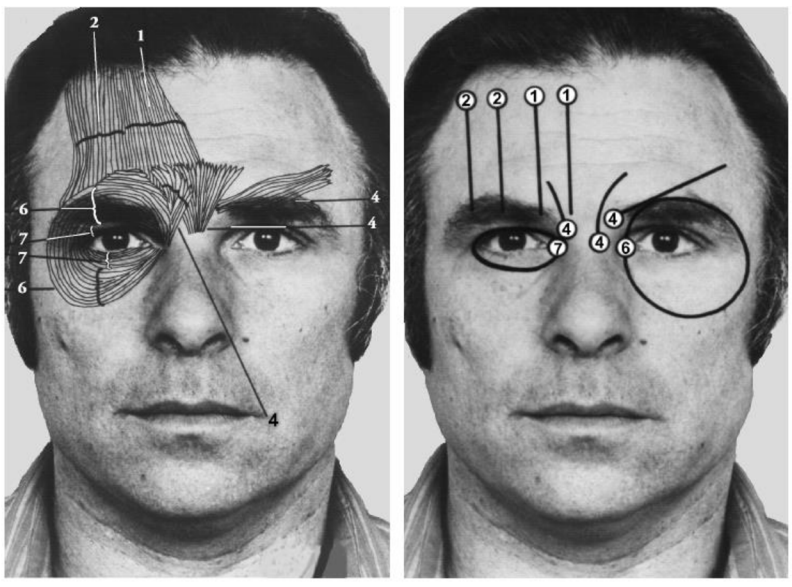
Tabulka 1 - Tabulka vyvinutých modelů [6]

Název modelu	**Vstup**	**Výstup**	**Vývojář**
VGG-Face	224 x 224 x 3	2622	Oxford university
FaceNet	160 x 160 x 3	128	Google
OpenFace	96 x 96 x 3	128	Carnegie Mellon University
DeepFace	152 x 152 x 3	4096	Facebook
DeepID2	55 x 47 x 3	160	University of Hong Kong
Dlib	150 x 150 x 3	128	Davis E. King

Reprezentace pomocí FACS

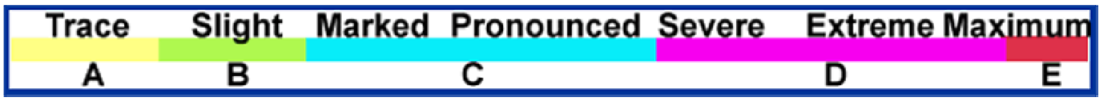
Toto řešení extrahuje příznaky z obličeje člověka. Zkratka FACS stojí pro „Facial Action Coding System“, volně přeloženo tedy „Kódovací systém akcí obličeje“. Akcí obličeje se rozumí svalová aktivita, která je vyvolána změnou výrazu obličeje. Každá obličejová akce sleduje konkrétní svaly obličeje. V případě, že u sledovaného svalu došlo ke kontrakci, akce obličeje je aktivní, v případě že je sval uvolněný akce obličeje je neaktivní [1]. Každá akce obličeje má své unikátní značení, které se skládá ze zkratky AU (vychází z anglického „Action unit“) a celého čísla. V pokračujícím textu práce bude libovolně používán buď pojem „akce obličeje“ nebo zkratka z anglického originálu AU.

FACS rozlišuje řádově desítky různých AU, příkladem může být AU1, která je pojmenována „Inner brow raiser“ a označuje zvednutí vnitřní strany obočí [1]. Svaly realizující tuto akci obličeje lze vidět na Obrázek 6, kdy čísla korespondují s kódováním FACS. Obrázek na levé straně zobrazuje jednotlivá svalová vlákna, jejichž aktivita je sledována. Na pravé straně je pak znázorněn směr pohybu svalu pro danou AU.



Obrázek 6 - Příklad sledovaných AU [1]

Mimo samotnou přítomnost sledované akce obličeje je dle FACS hodnocena také její intenzita. Akce obličeje lze z pohledu míry viditelnosti řadit do několika úrovní intenzity značených písmeny A až E [1]. Nulová nebo téměř zanedbatelná míra viditelnosti AU je přiřazena do úrovně A. Úroveň B se používá pro zařazení AU o mírné viditelnosti. Znatelná nebo zřetelná AU patří do úrovně C, vážné až extrémní viditelnost pak do úrovně D. Pro maximální vyjádření viditelnosti AU se používá písmeno E. Závislost mezí mírou viditelnosti a úrovní intenzity lze vidět na následujícím obrázku.



Obrázek 7 - Závislost míry viditelnosti AU a její úrovně intenzity [1]

Samotné akce obličeje nenesou žádnou informaci o aktuální emoci člověka. Přesto lze FACS použít pro interpretaci emoce, Kombinací 9 AU v horní oblasti obličeje a 18 AU v dolní oblasti obličeje lze zakódovat téměř libovolný výraz člověka. Jednotlivé kombinace AU pro vybrané emoce zobrazuje Tabulka 2.

Tabulka 2 - Tabulka AU pro jednotlivé emoce [2][3]

Emoce	**Akce obličeje**
Štěstí	12 6+12
Smutek	1+4 1+4+11/15 1+4+15+17 6+15 11+17 1
Překvapení	1+2+5+26/27 1+2+5 1+2+26/27 5+26/27
Strach	1+2+4 1+2+4+5+20+25/26/27 1+2+4+5+25/26/27 1+2+4+5 1+2+5+25/26/27 5+20+25/26/27 5+50 20
Zlost	4+5+7+10+22+23+25/26 4+5+7+10+23+25/26 4+5+7+17+23/24 4+5/7 17+24
Znechucení	9/10+17 9/10
Pohrdání	12+14

Konkrétní emoce mohou být interpretovány různými kombinacemi AU. Lze pozorovat, že například emoce strach má 8 různých možností zakódování. Každá osoba může projevovat tuto emoci jiným způsobem. Rozdíly lze pozorovat na Obrázek 8. Hlavní rozdíl v projevené emoci je pozice a tvar úst. Na levé straně obrázku pozorujeme ženu, která má ústa spíše přivřená a koutky úst směřují směrem dolů k bradě, která setrvává ve své přirozené pozici. Tomuto odpovídá AU20 – Lip Strecher a AU25 – Lips Part [2][3]. Naopak Kevin na pravé straně obrázku má ústa široce otevřená a jeho brada výrazně poklesla. Tento výraz odpovídá AU26 – Jaw Drop a AU27 – Mouth Stretch.

Obrázek 8 - Rozdílné vyjádření emoce strach [3][4]

Klasifikace

Posledním procesem v řetězci na Obrázek 2 je klasifikace. Způsob klasifikace emoce bude silně záviset na předchozím procesu reprezentace. V případě, kdy příznaky zkoumaného obličeje budou reprezentovány kódem FACS, lze při klasifikaci jednoduše vycházet z Tabulka 2. Klasifikovaná emoce bude vycházet z detekovaných akcí obličeje a jejich kombinací. Jako klasifikátor by také bylo možné využít rozhodovací strom, který by rozhodoval na základě detekovaných akcí obličeje.

V případě reprezentace obličeje jako vektoru příznaků, který se používá u naučených modelů CNN, je situace o něco komplikovanější. Pro klasifikaci by bylo možné použít natrénovanou neuronovou síť, která by zkoumané vektory příznaků řadila do několika emočních tříd. Pro naučení neuronové sítě by bylo nutné vhodně nastavit její topologii a parametry a také mít vhodný dataset.

1.2.2 Extrakce příznaků z gest a postojů

Jak již bylo v textu uvedeno výše, u této metody se s výhodou využívá faktu, že data, ze kterých jsou příznaky extrahovány mohou být anonymizována. Obličej snímaných osob může být cenzurován, protože pro získávání příznaku není jeho zkoumání relevantní. Z vědeckých výzkumů, ale i každodenní praxe je totiž zřejmé, že na projevu emocí se ve velké míře podílí i neverbální chování člověka [15]. Příkladem může být gestikulace, postoje a obecně projevy, které lze zařadit do oblasti zvané řeč těla. Vhodné příznaky pro rozpoznávání emocí lze extrahovat právě i z výše popsaných neverbálních projevů.

Mnoho výzkumů se zaměřovalo na gesta ilustrativní, jako je například zamávání při pozdravu či loučení. Pro rozpoznávání emocí však tyto typy gest nebyly příliš vhodné, protože se člověk často snaží své emoce skrývat – hlavně ty negativní. Člověk může úmyslně gestikulovat tak, aby jeho emoce nebyla jednoduše čitelná, což znemožňuje použití ilustrativních gest pro rozpoznávání emocí. Výzkumy však ukázaly, že existuje speciální podoblast gest, takzvaná mikro-gesta (MG). Mikro-gesta jsou pro rozpoznávání emocí mnohem vhodnější, protože umožňují poznat i potlačovanou nebo skrývanou emoci člověka. Rozdíl mezi ilustrativními a mikro gesty je ten, že MG jsou projevována nevědomky. Mezi příklady mikro-gest patří například promnutí očí, kousání nehtů, poškrábání se na zádech a další. Funkcí mikro-gest není doprovázení verbálního projevu, jako je tomu u ilustrativních gest. Naopak MG jsou spontánně vyvolána nějakým vnitřním pocitem člověka. Příkladem může být MG ruce křížem, které indikuje obranný postoj sledované osoby. Pokud se osoba usmívá, ale zároveň má ruce křížem, jde o určitou negaci. Úsměv na obličeji logicky vybízí emoci hodnotit jako kladnou – šťastný. Může se však jednat o úmyslné zastírání opravdové emoce a MG zkřížení rukou je pro přesnější vyhodnocení emoce vhodnější. [15]

Obrázek 9 - Příklady mikro-gest [15]

Někteří lidé mohou nesouhlasit s tím, že by u nich zkřížení rukou znamenalo obranný postoj nebo negaci. Mohou tvrdit, že je to gesto, při kterém se zkrátka cítí pohodlně, a že to je jejich přirozený reflex mít ruce křížem. Je nutné si však uvědomit, že ačkoliv se člověk se zkříženýma rukama může cítit pohodlně, neznamená to, že je uvolněný a má dobrou náladu. Není přirozené, aby člověk ve společnosti, se kterou se dobře baví měl ruce křížem. Toto gesto je běžně lidmi projevováno v nepříjemných situacích, kdy je člověk nervózní, čelí otázkám, na které nechce odpovědět, nebo se cítí zranitelný. Zkřížené ruce pak představují pomyslnou bariéru, za kterou se člověk schovává. V tu

chvíli mu to samozřejmě připadá přirozené a pohodlné, ovšem neznamená to, že se pohodlně cítí [38].

1.3 Možná řešení

Kapitola pojednává o již vyvinutých řešeních a nástrojích, které by pro rozpoznávání emocí bylo možné využít. Stručně popisuje princip fungování daného řešení a poskytuje detailnější informace o vývojáři, třídě klasifikovaných emocí, datasetu atd.

DeepFace

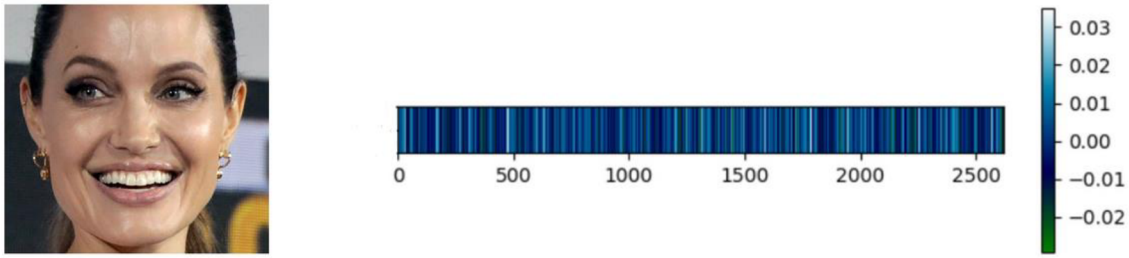
Je odlehčená python knihovna pro rozpoznávání obličejů a analýzu obličejových atributů, jako je věk, pohlaví, emoce a rasa. DeepFace realizuje všechny čtyři fáze řetězce rozpoznávání obličeje, které lze vidět na Obrázek 2, tedy detekci, zarovnání, reprezentaci a klasifikaci. Pro detekci obličeje a jeho zarovnání využívá metody jako je například: OpenCV, SSD, MTCNN, RetinaFace a Mediapipe. Metody pro detekci obličeje lze přepínat změnou příslušného argumentu volané funkce v kódu. Porovnání výstupního obrazu pro jednotlivé metody lze vidět na následujícím obrázku.

Obrázek 10 - Rozdíly metod detekce a zarovnání obličeje [8]

Všechny výstupní obrázky musí mít stejný rozměr, jelikož budou dále zpracovávány CNN modelem. Oříznuté obličeje však mají různé rozměry a natočení pro každou metodu. Aby nedocházelo k jejich deformaci, přidává DeepFace k oříznutým obrázkům černé pozadí, všechna výstupní data tedy mají stejný rozměr. Dle autora jsou nejpřesnějšími metodami RetinaFace a MTCNN, bohužel však na úkor rychlosti. Pro rychlé zpracovávání dat je doporučena metoda OpenCV nebo SSD. [20]

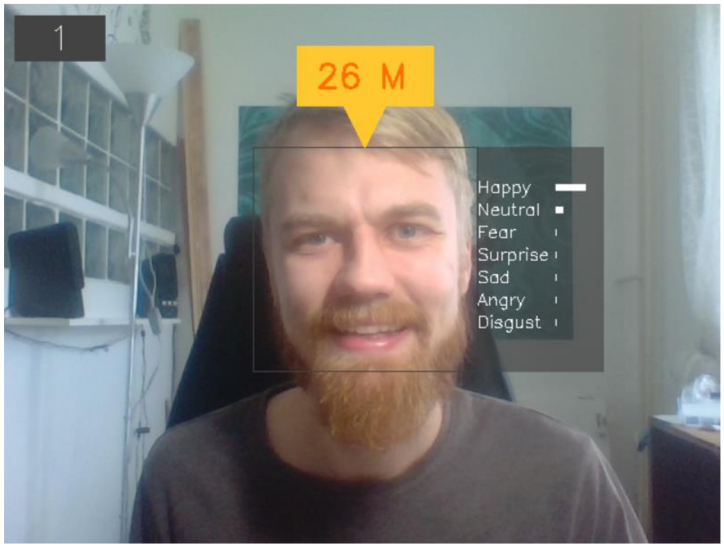
Pro reprezentaci a klasifikaci pak využívá již naučené CNN modely pro rozpoznávání obličejů. Ačkoliv je knihovna pojmenována po modelu DeepFace, lze pro reprezentaci a klasifikaci volit i modely jako je například: VGG-Face, Facenet, OpenFace, Dlib atd.

Vstupem pro tyto modely je obrázek o daném rozměru, výstupem je normalizovaný výstupní příznakový vektor, který reprezentuje daný obličej. Více informací o rozměrech dat pro zmíněné modely lze najít v kapitole 1.2.1 a Tabulka 1. Tento příznakový vektor je pak použit pro veškeré operace, například pokud je potřeba zjistit podobnost osoby na dvou různých snímcích, jsou porovnávány přímo příznakové vektory, a nikoliv fyzické obrázky [6][20]. Znázornění příznakového vektoru modelu VGG-Face je vidět na následujícím obrázku.



Obrázek 11 - Příznakový vektor modelu VGG-Face [6]

DeepFace nabízí již hotové funkce pro rozpoznávání obličejů a emocí v reálném čase. Pomocí knihovních funkcí lze detekovat na snímku obličej, porovnávat obličeje, odhadovat věk, pohlaví nebo emoci subjektu a další. Knihovna nabízí již naučené modely, které lze jen stáhnout a použít. Emoční model DeepFace dokáže rozpoznávat sedm druhů emocí, konkrétně zlost, znechucení, strach, štěstí, neutrální výraz, smutek a překvapení. Pro každou emoční třídu je emočním modelem přiřazena procentuální hodnota, přičemž suma všech hodnot dává 100 %. Ukázku výstupu z DeepFace lze vidět na následujícím obrázku.



Obrázek 12 - Ukázka výstupu DeepFace

Emoční model má dle vývojáře DeepFace celkovou správnost klasifikace 57 % [26]. Tato hodnota byla testována na datasetu FER2013. Hlavním vývojářem DeepFace je Sefik Ilkin Serengil, který vystudoval informatiku na turecké univerzitě Galatasaray. [6][20]

OpenFace

Je sada nástrojů určená pro vývojáře v oblasti počítačového vidění a strojového učení. Umožňuje analýzu obličeje člověka, konkrétně dokáže detekovat orientační body obličeje, odhadovat pozici hlavy, rozpoznávat akce obličeje a také odhadovat směr pohledu člověka. OpenFace umožňuje uživateli použít již hotové modely, ale i trénovat si své vlastní. OpenFace byl vyvinut na univerzitě v Cambridge, přičemž hlavní vývojář byl Tadas Baltrušaitis. Sada nástrojů se stále vylepšuje v CMU MultiComp Lab, přičemž původní vývojář Tadas Baltrušaitis se na vývoji aktivně podílí. [19]

Pro rozpoznávání emocí by bylo možné využít funkcionalitu rozpoznávání akcí obličeje. Jedná se o stejnou reprezentaci výrazů obličeje, jako využívá systém FACS, více informací lze nalézt v kapitole 1.2.1 – reprezentace pomocí FACS. OpenFace umožňuje rozpoznávat jak samotný výskyt akce obličeje, tak její intenzitu. Seznam akcí obličeje, který dokáže OpenFace detekovat jsou na Obrázek 13. [18][19]

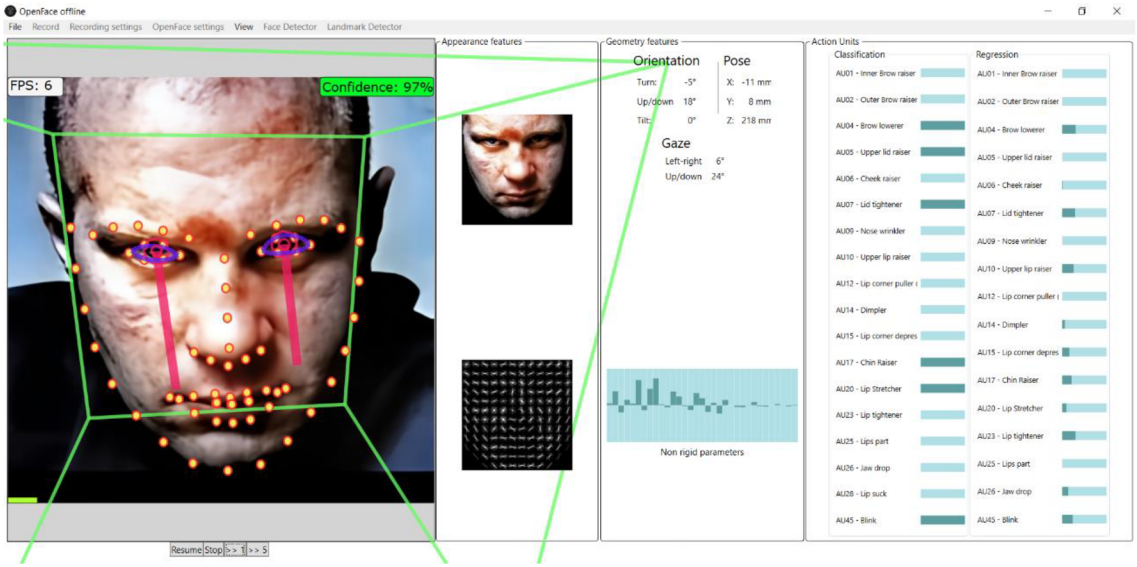
AU	Full name	Illustration
AU1	INNER BROW RAISER	
AU2	OUTER BROW RAISER	
AU4	BROW LOWERER	
AU5	UPPER LID RAISER	
AU6	CHEEK RAISER	
AU7	LID TIGHTENER	
AU9	NOSE WRINKLER	
AU10	UPPER LIP RAISER	
AU12	LIP CORNER PULLER	
AU14	DIMPLER	
AU15	LIP CORNER DEPRESSOR	
AU17	CHIN RAISER	
AU20	LIP STRETCHED	
AU23	LIP TIGHTENER	
AU25	LIPS PART	
AU26	JAW DROP	
AU28	LIP SUCK	
AU45	BLINK	

Obrázek 13 - Seznam akcí obličeje, které OpenFace dokáže rozpoznávat [18]

OpenFace neobsahuje žádnou knihovní funkci, kterou by bylo možné použít pro analýzu emoce na obrázku, nicméně pro vyhodnocení detekovaných AU by bylo možné využít Tabulka 2. Případně by pro klasifikaci emoce mohl být také použit jednoduchý klasifikátor, například rozhodovací strom. Výstupní formát dat OpenFace je CSV soubor.

Ten obsahuje informace o akcích obličeje ve formě dvou parametrů AUxx_r, který definuje intenzitu detekované akce obličeje v rozsahu 0-5 a AUxx_c, který nabývá hodnot 0 nebo 1, a definuje, zda je akce obličeje detekována. Označení parametrů xx je číslo dané akce obličeje, například AU01_r.

OpenFace nabízí grafické uživatelské rozhraní (GUI) ve formě Windows desktopové aplikace. Ta je vhodná pro rychlou analýzu vstupních dat, přičemž vstupními daty může být obrázek, sekvence obrázků, video či záznam z webkamery. Výběr vstupních dat je možné jednoduše v GUI nastavit. Rovněž lze nastavit možnosti nahrávání (volba parametrů, které se budou zapisovat do CSV souboru), zobrazování výsledků a volit mezi třemi druhy detektorů tváře (Haar, Hog-SVM, MTCNN) a třemi druhy detektorů orientačních bodů (CLM, CLNF, CE-CLM). Analyzovaný obličej v OpenFace lze vidět na následujícím obrázku.



Obrázek 14 - Ukázka OpenFace GUI

OpenFace uživatelské rozhraní se skládá ze čtyř sekcí. V sekci nejvíce nalevo je zobrazen analyzovaný obličej včetně detekovaných orientačních bodů (oranžové kroužky), polohy očí, směru pohledu (červené čáry) a 3D natočení obličeje (zelené čáry). V této sekci jsou také zobrazeny informace o počtu analyzovaných snímků za sekundu a důvěryhodnost pozic detekovaných orientačních bodů obličeje (confidence). V druhé sekci je zobrazen oříznutý obličej s černým pozadím a pod ním je zobrazen detekovaný obličej pomocí metody HOG (Histogram of Oriented Gradients). Další sekce je pro zobrazení orientace a pózy obličeje včetně informací o směru pohledu v číselných hodnotách. V poslední sekci jsou zobrazovány detekované akce obličeje a jejich intenzity.

Na Obrázek 14 lze vidět analýzu obličeje s emocí zlost. Detektor obličeje v tomto případě funguje velice dobře. Obrázek je oříznut zdola okolo brady a shora přibližně v polovině čela. Stejně tak detekce orientačních bodů obličeje je velice přesná. Emoce zlost by dle teorie FACS mohla být kombinací akcí obličeje zobrazených v levé části

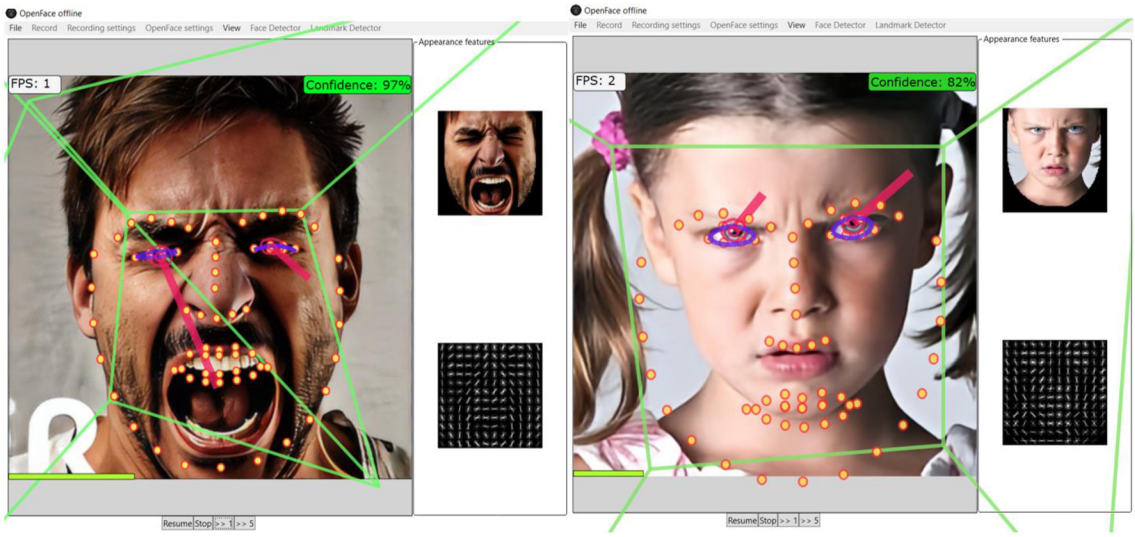
Tabulka 3, v pravé části tabulky jsou naopak zaznamenány detekované AU pomocí OpenFace.

Tabulka 3 - Kvalita detekce AU pomocí OpenFace

Možné kombinace	Detekovaná kombinace
4+5+7+10+22+23+25/26	AUc: 4, 5, 7, 17, 20, ~~45~~ AUr: 4, 7, 10, 17, 23
4+5+7+10+23+25/26	
4+5+7+17+23/24	
4+5/7	
17+24	

Detekovaná akce obličeje s číslem 45, tedy akce obličeje znázorňující mrknutí je chybná. Jedná se o statický obrázek, kde k mrknutí nedochází. Sami autoři OpenFace uvádějí, že na statických obrazcích není detekce akcí obličeje tak přesná, jako detekce z videí, kde je zaznamenána pouze jedna osoba [19]. Přesto detekované akce obličeje (AUc) a jejich intenzity (AUr) poměrně dobře odpovídají možným kombinacím emoce zlost. Detekovaná AU45 nebude uvažována, jelikož se jedná o chybnou detekci, a navíc se AU45 ani nevyskytuje v žádné z kombinací výrazů obličeje pro vyjádření hledaných emocí. Akce obličeje 20 je detekována, ale nevyskytuje se v žádné z možných kombinací, je tedy označena jako nesprávná. Zbylé detekované AU se vyskytují alespoň v jedné z možných kombinací emoce zlost, jsou tedy označeny jako správně detekované.

Analýza pomocí OpenFace však není bezchybná, na následujícím obrázku jsou vidět situace, kdy byla analýza obličeje nesprávná. V levé části je vidět obrázek s dobře detekovaným a oříznutým obličejem, nicméně pozice orientačních bodů detekující ústa jsou nesprávné. Přestože má člověk na obrázku doširoka otevřená ústa, OpenFace detekuje správně pouze horní část rtů a jako dolní ret nesprávně označí oblast. kde končí přední zuby. Ústa detekovaná pomocí OpenFace se tedy jeví jako zavřená, a tím pádem jsou chybně detekovány i akce obličeje s ústy spojené. V druhém případě je ukázka chybně detekovaného obličeje dívky. Z tohoto důvodu jsou i pozice orientačních bodů obličeje určeny nesprávně. Druhý případ nastává spíše ojediněle, OpenFace v drtivé většině případů obličej ořízne správně.



Obrázek 15 - Ukázky chybné analýzy OpenFace

U statických obrázků lze také vidět, že detekce natočení obličeje, je také poměrně nepřesná. Natočení obličeje má být znázorněno zelenými čarami, které mají tvořit krychli, která svým natočením odpovídá natočení obličeje. U výše uvedených obrázků však čáry vystupují z obrazu směrem ven a o krychli nelze moc hovořit. Detekce natočení z videí, nebo ze záznamu webkamery je o poznání přesnější.

iMiGUE

Je databáze, kterou vytvořila finská Univerzita v Oulu. Databáze je stále ve vývoji, pracuje na ní Centrum pro počítačové vidění a signálovou analýzu na Fakultě informačních technologií a elektrického inženýrství. Zkratka iMiGUE stojí za „identity-free video dataset for Micro-Gesture Understanding and Emotion analysis. Jedná se tedy o cenzurovanou databázi videí určenou pro studium emocí na základě pozorování mikro-gest. V současné době databáze obsahuje 359 odkazů na veřejně dostupná videa z YouTube kanálu Australian Open TV. Videa zachycují pozápasové rozhovory s hráči, přičemž je vždy zaznamenána informace, zda hráč daný zápas vyhrál nebo prohrál. Celkem databáze obsahuje 72 subjektů, kteří projevují různá mikro-gesta. Celkový počet 18 499 MG získaných z videosnímků je manuálně klasifikován a anotován do 32 tříd. Autoři detekovaná MG dělí dále ještě na pět podskupin, podle oblasti, ve které jsou pozorována. Podskupiny s příkladem mikro-gesta uvedeným v závorce jsou následující: tělo (vzpřímený sed), hlava (hlava vzpřímená), ruka (zkřížení prstů na rukou), tělo-ruka (poškrábání se na zádech), hlava-ruka (promnutí očí). Další informace o datasetu lze vidět v Tabulka 4. [15][16][17]

Tabulka 4 - Statistiky a vlastnosti databáze iMiGUE [17]

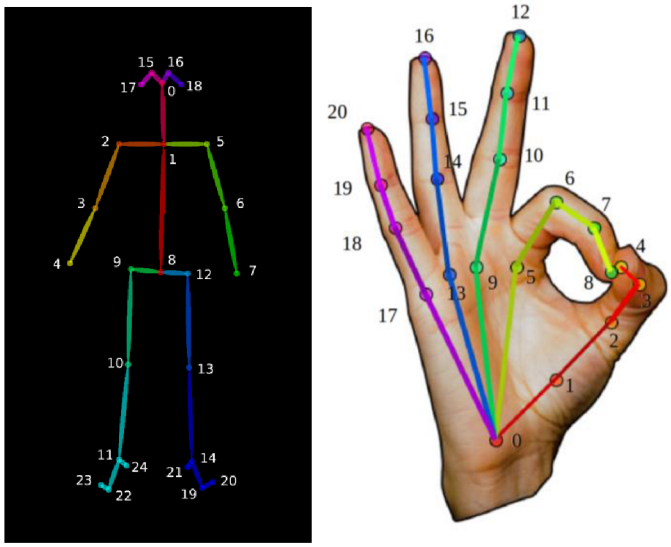
Třídy emocí	2 (Pozitivní/Negativní)
Třídy mikro-gest	32
Subjekty (ženy/muži)	72 (36/36)
Státy a regiony	28
Počet videí	359 (258 výher, 101 proher)
Počet vzorků mikro-gest	18 499
Rozlišení	1280 x 720
FPS	25
Průměrná délka snímku	2,5 sekund
Celková délka	2 092 minut
Počet anotátorů	5
Bio-info	Žádné
Způsob rozpoznávání	Holistický

Z Tabulka 4 lze vidět, že emoce jsou rozděleny do dvou tříd – pozitivní a negativní. Emoce sledovaného subjektu je manuálně anotována podle výsledku zápasu, předpokládá se, že po vyhraném zápase bude osoba projevovat pozitivní emoci a naopak. V databázi je mimo emoce anotován i časový údaj a mikro-gesto, které se ve snímku v daném čase nachází. [15]

Pro analýzu gest a pozic těla se využívá toolbox OpenPose. Podrobnější informace ohledně OpenPose budou uvedeny dále v textu. Pro rozpoznávání emocí z MG jsou pak naučeny modely, a to jak metodou učení s učitelem nebo bez učitele. Naučený model dosahuje přesnosti až 91,24 % na validační množině. Bohužel zatím není databáze z důvodu ochrany osobních údajů (GDPR) plně dostupná. Využití detekování mikro-gest za použitím toolboxu OpenPose však může být pro práci užitečné. [15]

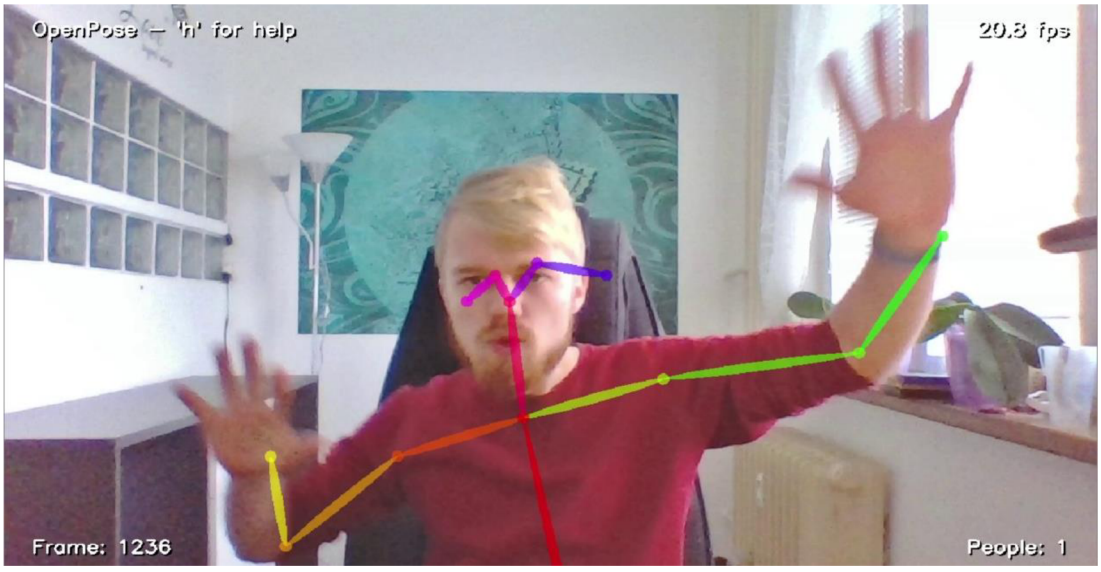
OpenPose

Je toolbox, který umožňuje v reálném čase detekovat pozice bodů lidského skeletu, rukou a také obličeje. OpenPose umožňuje detekovat hledané body u více osob najednou v jednom snímku, celkem je možné s ním sledovat až 137 různých bodů pro každou osobu. Sledovaný bod je častokrát lidský kloub, přičemž OpenPose umožňuje vykreslovat i spojnice mezi jednotlivými body. Spojnice mezi jednotlivými body reprezentují kosti člověka, vykreslením spojnic je tedy graficky znázorněn skelet sledované osoby a tímto způsobem lze analyzovat pohyb člověka. Příklad sledovaných bodů lze vidět na Obrázek 16. Počet analyzovaných bodů závisí na zvoleném modelu. S OpenPose lze detekovat body těla (klouby, hlava, zápěstí), body na obličeji (podobně jako to umí OpenFace), nebo jednotlivé klouby na rukou. Každý bod obsahuje tři údaje, a to pozici x-ové a y-nové souřadnice a její věrohodnost. Výstupní data jsou mimo jejich grafickou reprezentaci zapisována do textového souboru ve formátu JSON. Příklady detekovaných bodů lze vidět na následujícím obrázku. [15]



Obrázek 16 - OpenPose sledované body skeletu a ruky [29]

Bohužel je OpenPose poměrně náročný na výkon grafické karty. V rámci práce byl OpenPose testován na notebooku, který disponuje grafickou kartou GeForce GTX 1050. Ta bohužel nebyla dostatečná pro analýzu obrazu v plném rozsahu a uspokojivých výsledků bylo dosaženo pouze při detekci skeletu člověka. Vhodným nastavením toolboxu bylo dosaženo detekování jednotlivých kloubů člověka v reálném čase s dostačující přesností a frekvencí okolo 20 snímků za sekundu. Pro testovací účely byl použit záznam z webkamery. Výsledek lze pozorovat na Obrázek 17. Na úkor rychlosti byla omezena přesnost vyhodnocení toolboxu OpenPose.



Obrázek 17 - OpenPose - test na datech z webkamery

Na obrázku lze tedy pozorovat, že poloha očí a uší je vyhodnocena poměrně nepřesně (znázorněno fialovou a růžovou barvou). U těchto sledovaných bodů dochází k výrazným odchylkám oproti reálné pozici a z tohoto důvodu nebudou pro analýzu emoce člověka

použity. Naopak poloha kloubů horních končetin a trupu je vyhodnocená relativně přesně. Dochází tam k drobným odchylkám, nicméně je třeba uvažovat, že byl vyhodnocován dynamický obraz. Snímaný člověk je tedy v pohybu a drobné odchylky od reálné pozice detekovaného bodu jsou akceptovatelné.

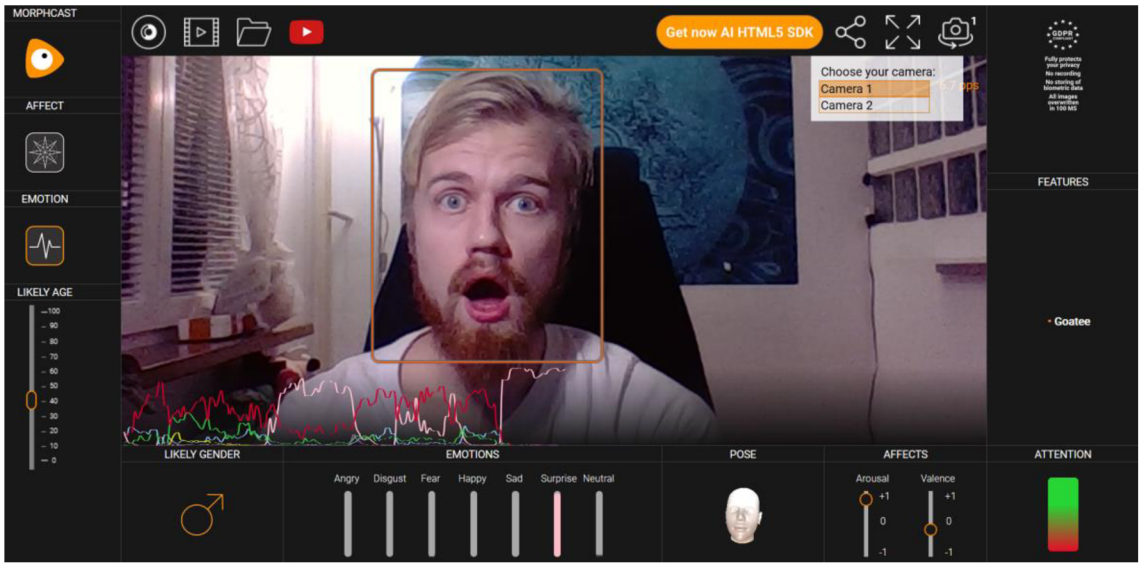
Bohužel v rámci dostupného hardwaru nebylo možné otestovat detekci jednotlivých kloubů rukou a bodů obličeje. Částečných výsledků bylo dosaženo pouze pro data ve formě statického obrazu a výsledky nebyly nijak uspokojivé. Pro data z webkamery pak analýza nebyla možná provést vůbec, a to z důvodu nedostatečné grafické paměti.

OpenPose se jeví jako užitečný nástroj, který by v rámci práce mohl být využit. Díky toolboxu lze získávat data o pozici horních končetin, která mohou být použita pro klasifikaci výsledné emoce. Příkladem mohou být například emoce strach a překvapení, při kterých se ruce častokrát nacházejí v oblasti obličeje, viz obrázek.

Obrázek 18 - Ruce v oblasti obličeje při projevení emoce [22]

Morphcast

Je sada produktů pro analýzu emocí vyvíjená italskou společností Cynny. Jedná se o komerční produkty, jako je například platforma pro tvorbu interaktivních videí, JavaScript engine nebo webové aplikace pro emoční analýzu. Morphcast umožňuje zákazníkům, kteří si předplatí jejich služby tvořit interaktivní videa, která jsou pak použita pro analýzu emocí cílové skupiny lidí. V praxi to funguje tak, že se člověk z cílené skupiny dívá na vytvořené video, zatímco ho snímá kamera. Interaktivní video dává člověku možnost volby jako je anketa, výběr kategorie, otázka ano/ne a další. Na základě zvolené možnosti se pak odvíjí scénář daného videa. Na pozadí je pak uživatel analyzován JavaScript enginem pro emoční analýzu. Kýžená data jsou pak zpracovávána v reálném čase a odeslána zákazníkovi pro jeho další zpracování. Morphcast zadarmo provozuje demo pro emoční analýzu, které je vyobrazeno na obrázku níže.



Obrázek 19 - Morphcast demo [28]

Demo umožňuje zpracovávat videa v reálném čase z webkamery a vyhodnocená data zobrazuje pomocí příslušných indikátorů. Morphcast software umí rozpoznávat sedm druhů emocí, detekovat naklonění hlavy, pozornost uživatele, určovat pravděpodobné pohlaví či věk. Bohužel ve verzi zdarma neumožňuje analyzovat uživatelem zvolená data (obrázky, nebo videa na osobním uložišti). Mimo vykreslování grafu v reálném čase Morphcast demo neumožňuje žádné jiné zaznamenávání dat, například do souboru. Tyto možnosti zůstávají placené.

Morphcast studio v minulosti spolupracovalo na mnoha projektech spojených s emoční analýzou. Pyšní se spolupracemi se světovými firmami jako je například CocaCola, Lexus či Yahoo. [11]

Další nástroje

Rozpoznávání emocí patří mezi populární oblasti počítačového vidění a nástroje pro jeho realizaci jsou neustále vyvíjeny. Mezi projekty, které je možné veřejně nalézt na GitHubu, ale nebyly v rámci práce nijak dále testovány a zkoumány, lze zařadit například Human [30] , Emotion detection [31], Multimodal Emotion Recognition [32] a nespočet dalších.

1.4 Datasety pro emoční rozpoznávání

Pro emoční rozpoznávání byla vytvořena řada databází obrázků, přičemž některé z nich jsou veřejně dostupné. Obecným problémem pro tyto veřejně dostupné datasety je kvalita jejich anotace. V mnoha případech byla zaznamenána sporně nebo dokonce špatně anotovaná data. Tvůrci datasetu uvádějí mnohdy několik různých anotátorů, toto může být problém, protože každý člověk může vyhodnotit emoci na obrázku jiným způsobem. Kapitola popisuje některé z vybraných datasetů, které budou v rámci práce používány, uvádí informace o rozlišení, velikosti, počtu emočních tříd atd.

FER2013

Je databáze obrázků určená pro použití v oblasti rozpoznávání emocí obličeje. Data jsou tvořena šedotónovými obrázky o velikosti 48x48 pixelů. Obličeje jsou již vycentrovány, každý obličej tedy zabírá v obrázku přibližně stejnou oblast. Databáze obsahuje okolo 30 000 obrázků, které jsou roztříděny do sedmi skupin, které odpovídají projevované emoci. Skupiny jsou angry (zlost), disgust (znechucení), fear (strach), happy (štěstí), sad (smutek), surprise (překvapení) a neutral (neutrální). Dataset byl vytvořen pány Pierre-Luc Carrier a Aaron Courville, jako součást jejich výzkumného projektu. Zkratka FER vychází z anglického „Facial Expression Recognition". Dataset byl mimo jiné využit i pro soutěž vyhlášenou komunitou Kaggle, která proběhla v roce 2013 [21]. Následující koláž představuje ukázku obrázků datasetu FER2013. Každý sloupec představuje emoční třídu. První sloupec představuje emoční třídu zlost, následuje znechucení, strach, štěstí, neutrální výraz, smutek a překvapení.

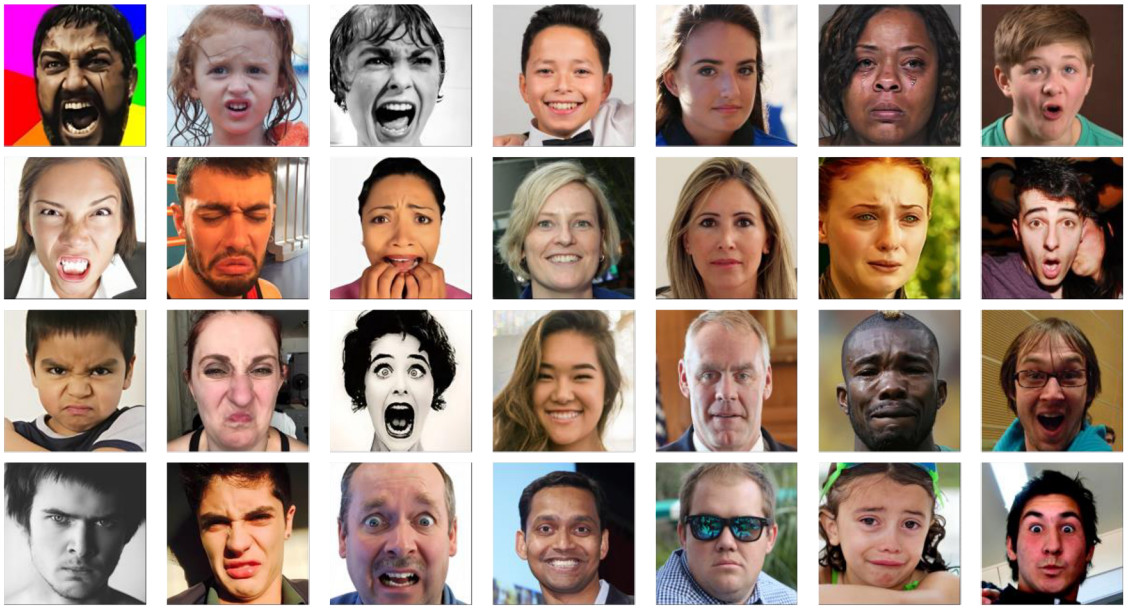
Obrázek 20 - Ukázka datasetu FER2013

AffectNet

Je rozsáhlá databáze barevných obrázků zobrazujících emoce obličeje. Je tvořena daty o různých rozměrech, jedná se však o obrázky s podstatně vyšším rozlišením, než je tomu u datasetu FER2013. Typický rozměr obrázku v databázi AffectNet je 512x512 pixelů. Databáze se skládá z přibližně jednoho milionu obrázků, přičemž 440 tisíc jich bylo manuálně anotováno a rozřazeno do 8 emočních tříd. AffectNet obsahuje stejné emoční třídy jako dataset FER2013, přičemž navíc je rozšířen o třídu emoce contempt (pohrdání). Jedná se pravděpodobně o největší databázi obličejových emocí, který byla vytvořena.

Data o velikosti 120 GB jsou pod správou akademických pracovníků na Univerzitě Denver. [22]

Bohužel jsou data často nepřesně anotována. Například lidé na obrázcích označených jako „zlost“ se usmívají nebo projevují jinou emoci. To stejné platí i pro jiné emoční třídy. Rozsah dat je i problém z hlediska výpočetní náročnosti. Z těchto důvodů byl vybrán pouze vzorek dat z databáze, který byl přezkoumán a nehodící se data byla smazána. Data zařazená ve špatných kategoriích byla odstraněna, stejně tak jako obrázky, kdy obličej byl zachycen z profilu nebo jiným způsobem nevyhovoval. Jako vzorek databáze AffectNet byl použit dataset s 31 tisíci obrázky o velikosti přibližně 9 GB [23]. Vzhledem k časové náročnosti manuální anotace dat bylo pro každou třídu extrahováno pouze 250-300 obrázků. Ty byly pak rozděleny na testovací množinu - 50 obrázků a trénovací množinu cca 200 obrázků. Následující koláž představuje ukázku obrázku datasetu AffecNet. Každý sloupec opět odpovídá jedné emoční třídě a je zachováno stejné pořadí jako u koláže FER2013.



Obrázek 21 - Ukázka datasetu AffectNet

2. VÝBĚR VHODNÉ METODY

V rámci rešerše byly nastíněny metody, jak by bylo možné systém pro rozpoznávání emocí realizovat a také byly popsány datasety, které by bylo možné využít. V následující kapitole jsou popsány výsledky zmíněných metod a jejich modifikací na datasetech FER2013 a AffectNet. Výsledky kvality klasifikace jsou prezentovány maticemi záměn.

2.1 Vyhodnocování

Jako nástroj pro vyhodnocení správnosti klasifikace dané metody byla použita matice záměn. Ta lze obecně použít jak pro binární klasifikaci, kde jsou výstupem pouze dvě třídy, tak pro více třídovou klasifikaci, kde počet výstupních tříd není teoreticky ničím omezen. V rámci práce je uvažována více třídová klasifikace, kde je uvažováno sedm výstupních tříd. Konkrétně se jedná o zlost, znechucení, strach, štěstí, neutrální emoci, smutek a překvapení.

Matice záměn je vhodná především pro vyhodnocení predikcí modelu u dat, která již máme anotována, tedy známe jejich skutečnou výstupní třídu. Informace o počtech predikcí a skutečných hodnot pro dané třídy jsou pak vhodně zapisovány do sloupců a řádků matice záměn. Příklad lze vidět v Tabulka 7, kde v řádku „(6) – Celkem“ je uveden celkový počet anotovaných snímků pro danou třídu. Pro emoční třídu zlost je dle matice záměn dostupných 940 snímků. Obdobně by byly zapsány i počty pro ostatní emoční třídy, pro přehlednost však nejsou v ukázkové matici záměn další počty uvedeny. Naopak ve sloupci „(3) – Celkem“ je uveden počet predikcí pro danou emoční třídu. V případě Tabulka 7 má celkový počet dat predikovaných jako třída zlost hodnotu 861.

U matice záměn lze pro predikované a skutečné hodnoty najít právě čtyři možné kombinace. Tedy predikovaná hodnota správně pozitivní (True Positive – TP), nesprávně pozitivní (False Negative – FN), nesprávně negativní (False Negative – FN) a správně negativní (True Negative – TN). Tabulka 5 zobrazuje nejjednodušší případ, tedy matici záměn pro dvě výstupní třídy. Z hodnot TP, FP, FN a TN jsou pak dále počítány další parametry, které jsou používány pro hodnocení klasifikace. Mezi tyto parametry patří přesnost (precision), senzitivita (recall), správnost (accuracy) a F1 míra (F Score).

Tabulka 5 - Matice záměn pro dvě třídy

		Skutečnost	
		Třída 1	Třída 2
Predikce	Třída 1	TP	FP
	Třída 2	FN	TN

Pro více třídovou klasifikaci pak vypadá matice záměn poněkud komplikovaněji. Hodnoty TP se nacházejí vždy na hlavní diagonále tabulky, stejně jako u matice záměn pro dvě třídy se jedná o jedno číslo v tabulce, které vyjadřuje počet správně predikovaných hodnot pro danou třídu [39]. Tyto hodnoty jsou v dále v maticích záměn podbarveny šedou barvou, viz Tabulka 7. Naopak další hodnoty, tedy FP, FN a TN, už jsou dány součtem více buněk tabulky. Pro více třídovou klasifikaci platí, že výpočet parametrů bude odlišný pro každou třídu. Bude tedy potřeba vypočítat parametry TP, FP, FN, a TN zvlášť pro každou třídu. Příklad výpočtu TP, FP, FN a TN pro první třídu je znázorněn na Tabulka 6.

Tabulka 6 - Matice záměn pro N tříd

		Skutečnost			
		Třída 1	Třída 2	…	Třída N
Predikce	Třída 1	TP		FP	
	Třída 2				
	⋮	FN		TN	
	Třída N				

Daný parametr je vždy vypočítán jako sumace všech čísel v buňkách ohraničených v rámečku. Parametr FN je v tomto případě vypočítán jako suma všech nesprávně predikovaných hodnot, tedy hodnoty ve sloupci „Třída 1“, které jsou predikovány jako „Třída 2“ - „Třída N“. Analogicky jsou vypočítány i parametry FP a TN.

Na tabulce níže je zobrazena ukázka matice záměn s používanými pojmy. Tabulka 8 obsahuje ve sloupcích informaci o skutečném počtu obrázků v dané třídě (anotované dle datasetu FER2013). V řádcích pak informuje o počtu obrázků s predikovanou emocí. Například pro emoci „zlost“ je v datasetu FER2013 dostupných 940 obrázků, přičemž správně klasifikovaných bylo 734. Jako nesprávně byly predikovány 4 obrázky jako znechucení, 59 obrázků jako strach, 22 obrázků jako štěstí, 61 jako neutrální, 72 jako smutek a 8 jako překvapení. Všechny tyto obrázky ale ve skutečnosti zachycovaly emoci zlost. Naopak 127 obrázků bylo nesprávně predikováno jako emoce zlosti, přitom ve skutečnosti se jednalo o jinou emoční třídu. Konkrétně tedy o 7 obrázků s reálnou emocí znechucení, 40 s emocí strach, 14 s emocí štěstí, 20 s emocí neutrální, 41 s emocí smutek a 5 s emocí překvapení. Ostatní parametry matice záměn, mezi které patří přesnost, senzitivita, správnost a F1 míra jsou vypočteny dle níže zmíněných vzorců.

Tabulka 7 - Ukázka matice záměn s používanými pojmy

(1) Celková správnost		(2) Skutečnost							(3) Celkem	(4) Přesnost
		Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení		
(5) Predikce	Zlost	734	7	40	14	20	41	5	861	85 %
	Znechucení	4	89	2	2	1	2	0		
	Strach	59	7	822	49	41	70	43		
	Štěstí	22	1	9	1679	48	27	24		
	Neutrální	61	2	53	57	1026	117	19		
	Smutek	72	5	65	19	74	877	6		
	Překvapení	8	0	27	11	6	5	700		
	(6) Celkem	940								
	(7) Senzitivita	78 %								
	(8) Správnost	95 %								
	(9) F1 míra	81 %								

Legenda [39]:

(1) Celková správnost

- Vyjadřuje procentuální správnost predikce všech tříd
- Její hodnota určuje, kolik procent ze všech hodnot bylo správně predikováno

$$Celková\ správnost = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + FN_i} \tag{2.1}$$

(2) Skutečnost

- Sekce pro informace o skutečných počtech dat

(3) Celkem

- Udává celkový počet predikcí pro danou třídu

(4) Přesnost

- Vyjadřuje v procentech přesnost predikce
- Její hodnota určuje, kolik procent dat z celkového počtu predikcí bylo správně predikovaných

$$Přesnost = \frac{TP}{TP + FP} \times 100 \tag{2.2}$$

(5) Predikce
- sekce pro informace o predikovaných počtech dat

(6) Celkem
- Udává celkový počet dat v dané třídě

(7) Senzitivita
- Vyjadřuje, jak je schopen klasifikátor danou emoční třídu rozpoznat, tedy jak je na ní „senzitivní“
- Její hodnota určuje, kolik procent dat z dané emoční třídy klasifikátor rozpoznal

$$Senzitivita = \frac{TP}{TP + FN} \times 100 \tag{2.3}$$

(8) Správnost
- Vyjadřuje procentuální hodnotu správného rozpoznání, zda se jedná o danou třídu nebo se jedná o jinou třídu

$$Správnost = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.4}$$

(9) F1 míra
- Počítá se z hodnot přesnosti a senzitivity

$$F1\ míra = \frac{2 \times Přesnost \times Senzitivita}{Přesnost + Senzitivita} \tag{2.5}$$

2.2 DeepFace

DeepFace knihovna obsahuje naučený emoční model, který byl pro vyhodnocení kvality klasifikace použit. Bylo zjištěno, že klasifikace tohoto modelu silně závisí na datech, která byla použita autorem pro naučení emočního modelu. Pro lepší výsledky by bylo možné DeepFace řešení modifikovat a naučit emoční model na jiných datech. V kapitole jsou výsledky klasifikace DeepFace modelu na datasetech FER2013 a AffectNet.

DeepFace model a dataset FER2013

V programovacím jazyce Python byl vytvořen skript, který umožňuje postupně zpracovávat vstupní data, která jsou uživatelem předložena. Skript zpracovává postupně všechny obrázky obsažené ve složce, jejíž cestu uživatel zvolí pomocí proměnné *path*. S využitím emočního modelu DeepFace a knihovní funkce *DeepFace.analyze()* je pro daný obrázek vyhodnocena dominantní emoce. Funkce *DeepFace.analyze()* dokáže rozpoznat všech sedm typů emocí, které dataset FER2013 obsahuje. Skript byl použit pro všech sedm tříd testovacího datasetu. Celkem validační množina obsahuje 6315 fotografií. Výsledky úspěšnosti klasifikace emoce jsou vyjádřeny maticí záměn, kterou lze vidět níže.

Pro každou třídu je vypočtena hodnota senzitivity (recall), přesnosti (precision), správnosti (accuracy) a F1 míry. Emoční model vykazuje poměrně konzistentní senzitivitu napříč všemi třídami. Hodnota průměrné senzitivity je okolo 83 % přičemž nejvyšších hodnot dosahuje senzitivita pro třídy překvapení (88 %) a štěstí (92 %). Stejně tak přesnost klasifikace dosahuje u tříd podobných hodnot, průměrně 84 %. Nejpřesněji model klasifikuje emoci překvapení (92 %) a štěstí (93 %). Správnost dosahuje u všech tříd hodnotu nad 92 % a nejsprávněji je klasifikována emoce znechucení (99,53 %). F1 míra také nevykazuje znatelné rozdíly mezi klasifikací jednotlivých tříd, hodnota, F1 míry je také poměrně konzistentní. Nejvyšší F1 míry dosahuje opět třída překvapení (90 %) a štěstí (92 %). Celková správnost klasifikace emočního modelu DeepFace na datasetu FER2013 je 83,81 %.

Tabulka 8 - Matice záměn, DeepFace + FER2013

Celková správnost 83,81 %		**Skutečnost**								
		Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení	Celkem	Přesnost
Predikce	Zlost	734	7	40	14	20	41	5	861	85 %
	Znechucení	4	89	2	2	1	2	0	100	89 %
	Strach	59	7	822	49	41	70	43	1091	75 %
	Štěstí	22	1	9	1679	48	27	24	1810	93 %
	Neutrální	61	2	53	57	1026	117	19	1335	77 %
	Smutek	72	5	65	19	74	877	6	1118	78 %
	Překvapení	8	0	27	11	6	5	700	757	92 %
	Celkem	940	111	1018	1825	1216	1139	797		
	Senzitivita	78 %	80 %	81 %	92 %	84 %	77 %	88 %		
	Správnost	95 %	99 %	93 %	96 %	92 %	93 %	98 %		
	F1 míra	81 %	84 %	78 %	92 %	80 %	78 %	90 %		

DeepFace model a dataset AffectNet

Emoční model DeepFace dosahoval při analýze obrázků z datasetu FER2013 nad očekávání dobrých výsledků. Hodnoty senzitivity, přesnosti, správnosti a F1 míry byly konzistentní napříč třídami a celková správnost klasifikace byla 83,81 %. Pro potvrzení kvality klasifikace byl DeepFace model otestován ještě na jiném datasetu – AffectNet. Pro vyhodnocení dat byl opět využit skript napsaný v jazyce python, který je blíže popsaný v předcházející kapitole. Výsledky jsou zaznamenány v matici záměn níže.

Vypočtené hodnoty senzitivity, přesnosti, správnosti a F1 míry dosahují mnohem nižších hodnot, než tomu bylo u datasetu FER2013. Průměrná senzitivita je 42 %, což je přibližně poloviční hodnota, než jaké bylo dosaženo u datasetu FER2013. Senzitivita navíc vykazuje značné výkyvy napříč třídami. Nejvyšší hodnoty senzitivity bylo opět dosaženo pro emoční třídu štěstí (80 %), naopak nejnižší hodnota byla pro třídu znechucení a to pouze 12 %. I přes velice nízkou hodnotu senzitivity však model třídu znechucení klasifikoval poměrně přesně. Konkrétně tedy s přesností 75 %, což je nejvyšší hodnota přesnosti v dané matici záměn. Model tedy velice zřídka emoci znechucení rozpoznal, ale když už se tak stalo, pravděpodobnost správné klasifikace byla poměrně vysoká. Nejméně přesně model vyhodnotil emoci smutek, dosáhl přesnosti pouhých 29 %. Průměrná přesnost modelu byla pouze 46 %, což je opět o polovinu méně, než tomu bylo u datasetu FER2013. Správnost klasifikace byla však poměrně konzistentní, průměrná hodnota nabývala zhruba 83 %. Ani v jednom případě však nebylo dosaženo správnosti vyšší než 90 %. Hodnota F1 míry se pro různé třídy značně lišila. Nejvyšší F1 míry bylo dosaženo pro emoci štěstí (63 %) a nejnižší pro emoci znechucení (21 %). Celková správnost klasifikace emočního modelu DeepFace na datasetu AffectNet byla pouze 42 %.

Tabulka 9 - Matice záměn, DeepFace a AffectNet

Celková správnost 42 %		**Skutečnost**								
		Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení	Celkem	Přesnost
Predikce	Zlost	22	14	3	1	5	3	5	53	42 %
	Znechucení	2	6	0	0	0	0	0	8	75 %
	Strach	9	4	18	3	4	9	6	53	34 %
	Štěstí	4	8	10	40	2	10	3	77	52 %
	Neutrální	6	10	4	4	26	12	11	73	36 %
	Smutek	6	8	3	2	13	15	5	52	29 %
	Překvapení	1	0	12	0	0	1	20	34	59 %
	Celkem	50	50	50	50	50	50	50		
	Senzitivita	44 %	12 %	36 %	80 %	52 %	30 %	40 %		
	Správnost	83 %	87 %	81 %	87 %	80 %	80 %	86 %		
	F1 míra	43 %	21 %	35 %	63 %	42 %	29 %	48 %		

Výsledek klasifikace emoce pomocí modelu DeepFace byl pro dataset AffectNet mnohem horší než pro dataset FER2013. Tento výsledek ukazoval, že kvalita klasifikace

pomocí modelu DeepFace nějakým způsoben závisí na vstupních datech. Později bylo odhaleno, že autor DeepFace použil jako trénovací množinu právě část datasetu FER2013. [26]

Vysvětlení, proč pro tyto dva datasety došlo k takto rozdílným výsledkům, mohou být následující. Modelu byla předložena data, která byla autorem použita jako trénovací. Druhé vysvětlení je, že je daný model na dataset FER2013 přeučen. Třetí možností vysvětlení je v odlišnosti obou datasetů, obrázky v datasetech mají totiž velice rozdílné rozlišení. Ani jedno z možných vysvětlení nebylo v práci dále zkoumáno. Výsledek klasifikace modelu DeepFace pro dataset FER2013 byl vyhodnocen jako zavádějící a pro další testování kvality klasifikace bude využíván dataset AffectNet. Ten na rozdíl od datasetu FER2013, který má rozlišení pouze 48x48 pixelů, lépe odpovídá vstupům, které pro systém rozpoznávání emocí budou použity, například snímky z webkamery.

2.3 OpenFace

Pro určení emocí byla využita detekce akcí obličeje. OpenFace bohužel neumí rozpoznat všechny typy akcí obličeje, které jsou nutné pro určení emoce - Tabulka 2. Chybí detekce AU11, AU22, AU24, AU27 a AU50. Z tohoto důvodu nebyl pro vyhodnocení emoce použit FACS, ale byl navržen jednoduchý klasifikátor ve formě rozhodovacího stromu.

OpenFace model a dataset FER2013

Pro každou ze sedmi emočních tříd datasetu FER2013 byla provedena analýza fotek obličejů, s cílem určit akce obličeje. Bohužel výsledky nebyly příliš dobré a OpenFace měl s analýzou dat problém. OpenFace určuje při analýze obličeje dva parametry kvality, a to confidence a success. Confidence určuje, jaká je jistota odhadu detekce orientačních bodů, což se projevuje i na přesnosti detekce akcí obličeje. Vyjadřuje jistotu v procentech a může nabývat hodnot v rozsahu 0–100 %. Success pak určuje, jak úspěšně analýza proběhla, tedy zda byl obličej detekován a správně zanalyzován a je vyjádřen binární hodnotou 0/1. Průměrná hodnota parametru success byla pro třídu angry pouze 0,47, tedy pouze u 47 % obrázku byl úspěšně detekován obličej. U více než poloviny dat tedy nebylo možné detekovat akce obličeje, a tedy ani určit emoci. Průměrná hodnota parametru confidence se pak pohybovala okolo 48 %. Pro data, kde byla detekce obličeje úspěšná (success byl roven jedné) se parametr confidence pohyboval okolo 87 %. Experimentální metodou bylo zjištěno, že nejlepších výsledku je dosaženo při použití Haarova detektoru obličeje a detektoru orientačních bodů obličeje CE-CLM. Parametr success při této konfiguraci dosáhl 52,5 %, což je stále velice málo. OpenFace tedy pro tento dataset nebylo možné řádně otestovat. [19]

OpenFace model a dataset dataset AffectNet

Analýza obrázku z datasetu AffectNet byla mnohem úspěšnější, než tomu bylo u datasetu FER2013. OpenFace dokázal extrahovat příznaky z více než 80 % dat, u některých emočních tříd dokonce se 100% jistotou (štěstí a neutrální). U některých fotografií měl však problém rozpoznat obličej a segmentoval například pouze část nosu. V tomto případě logicky nemohl ani rozpoznat orientační body nebo akce obličeje. Neúspěšně analyzovaná data byla tedy smazána a trénink a validace byla provedena pouze na datech s úspěšně extrahovanými příznaky. Při zpracování snímků v reálném čase bude k datům přistupováno obdobně, pokud nebude možné extrahovat příznaky, daný snímek nebude vyhodnocován. Vzhledem k frekvenci v řádu desítek snímků za sekundu, by toto v praxi neměl být problém.

Výstupem OpenFace byly dva CSV soubory pro trénovací a testovací množinu. Soubor obsahoval informace o čísle snímku, identifikátoru obličeje, časovém razítku, věrohodnosti detekce bodů obličeje, úspěšnosti extrakce příznaků, detekovaných AU a emoční třídě. Formát dat uložených v souboru lze vidět v tabulce níže.

Tabulka 10 - Formát dat pro OpenFace

Frame	FaceID	TimeStamp	Confidence	Success	AU1	…	AU45	Class
1	0	207.241	0.92	1	1	…	1	Angry
2	0	207.299	0.92	1	0	…	0	Sad
3	0	207.345	0.92	1	1	…	0	Happy
4	0	207.375	0.92	1	1	…	1	Neutral

V programovacím jazyce Python byl s pomocí scikit-learn knihovny napsán kód realizující trénink rozhodovacího stromu na trénovacích datech a vyhodnocení jeho klasifikace na testovacích datech. Knihovna scikit-learn využívá optimalizovaného CART (Classification and Regression Trees) algoritmu a umožňuje nastavit kriteriální funkci, typ dělení, maximální hloubku stromu, minimální počet vzorků v listu stromu atd [24][25]. S pomocí knihovny matplotlib lze vytvořený rozhodovací strom i vykreslit do obrázku. Nejlepších výsledků celkové správnosti klasifikace bylo dosaženo při použití kriteriální funkce výpočtu entropie, maximální hloubce stromu 9 a minimálního počtu vzorků v listu 9. Při této konfiguraci bylo dosaženo celkové správnosti 64,11 % a podrobnější informace o úspěšnosti klasifikace popisuje matice záměn níže.

Tabulka 11 - Matice záměn, OpenFace a AffectNet

Celková správnost 64,11 %		Skutečnost								
		Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení	Celkem	Přesnost
Predikce	Zlost	14	5	1	0	0	0	0	20	70 %
	Znechucení	6	24	1	0	4	2	0	37	65 %
	Strach	0	0	18	0	3	1	6	28	64 %
	Štěstí	5	1	0	50	1	0	0	57	88 %
	Neutrální	2	0	2	0	30	10	2	46	65 %
	Smutek	11	12	6	0	5	23	4	61	38 %
	Překvapení	0	0	6	0	7	0	25	38	66 %
	Celkem	38	42	34	50	50	36	37		
	Senzitivita	37 %	57 %	53 %	100%	60 %	64 %	68 %		
	Správnost	90 %	89 %	91 %	98 %	87 %	82 %	91 %		
	F1 míra	48 %	61 %	58 %	93 %	63 %	47 %	67 %		

Vypočtené hodnoty senzitivity, přesnosti, správnosti a F1 míry dosahují vyšších hodnot, než tomu bylo se stejným datasetem a modelem DeepFace. Nejvyšší senzitivity bylo dosaženo u emoční třídy štěstí, kdy všech 50 vzorků dat bylo klasifikováno správně. Senzitivita tedy v tomto případě byla 100 %. Hodnoty senzitivity pro další třídy byly znatelně horší a mimo třídu zlost se pohybovaly v rozmezí 50-70 %. Hodnoty přesnosti byly napříč třídami vyváženější a průměrná přesnost dosahovala 65 %. Nejvyšší přesnost byla opět dosažena pro třídu štěstí. Stejně tak správnost byla poměrně konzistentní, její hodnoty se pohybovaly v rozmezí 82-98 % a její průměrná hodnota byla 90 %. Vypočtené hodnoty F1 míry se napříč třídami lišily, nejvyšší hodnoty dosahovala třída štěstí s 93 %.

Výsledek klasifikace pomocí detekce akcí obličeje a klasifikace rozhodovacím stromem dosahovala vyšší celkové správnosti, než tomu bylo u stejného datasetu a použití konvoluční neuronové sítě (model DeepFace). Hodnota celkové správnosti byla vyšší o 22 %.

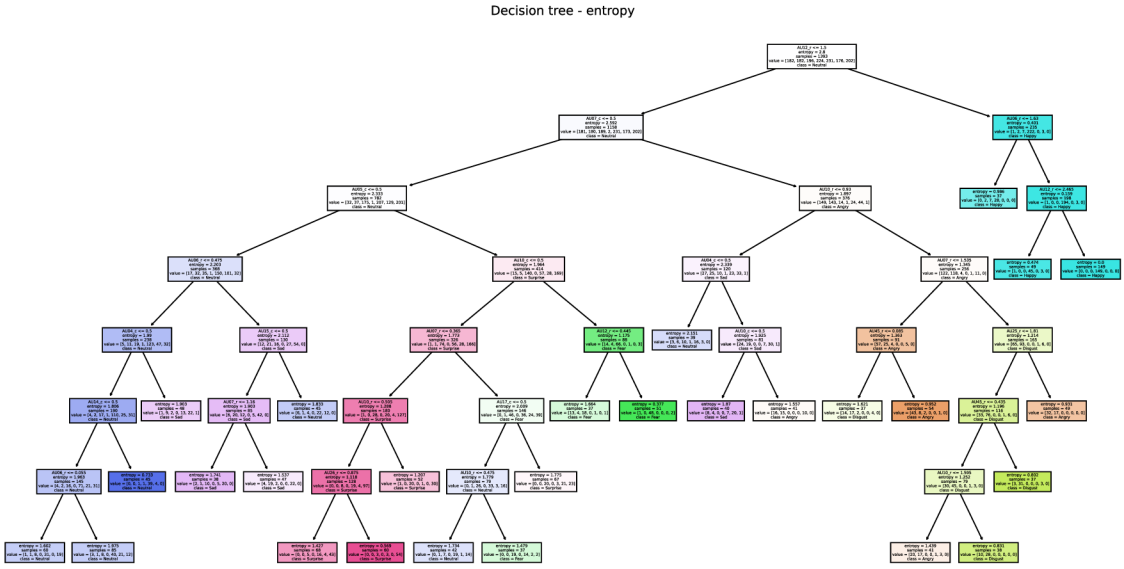
Pro učení výše uvedeného rozhodovacího stromu byly použity pouze informace o výskytu dané akce obličeje, které byly reprezentovány stavy 0 nebo 1. OpenFace však dokáže určit i intenzitu sledované akce obličeje jako číslo v rozsahu 0-5. Intenzita projevení dané akce obličeje může mít přínos pro trénování stromu, a i výslednou klasifikaci. Následující rozhodovací strom byl tedy naučen na datech o výskytu i intenzitě

dané AU. Trénovací i testovací data byla opět ve formě CSV souborů s obdobným formátem dat, jako je na Tabulka 10. Testovací i trénovací data obsahovala 17 parametrů intenzity AU a 18 parametrů výskytu AU. Pro AU28 byl v rámci OpenFace detekován pouze výskyt dané AU, parametr vyjadřující intenzitu AU28 nebylo možné získat. Pravděpodobně se jedná o chybu autora OpenFace, který opomenul intenzitu AU28 určit nebo vypsat do výstupního souboru. Konkrétně se jedná o „lip suck“, tedy akci obličeje, kdy člověk směruje rty do dutiny ústní, jako je vidět na následujícím obrázku.

Obrázek 22 - Ukázka AU28 [2]

Absence informace o intenzitě AU28 není pro určování požadovaných emočních tříd nijak kritická. Jelikož se daná AU nevyskytuje na seznamu AU pro dané emoce na Tabulka 2, lze tento parametr pro učení stromu vynechat. Pro učení i vyhodnocení rozhodovacího stromu byl použit obdobný kód jako v předcházejícím případě. Jediná změna se týkala úpravě funkce pro načtení trénovacích a testovacích dat, jelikož v tomto případě došlo k nárustu počtů parametrů.

U naučeného rozhodovacího stromu byla sledována celková správnost klasifikace, přičemž proces učení splňoval následující omezující podmínky. Hloubka stromu byla omezena v rozsahu 1-15 a minimální počet vzorků v listu stromu byl omezen v rozsahu 5-40. Byly testovány všechny kombinace pro dané omezující podmínky a nejlepší celkové správnosti bylo dosaženo pro hloubku stromu 7 a minimální počet vzorků v listu 37. Celková správnost klasifikace byla překvapivě nižší než v předchozím případě, její hodnota byla 62,72 %. Předpoklad byl, že strom, kterému bylo předloženo v rámci učení více informací, bude mít u validace lepší výsledky. Tento předpoklad byl ale mylný, nicméně zhoršení celkové správnosti bylo pouze o 1,39 %, což není nijak výrazný pokles. Pozitivním zjištěním však bylo, že nově naučený strom byl schopný s jednodušší architekturou dosáhnout téměř totožných výsledků jako strom předchozí. Hloubka stromu se z 9 snížila na 7 a minimální počet vzorků v listu stromu se zvýšil z 9 na 37. Nově naučený strom se jeví jako robustnější než jeho předchůdce, který sice dosahoval o něco lepších výsledku, ale s velkou pravděpodobností tam mohlo dojít k jeho přeučení. Architekturu nově naučeného rozhodovacího stromu je možné vidět na následujícím obrázku.



Obrázek 23 - Rozhodovací strom, hloubka = 7, počet vzorků na list = 37

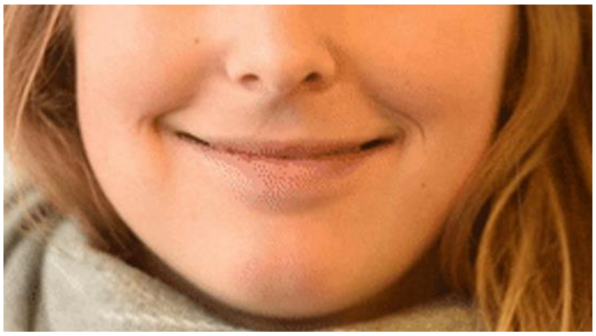
Obdélníky na obrázku představují jednotlivé uzly a listy rozhodovacího stromu. Každý uzel pak obsahuje informaci o rozhodovacím parametru, hodnotě entropie, počtu vzorků, rozdělení vzorků v rámci emočních tříd a výslednou emoční třídu daného uzlu/listu. Rozhodovací parametr je realizován logickou podmínkou s využitím parametrů výskytu akce obličeje (AU_r) a intenzity akce obličeje (AU_c). Je-li daná podmínka splněna dochází k rozvětvení stromu doleva, je-li neplatná tak doprava. Hodnota entropie vyjadřuje míru neuspořádanosti v daném uzlu/listu, vyšší hodnota entropie představuje vyšší míru neuspořádanosti. Počet vzorků udává celkový počet dat, které se budou v rámci uzlu dělit, v případě listu je to konečný počet vzorků, který se už dále nedělí. Počet prvků v prvním uzlu v rozhodovacím stromu také udává celkový počet dat, která byla pro učení stromu použita. Rozdělení vzorků v rámci emočních tříd udává, jaký počet vzorků platí pro danou třídu. Rozdělení je vyjádřeno jednorozměrným polem o sedmi prvcích. Suma všech vzorků pole dává celkový počet vzorků v daném uzlu/listu. Výsledná emoční třída daného uzlu/listu je dána třídou s nejpočetnějším zastoupením. Emoční třída je mimo jiné reprezentována i barvou daného obdélníku. Následující obrázek ukazuje příklad uzlu rozhodovacího stromu.

AU12_r <= 1.5
entropy = 2.8
samples = 1393
value = [182, 182, 196, 224, 231, 176, 202]
class = Neutral

Obrázek 24 - Ukázka uzlu rozhodovacího stromu

Konkrétní ukázka uzlu na Obrázek 24 představuje první uzel naučeného rozhodovacího stromu na Obrázek 23. Rozhodovacím parametrem pro tento uzel bylo splnění logické podmínky, která platí, pokud je intenzita AU12 menší nebo rovna hodnotě 1,5. Tímto

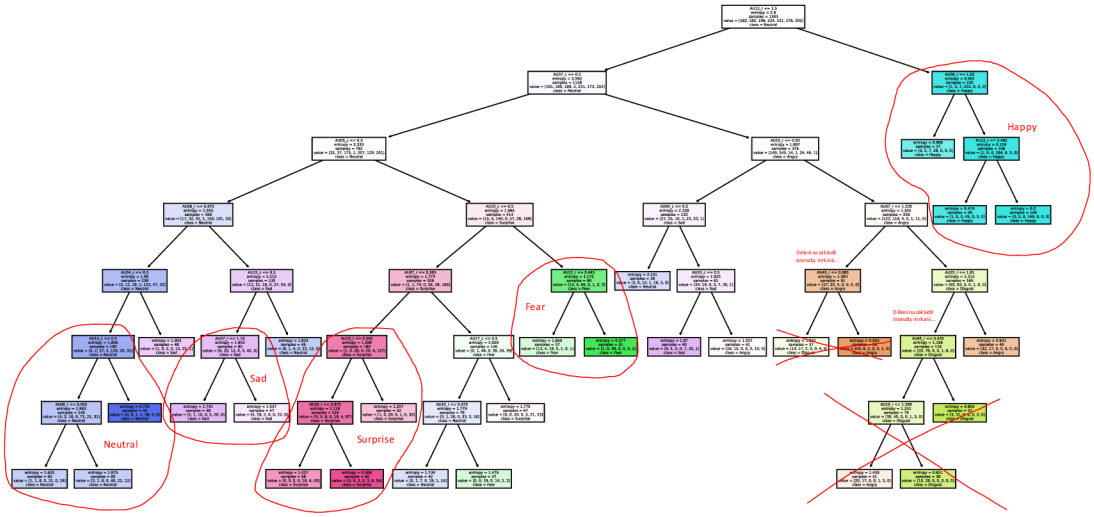
způsobem rozděluje strom množinu všech vzorků na 1158 vlevo pro které podmínka platí a 235 vpravo pro které neplatí. Akce obličeje s číslem 12 se nazývá „lip corner puller“ a je typická pro vyjádření úsměvu. Množina 235 obrázků, pro které bylo splněno, že intenzita vyjádření AU12 je vyšší než 1,5, by měla teoreticky obsahovat úsměvy člověka. Tato hypotéza je potvrzena výslednou emoční třídou dále pokračujících uzlů a listů. U všech pěti je emoční třída štěstí zastoupena nejpočetněji, konkrétně se jedná o 222 případů z 235. Příklad AU12 lze pozorovat na Obrázek 25.



Obrázek 25 - Ukázka AU12 [2]

Dále lze na Obrázek 24 pozorovat, že entropie pro daný uzel je poměrně vysoká, konkrétně s hodnotou 2,8. To je dáno podobným počtem dat napříč všemi třídami. Pro daný uzel konkrétně platí, že 182 patří do třídy zlost, 182 do znechucení, 196 do strach, 224 do štěstí, 231 do neutrální, 176 do smutek a 202 do třídy překvapení. Mezi třídami tedy neplatí téměř žádní uspořádanost a hodnota entropie je vysoká. Výsledná emoční třída je neutrální, protože emoce zaujímá nejpočetnější skupinu s 231 prvky.

Výsledný strom častokrát dělí uzly na další větve, přestože už všechny další uzly a listy mají stejnou třídu. Strom by tedy šlo zjednodušit sjednocením několika uzlů do jednoho. Uzly které, by mohly být sjednoceny jsou na dalším obrázku zakroužkovány červenou barvou, přičemž výsledná emoční třída je u nich také uvedena červeně. Analýzou výsledného stromu bylo rovněž zjištěno, že u dvou uzlů došlo k náhodnému dělení na základě nerelevantního parametru. V obou případech dochází na dělení na základě parametru AU45_r, který představuje intenzitu mrkání. Uvažování intenzity mrkání v tomto případě nemá žádný smysl, a to hned z několika důvodů. Zaprvé se AU45 vůbec nevyskytuje v seznamu AU pro dané emoční třídy v Tabulka 2, a dle teorie FACS by tedy tato akce obličeje neměla mít na tyto emoce žádný vliv. Druhým důvodem je to, že detekce AU pomocí OpenFace byla použita na statické snímky různých osob a jejich obličejů. Jelikož je mrkání proces, který standartně trvá nějakou dobu, pro jeho spolehlivé detekování by byla potřeba sekvence snímků. Detekce mrkání je pro použita data pravděpodobně chybná a neměla by tedy být použita pro trénování stromu. Pro ostatní uzly nebyla nalezena žádná nesrovnalost a ve většině případů se dělící parametry pro dané emoční třídy shodují s Tabulka 2.



Obrázek 26 - Zjednodušený rozhodovací strom, hloubka = 7, počet vzorků na list = 37

Následující tabulka představuje matici záměn pro nově naučený rozhodovací strom. Nejúspěšněji klasifikovaná emoce je opět štěstí, z celkového počtu 50 vzorků bylo 49 rozpoznáno a senzitivita této třídy byla 98 %. Vysokých hodnot bylo dosaženo i pro správnost, přesnost, či F-míru, ve všech případech bylo dosaženo hodnoty vyšší než 90 %.

Tabulka 12 - Matice záměn, OpenFace a AffectNet (výskyt + intenzita akce obličeje)

Celková správnost 62,72 %		**Skutečnost**							
		Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení	Celkem	Přesnost
Predikce	Zlost	14	9	0	0	1	1	0	25	56 %
	Znechucení	5	21	0	0	2	0	0	28	75 %
	Strach	1	0	15	0	0	0	5	21	71 %
	Štěstí	0	0	4	49	0	1	0	54	91 %
	Neutrální	6	4	10	1	38	16	5	80	47 %
	Smutek	12	8	1	0	2	18	2	43	42 %
	Překvapení	0	0	4	0	7	0	25	36	69 %
	Celkem	38	42	34	50	50	36	37		
	Senzitivita	37 %	50 %	44 %	98 %	76 %	50 %	68 %		
	Správnost	88 %	90 %	91 %	98 %	81 %	85 %	92 %		
	F1 míra	44 %	60 %	55 %	94 %	58 %	46 %	68 %		

Co se týče ostatních emočních tříd, nedocházelo k žádným významným změnám oproti předešlému výsledku. Když porovnáme výsledky F-míry, která zahrnuje jak senzitivitu, tak přesnost klasifikace, jsou rozdíly u jednotlivých tříd maximálně 5 %.

Dle dosažených výsledků nemělo přidání příznaku intenzity akce obličeje velký vliv na kvalitu klasifikace. Toto je zvláštní výsledek, protože očekáváním bylo, že se klasifikace o něco zpřesní. Důvod proč se přidání intenzity akce obličeje nijak významně neprojevilo může být ten, že pro validaci nebyl použit dostatečně velký počet dat. Na testování bylo použito 350 obrázků, tedy 50 obrázků pro každou třídu. Bohužel se ale pro každý nepodařilo získat kvalitní příznaky, a tak je reálný počet dat ještě menší. Problém s kvalitou datasetu AffectNet byl nastíněn v kapitole 1.4 a část datasetu, která byla v rámci práce reanotovaná nebyla dostatečná. Pro spolehlivější porovnání OpenFace na datasetu AffectNet ho bude třeba rozšířit.

2.4 OpenPose

Jelikož datasety FER2013 a AffectNet neobsahují fotografie zachycující celého člověka včetně končetin, ale pouze oříznutý obličej s projevovanou emocí, nebylo možné OpenPose na těchto datasetech otestovat. Spolehlivost detekování bodů zájmu však byla otestována na záznamu webkamery a řadě jiných obrázků (viz Obrázek 27), získaných například z knihovny Microsoftu [40]. Pro statické obrázky bylo dosaženo spolehlivého určení jednotlivých bodů lidského skeletu. Částečných výsledků bylo dosaženo i pro detekování kloubů jednotlivých prstů na ruce. Přesnost detekování bodů v reálném čase z webkamery však musela být snížena, z důvodu nedostatečného výpočetního výkonu. Detekování kloubů jednotlivých prstů pak pro záznam z webkamery nemohlo být provedeno vůbec.

Obrázek 27 - Test spolehlivosti detekce pomocí OpenPose [40]

Použití OpenPose pro systém pro rozpoznávání emocí je tedy limitované, není možné detekovat například mikro-gesta zobrazena na Obrázek 9. Částečně lze však OpenPose využít například pro detekování pozic horních končetin. I tyto údaje by mohly vést ke zpřesnění klasifikace emoce.

3. REALIZACE

Kapitola popisuje realizaci systému pro rozpoznávání emocí na základě předešlých pokusů a zjištěných informací. V rámci kapitoly jsou popsány zvolené snímací a vyhodnocovací prostředky a scéna, kterou bude systém snímat. Dále kapitola popisuje rozšíření datasetu AffectNet pomocí augmentace a shrnuje výsledky DeepFace a OpenFace na rozšířeném datasetu. Převážná část kapitoly se věnuje spojení všech tří vyhodnocovacích metod, tedy DeepFace, OpenFace a OpenPose a návrhu jednoduchého uživatelského rozhraní. V rámci realizace systému je také uvažován proces přeučení na konkrétní osobu.

3.1 Návrh snímacího a vyhodnocovacího HW

Jak už bylo uvedeno v textu výše, systém má být používán v rámci dialogu pedagoga a dítěte, případně jiných dvou osob. První osoba (pedagog) je snímaná kamerou a její emoce je vyhodnocována systémem. Výstup systému je vhodným způsobem zobrazován druhé osobě, tedy dítěti s poruchou autistického spektra. Je uvažováno, že kamera bude osobu snímat ze vzdálenosti přibližně jednoho metru. Přepokládá se, že v záběru kamery bude zachycena osoba od pasu nahoru. Bude tedy možné zaznamenávat a extrahovat příznaky z obličeje snímané osoby. a stejně tak bude možné vyhodnocovat pohyby horních končetin. Rovněž se předpokládá, že snímaná osoba bude vhodným způsobem osvětlena.

Jako snímací a vyhodnocovací hardwarový prostředek lze s výhodou využít osobní počítač nebo notebook s webkamerou. Pro tuto diplomovou práci byl konkrétně použit notebook Acer Nitro s procesorem Intel Core i5 7300HQ a dedikovanou grafickou kartou NVIDIA GeForce GTX 1050 a operačním systémem Windows 10. Jako snímací zařízení byla použita vestavěná webkamera s maximálním rozlišením 1280x720 pixelů. K dispozici byla také druhá externí webkamera s rozlišení 1920x1080 pixelů.

3.2 Augmentace dat

Z důvodu nedostatečného počtu dat byly hledány způsoby, jak současně získané datasety rozšířit. Pro to se hodí technika augmentace dat, která se používá právě pro rozšíření datasetu. Augmentace spočívá ve vygenerování nových dat z již předloženého obrazu. Podmínkou kvalitní augmentace je změnit obraz tak, aby se pro algoritmus počítačového vidění jevil jako obraz nový, ale zároveň zachovat atributy obrazu, které je potřeba pro spolehlivé extrahování příznaků. Způsobů, jak augmentovat data je mnoho, může se jednat o různé deformace obrazu, transformace, změny barev, jasových úrovní, přidání šumu, oříznutí obrazu atd. Pro účely rozšíření datasetu byly použity následující metody:

a) **Vodorovné překlopení obrazu**
   - na extrahování příznaků z obličeje nemá překlopení žádný vliv
   - metoda je vhodná i pro obrazy zachycující celé tělo člověka
   - dostatečně změní původní obraz, v podstatě všechny pixely původního obrazu se změní

b) **Změna jasu obrazu**
   - změna jasu bude v omezeném intervalu, aby nedocházelo k přesvícení, nebo přílišnému ztmavení snímku, které by mohlo ovlivnit extrakci příznaků
   - nemusí však vždy dostatečně změnit daný obraz

c) **Kombinace překlopení a změny jasu**
   - prosté spojení dvou předchozích metod

d) **Transformace na šedotónový obraz**
   - výraz člověka nebo jeho postoj budou stále dobře čitelné
   - dostatečně změní původní obraz

Všechny výše zmíněné metody augmentace dat byly současně kombinovány s náhodným natočením obrazu. Úhel natočení byl generován s náhodným rovnoměrným rozložením z intervalu 0-6 °. Cílem přidání náhodného natočení obrazu bylo zajištění unikátnosti nově generovaného obrazu. V případě, že například změna jasu augmentovaného obrazu bude minimální, rotace o náhodný úhel zajistí, že nově generovaný obraz se bude lišit od originálu.

Jednotlivé augmentační metody byly realizovány jako jednotlivé funkce [33] v Pythonu. Vstupními argumenty funkcí byly omezující parametry pro augementaci, díky kterých se nastavil například maximální úhel natočení obrazu a jiné parametry pro dané funkce. Jednotlivé funkce vždy vracely změněný obraz jako proměnnou *img*.

Pomocí proměnných *inputDir* a *outputDir* byly definovány absolutní cesty ke složkám v počítači. Vstupní složka obsahovala data, která byla předložena skriptu a tyto data byly augmentovány. Skript vždy načetl danou fotografii, provedl patřičnou augmentační metodu a pod novým názvem fotografii uložil do výstupní složky. Díky čtyřem metodám augmentace dat bylo možné AffectNet 5x rozšířit. Ukázka augmentovaných dat je zobrazena na následujícím obrázku.

Obrázek 28 - Ukázka augmentovaných dat

3.2.1 Výsledky na augmentovaných datech

U předchozích výsledků byla klasifikace pomocí DeepFace a OpenFace testována na validační množině datasetu AffectNet, která obsahovala 50 obrázků pro každou emoční třídu. Takto malý počet dat mohl vést ke zkresleným výsledkům, a proto byla kvalita klasifikace otestována na augementovaném validačním datasetu. Ten už čítal 250 vzorků dat pro každou emoční třídu.

Následující matice záměn zobrazuje výsledky klasifikace DeepFace na rozšířené validační množině. Na rozdíl od předešlé klasifikace na Tabulka 9, je emoce vyhodnocena pouze u snímku, u kterého byla pravděpodobnost určené emoce vyšší než 50 %. Z tohoto důvodu jsou celkové počty skutečných snímků nižší, než je velikost validační množiny. Celková správnost klasifikace modelu DeepFace na augmentovaném datasetu byla téměř shodná jako v předešlém případě. Taktéž hodnoty přesnosti, senzitivity, správnosti a F1 míry u jednotlivých tříd dosahovaly podobných úrovní jako u předešlého testování na-Matice záměn, DeepFace a AffectNet Tabulka 9. Předchozí výsledky klasifikace jsou tedy pravděpodobně správné a nejsou zkreslené malým počtem validačních dat.

Tabulka 13 - Matice záměn, DeepFace a AffectNet (augmentovaný)

Celková správnost 42,67 %		**Skutečnost**								
		Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení	Celkem	Přesnost
Predikce	Zlost	103	60	9	3	15	18	24	232	44 %
	Znechucení	2	15	1	0	0	0	0	18	83 %
	Strach	48	21	94	7	21	42	35	268	35 %
	Štěstí	27	53	47	195	14	35	26	397	49 %
	Neutrální	30	28	17	18	111	55	45	304	37 %
	Smutek	19	39	18	8	32	70	21	207	34 %
	Překvapení	6	7	41	5	4	10	90	163	55 %
	Celkem	235	223	227	236	197	230	241		
	Senzitivita	44 %	7 %	41 %	83 %	56 %	30 %	37 %		
	Správnost	84 %	87 %	81 %	85 %	82 %	81 %	86 %		
	F1 míra	44 %	12 %	38 %	62 %	44 %	32 %	45 %		

Na augementované validační množině byl vyhodnocen také rozhodovací strom a OpenFace. Byla snaha použít pro učení stromu také augmentovanou trénovací množinu. Bohužel se však nepodařilo dosáhnout nijak lepších výsledků. Pro trénování stromu tedy byla použita původní trénovací množina. Pro OpenFace byl také nastaven omezující

parametr, a to v podobě podmínky určující minimální věrohodnost detekce orientačních bodů obličeje. Emoce byla vyhodnocena pouze v případě, kdy byla věrohodnost detekce vyšší než 76 %. Tato hranice byla určena experimentálně. I přes omezující podmínku byla vyhodnocena téměř všechna data pro emoční třídu. Celková správnost klasifikace byla na augmentovaných datech o 12,2 % nižší než u předešlé klasifikace na Tabulka 12. Předchozí vyhodnocení tedy bylo pravděpodobně zkreslené nízkým počtem testovacích dat. Podrobné výsledky lze vidět v matici záměn níže.

Tabulka 14 - Matice záměn, OpenFace a AffectNet (augmentovaný)

Celková správnost 50,55 %		**Skutečnost**								
		Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení	Celkem	Přesnost
Predikce	Zlost	95	52	39	10	14	6	0	216	44 %
	Znechucení	59	100	22	2	7	37	1	228	44 %
	Strach	24	15	54	1	10	7	27	138	39 %
	Štěstí	12	4	32	212	0	13	0	273	78 %
	Neutrální	5	6	26	5	136	39	39	256	53 %
	Smutek	46	50	20	20	42	112	21	311	36 %
	Překvapení	3	13	47	0	41	30	157	291	54 %
	Celkem	244	240	240	250	250	244	245		
	Senzitivita	39 %	42 %	23 %	85 %	54 %	46 %	64 %		
	Správnost	84 %	84 %	84 %	94 %	86 %	81 %	87 %		
	F1 míra	41 %	43 %	29 %	81 %	54 %	40 %	59 %		

3.3 Spojení vybraných metod

V rámci diplomové práce byly otestovány tři metody, které se jeví jako vhodné pro navrhovaný systém. Konkrétně se jedná o DeepFace a OpenFace pro vyhodnocení dat zaznamenaných z obličeje a OpenPose, které by mohlo být použito pro analýzu pohybu horních končetin. Metody byly testovány na obrázcích i videích, ovšem vždy samostatně. Výsledky jednotlivých metod byly uspokojivé, nicméně se přepokládá, že vhodným spojením zmíněných metod bude dosaženo lepších výsledků při klasifikaci emoce. Jednou z komplikací při spojení zmíněných metod je fakt, že nejsou napsány ve stejném programovacím jazyce. V případě DeepFace se jedná o knihovnu napsanou v Pythonu. OpenFace je zase naopak projekt napsaný čistě v C++. Pouze OpenPose obsahuje aplikační rozhraní (API) jak pro C++, tak pro Python.

3.3.1 Spojení DeepFace a OpenFace

Jak již bylo napsáno v úvodu kapitoly 3.3, OpenFace je celý napsán v C++ a v současné době neexistuje žádné API pro Python. Spojení těchto dvou metod bylo nutné provést jiným způsobem.

Prvním „naivním“ pokusem bylo paralelní spuštění obou aplikací a jejich analýza dat z webkamery. Předpoklad byl, že by obě aplikace zapisovaly svůj výstup do souboru, který by byl dále vyhodnocován. Bohužel bylo záhy zjištěno, že po spuštění jedné z aplikací dojde k zablokování přístupu k webkameře všem dalším aplikacím. V případě, že byla jako první aplikace spuštěna OpenFace, došlo k zamezení čtení dat z webkamery, a následná analýza pomocí DeepFace již nebyla možná. Byla snaha tento problém řešit virtuální webkamerou, tedy streamováním stejných dat zaznamenávaných fyzickou webkamerou na virtuální kanál. Pro tyto účely byly otestovány programy, které se primárně používají pro modifikaci dat zaznamenaných webkamerou. Jedná se především o přidávání grafických elementů nebo grafických filtrů do původního streamu dat. Využití těchto programů je například při video rozhovorech, kdy si uživatel může například změnit nebo rozmazat pozadí, které za ním webkamera snímá. Tyto programy fungují způsobem, že dokáží číst data z webkamery a replikovat je jako další zdroj dat, ze kterého mohou ostatní aplikace číst. Přesně tato funkcionalita by se hodila pro vyřešení spojení DeepFace a OpenFace. Proto byly otestovány programy ManyCam a YouCam a bylo zjištěno, že OpenFace je možné použít pro čtení dat generovaným zmíněným softwarem, bohužel však s velice nízkým počtem snímků za sekundu. Streamování dat dosahovalo rychlosti pouhých 10 FPS (Frames per secod), což je velice málo, úváží-li se využití pro zpracování dat v reálném čase. Toto řešení s virtuální webkamerou nevedlo k uspokojivým výsledkům a bylo vyhodnoceno jako nevhodné.

Druhý pokus, jak spojit DeepFace a OpenFace, byl pomocí skriptu v Pythonu. V tom byly DeepFace předloženy jednotlivé snímky z webkamery, které mohly být vyhodnoceny funkcí *DeepFace.analyze()*. Zároveň byly jednotlivé snímky ukládány do složky, kde mohly být analyzovány pomocí OpenFace. Pro pojmenovávání ukládaných snímků byl zvolen unikátní identifikátor ve formě inkrementovaného čísla. Jednotlivé snímky byly ukládány ve formátu .jpg. Z důvodu absence Python API nebylo možné analyzovat snímky přímo v kódu, nicméně OpenFace obsahuje již zkompilované EXE soubory pro jednotlivé funkce, jako například analýza obličeje, analýza pohledu, extrakce příznaků z obličeje atd. Pro analýzu obličeje byl použit *FeatureExtraction.exe* soubor, který umožňuje analyzovat jednotlivé fotky nebo celou sekvenci a extrahované příznaky zapíše do CSV souboru. V rámci Python kódu byl tedy uložený obrázek z webkamery analyzován právě pomocí *FeatureExtraction.exe*. Ten byl v Pythonu spuštěn pomocí modulu subprocess [34]. Subprocess modul umožňuje spustit soubory stejným způsobem jako je to možné pomocí příkazové řádky. Každý snímek z webkamery byl tímto způsobem analyzován a extrahované příznaky byly uloženy do samostatného CSV souboru, který měl stejný název jako analyzovaný snímek. Tento CSV soubor byl po

analýze přečten pomocí CSV modulu [35] a extrahované příznaky mohly být použity v kódu.

Bohužel byla analýza dat tímto způsobem velice pomalá. Průměrná rychlost zpracování dat byla okolo dvou snímků za sekundu, což je pro analýzu emoce v reálném čase příliš nízká hodnota. Vzhledem k tomu, že pro každý snímek byl zapisován nový CSV soubor, byla snaha omezit počet zapisovaných výstupních dat. Úvaha byla, že pokud by se podařilo zrychlit zápis, který se provádí při analýze každého snímku, bylo by možné zvýšit celkovou rychlost analýzy. Výchozí nastavení OpenFace, tedy *FeatureExtraction.exe,* totiž zapisuje všechny extrahované příznaky, ať už jde o směr pohledu, natočení hlavy nebo samotné pozice jednotlivých orientačních bodů. Jde o přibližně 700 různých parametrů, které se ovšem nevyužívají pro samotnou klasifikaci emoce. Pro určení emoce pomocí naučeného rozhodovacího stromu je relevantních pouze 35 příznaků, tedy výskyt a intenzita detekovaných akcí obličeje. Omezení zapisovaných dat však nemělo na rychlost analýzy téměř žádný vliv. Počet analyzovaných snímků za sekundu se v podstatě nezměnil. Řešení je však možné použít pro analýzu fotografií a vyhodnocení emoce s kombinovaným použitím DeepFace i OpenFace. Použití skriptu pro data z webkamery, a jejich analýzu v reálném čase, je ale nevhodné.

Dalším zkoumáním a testováním bylo zjištěno, že rychlost analýzy snímků pomocí OpenFace a souboru *FeatureExtraction.exe*, je odlišná pro analýzu jednotlivých snímků nebo jejich sekvenční analýzu. OpenFace totiž umožňuje volit módy, jakým způsobem spustí analýzu pomocí *FeatureExtraction.exe.* V případě analýzy konkrétního obrázku se na začátku pokaždé načtou a inicializují modely, které jsou nutné pro analýzu. Konkrétně jde o modely pro detekci obličeje, nebo detekci orientačních bodů obličeje. Poté následuje analýza snímku a extrakce příznaků, případná vizualizace, zápis dat do souboru a na úplném konci jsou modely resetovány a analýza je ukončena. V případě analýzy dalšího snímku tímto způsobem se celý proces opakuje od načtení modelů až po jejich resetování. Naopak u sekvenční analýzy se modely načtou a inicializují pouze jednou a v rámci inicializační fáze se také určí, kolik dat se bude analyzovat (na základě počtu obrázků v dané složce). Následující extrakce příznaků, vizualizace a zápis dat pak probíhá ve smyčce, která je ukončena v případě, že byly analyzovány všechny obrázky ve složce. Proces je ukončen resetem modelů a ukončením analýzy. V případě sekvenční analýzy bylo možné dosahovat mnohem vyššího počtu analyzovaných snímků za sekundu než v případě analýzy jednotlivých snímků, jednalo se o hodnotu, která se pohybovala mezi 15-30 FPS. Tento výsledek byl už mnohem slibnější pro analýzu dat v reálném čase, nicméně i tento přístup měl svá omezení.

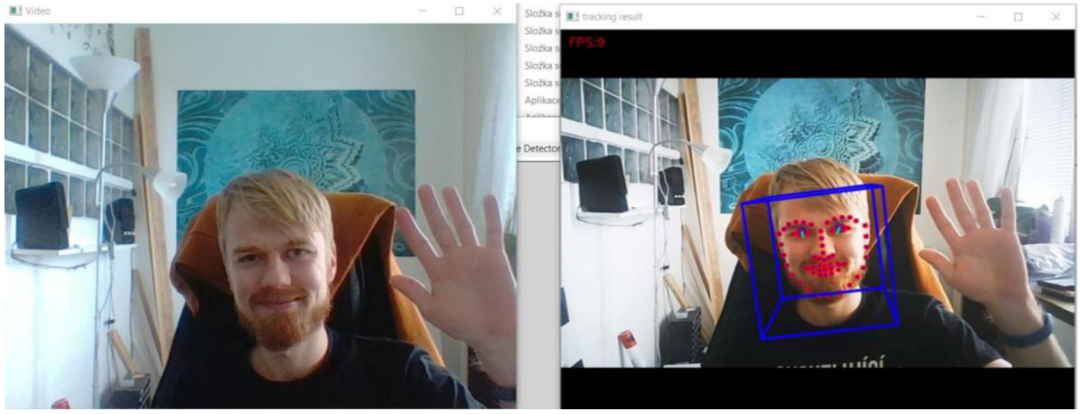
Díky sekvenční analýze dat bylo dosaženo toho, že načítání modelů, které zabírá mnoho času, bylo provedeno pouze jednou a proces extrakce příznaků ze snímků byl výrazně urychlen. Problém byl ovšem ten, že počet obrázků ve složce, které bylo nutné analyzovat nebyl konstantní ale neustále rostl. Jednotlivé snímky byly ukládány díky Python skriptu, který je četl jako záznam z webkamery. OpenFace ovšem v případě

spuštění *FeatureExtraction.exe* očekával neměnný počet analyzovaných snímků. Pokud například v době spuštění *FeatureExtraction.exe* bylo ve složce uloženo 10 snímků, EXE soubor analyzoval pouze těchto 10 snímků a další, které byly dále ukládány už nebral v potaz. Cílem tedy bylo, pokusit se modifikovat C ++ kód pro *FeatureExtraction.exe* a kompilací a buildem získat novou .exe aplikaci, která by fungovala pro sekvenční analýzu obrázku, jejichž počet se v čase bude měnit.

Pro sekvenční analýzu dat je v rámci OpenFace a aplikace *FeatureExtraction.exe* využívaná třída *SequenceCapture*, která není součástí standartních knihoven, ale je unikátním řešením právě pro OpenFace. Objekt *SequenceReader* je instancí dané třídy a využívá se mimo jiné pro postupné procházení obrázků ve složce. Pro tyto účely je sestavena datová struktura typu fronta, která je inicializována v počáteční fázi procesu (kdy jsou například načítány modely). Objekt *SequenceReader* pro procházení a načítání dat využívá právě tuto frontu, v případě, že fronta už neobsahuje žádný prvek, vrátí *SequenceReader* hodnotu reprezentující prázdný snímek. V případě prázdného snímku je ukončena i hlavní smyčka, ve které jsou jednotlivé obrázky analyzovány a dojde k resetu modelu a ukončení analýzy. Toto je důvod, proč v předešlém příkladu popisu fungování *FeatureExtraction.exe* byla analýza ukončena už po 10 snímcích. Fronta byla inicializována na začátku procesu, kdy bylo ve složce pouze 10 obrázků. V případě nově uložených obrázku už nebyla aktualizována, a tím pádem byla smyčka po 10 iteracích ukončena.

Vzhledem k tomu, že OpenFace je komplexní projekt, na němž pracovala řada lidí, nebylo snadné projekt celý do detailu pochopit. Aplikace *FeatureExtraction.exe* používá mimo třídu *SequenceReader* i další třídy, které jsou navzájem provázány a jsou důležité pro celkové správné fungování. Ačkoliv byla snaha upravit logiku fungování smyčky pro analýzu obrázku, samotnou třídu *SequenceReader* nebo přidat kód pro průběžnou aktualizaci fronty, nebylo dosaženo funkčního řešení.

Vzhledem k neúspěšným pokusům o úpravu softwaru byla potřeba vyřešit spojení metod jiným způsobem. Ačkoliv předešlé pokusy s virtuální webkamerou neměly očekávané výsledky, bylo jimi potvrzeno, že lze OpenFace použít jednoduše na jiný stream dat než zabudovanou webkameru. Byla tedy pořízena externí webkamera, která byla k notebooku jednoduše připojena přes USB rozhraní a byla připevněna hned vedle zabudované webkamery. Tímto způsobem bylo docíleno téměř stejného záznamu na obou zařízeních a aplikace bylo možné spustit paralelně. Na následujícím obrázku je vidět záznam z obou webkamer. V levé části obrázku je vidět záznam ze zabudované webkamery, který je využíván pro analýzu pomocí DeepFace a v pravé části je výstup OpenFace, který pro analýzu využívá externí webkameru.



Obrázek 29 - Paralelní běh DeepFace a OpenFace

Výsledné spojení obou metod je řešeno v Pythonu. Pro větší přehlednost byl vytvořen soubor *openFace.py*, který obsahuje jednotlivé funkce pro snazší používání OpenFace v Pythonu. Mezi hlavní funkce *openFace.py*, které stojí za zmínku patří:

openFace.featuresExtractionWebcam()

- funkce, která slouží pro spuštění analýzy dat z webkamery pomocí OpenFace
- analýza je provedena souborem *FeatureExtraction.exe*, který je spuštěn pomocí Python modulu Subprocess
- díky funkce *subprocess.Popen()* je proces spuštěn způsobem, který nijak neomezuje pokračující Python kód, proces pouze spustí danou EXE aplikaci a ta běží na pozadí
- *subprocess.Popen()* funguje obdobným způsobem, jako by byl EXE soubor spuštěn z příkazové řádky, argumenty pro spuštění jsou definovány proměnnou *args*, cesta k EXE souboru proměnnou *exePath* a výstupní složka proměnnou *outDir*
- před spuštěním procesu dojde ještě ke smazání předešlého obsahu výstupní složky, která může obsahovat stará data z předešlé analýzy
- funkce vrací proměnnou proces, kterou lze pak dále použít například pro ukončení procesu v hlavním zdrojovém souboru

openFace.checkCSV()

- funkce, která kontroluje, zda už existuje CSV soubor s výstupními daty z OpenFace
- pokud je soubor nalezen, vypíše tuto informaci do terminálu a vrátí cestu k vytvořenému CSV souboru

openFace.createCustomCsv()

- funkce pro vytvoření nových CSV souborů pro přeučení rozhodovacího stromu
- nové soubory jsou uloženy do složky definované uživatelem
- soubory jsou po ukončení funkce připraveny na zápis nových dat z OpenFace

openFace.writeToCustomCSV(csvReadPath, emotion)

- funkce zapisuje data extrahovaná během procesu trénování na konkrétní emoci
- na základě výběru emoce zapíše data včetně anotované emoce na příslušnou pozici trénovacího a testovacího CSV souboru, který byl vytvořen pomocí funkce *openFace.createCustomCsv()*
- argumenty *csvReadPath* představují cestu k CSV souboru, do kterého OpenFace data zapisuje, *emotion* pak představuje emoci vybranou uživatelem

openFace.predict(csvFilePath, treeClass, lastPosition, skipHeader)

- funkce pro predikci výstupní emoce na základě dat přečtených z CSV souboru
- zjišťuje, zda se zvětšila velikost čteného CSV souboru, a pokud ano, přečte nová data, která do něj byla zapsána
- pro dynamické čtení CSV souboru se používá proměnná *lastPosition*, která definuje pozici, na kterou je třeba se před začátkem čtení posunout
- na základě přečtených dat vrátí dominantní emoci a pozici pro příští čtení CSV souboru
- funkce umožňuje určit i pravděpodobnost dané emoce v procentech, například pokud bylo přečteno 10 snímků a u 7 z nich byla predikována emoce štěstí, bude výstup „happy“ a její pravděpodobnost „70 %”.

Funkce *openFace.predict()* využívá naučeného rozhodovacího stromu. Pro ten byl také vytvořen zvláštní Python soubor – *decTree.py*, který obsahuje relevantní funkce k rozhodovacímu stromu. Funkcí *decTree.trainTree(plotTree)*, je realizováno načtení trénovacích a testovacích CSV souborů pro učení stromu, rozdělení data na příznaky a výstupní třídy a samotné učení. Pomocí parametru *plotTree* pak lze zvolit, zda má být daný strom také graficky zobrazen. Funkce vrací objekt naučeného stromu, který lze využit pro predikování emocí. Díky možnosti přetrénování stromu pomocí jedné funkce je možné jednoduše strom přeučovat během chodu programu.

Tímto způsobem bylo úspěšně realizováno spojení DeepFace a OpenFace i pro analýzu dat z webkamery. Pro porovnání s předchozími výsledky byl proveden test kvality klasifikace na statických snímcích. Pro vyhodnocení byla opět využita augmentovaná validační množina datasetu AffectNet. Bylo však nutné výstupy obou metod vhodným způsobem spojit a vyhodnotit finální emoci. DeepFace mimo dominantní emoci vyhodnocuje i její procentuální pravděpodobnost, ovšem rozhodovací strom u OpenFace pro statický snímek vrací pouze výslednou emoci. Nelze tedy porovnávat, který výstup je pravděpodobnější, a z tohoto důvodu bylo spojení metod vyhodnoceno způsobem, že jako správně klasifikovaná emoce bude vyhodnocen případ, kdy alespoň jedna z metod klasifikovala správně. Následující tabulka zobrazuje podrobné výsledky pro danou klasifikaci. Omezující podmínky pro DeepFace byly minimální pravděpodobnost dominantní emoce 90 %. Tato hodnota byla zvýšena z 50 % na 90 %, protože DeepFace vykazovalo horší výsledky klasifikace než OpenFace. Pro OpenFace

bylo stanoveno omezení minimální jistoty detekce bodů obličeje 77 %. Celková správnost klasifikace se zvýšila o více než 8 %. Zvýšily se i průměrné hodnoty přesnosti, senzitivity a F1 míry.

Tabulka 15 - Matice záměn, DeepFace + OpenFace a AffectNet (augmentovaný)

Celková správnost 58,79 %		**Skutečnost**								
		Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení	Celkem	Přesnost
Predikce	Zlost	130	57	25	7	12	3	2	236	55 %
	Znechucení	36	104	18	0	6	24	1	189	55 %
	Strach	20	16	89	1	13	17	17	173	51 %
	Štěstí	21	20	31	230	5	20	4	331	69 %
	Neutrální	6	7	26	2	157	35	24	257	61 %
	Smutek	29	30	14	10	30	121	18	252	48 %
	Překvapení	2	8	40	0	27	24	179	280	64 %
	Celkem	244	242	243	250	250	244	245		
	Senzitivita	53 %	43 %	37 %	92 %	63 %	50 %	73 %		
	Správnost	87 %	87 %	86 %	83 %	89 %	85 %	90 %		
	F1 míra	54 %	48 %	43 %	79 %	62 %	49 %	68 %		

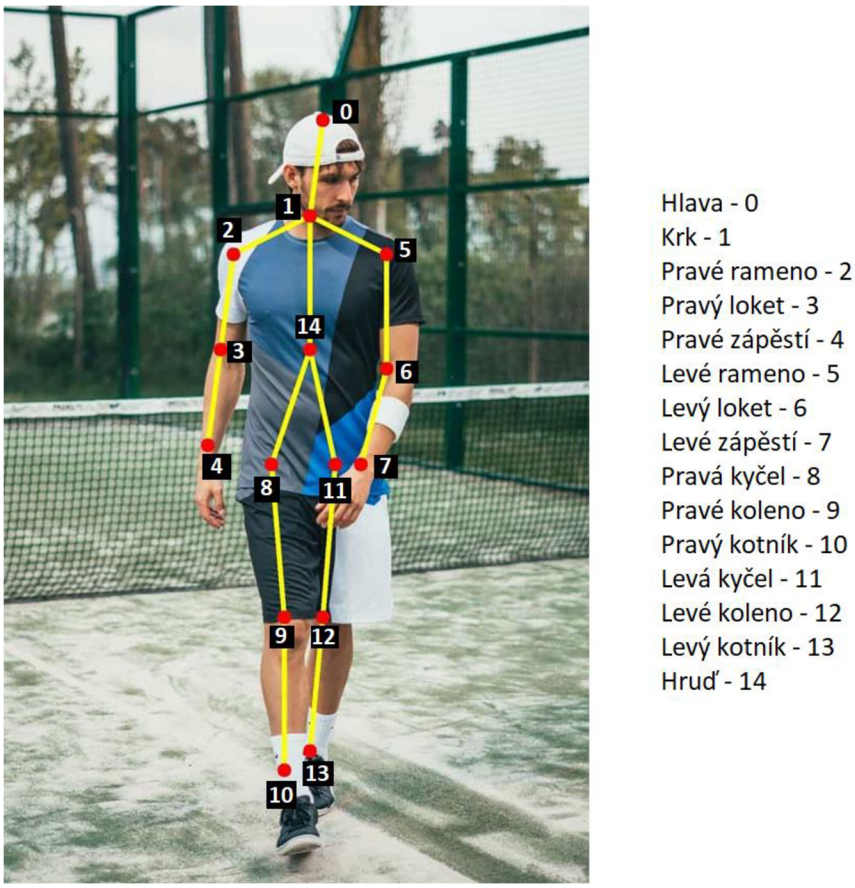
Spojením DeepFace a OpenFace bylo docíleno lepší klasifikace, nicméně očekávání byla vyšší. Přestože bylo mnoho úsilí věnováno pečlivé filtraci a opravě anotací, docházelo ve spoustě případech k tomu, že některé emoce byly z hlediska obličeje velice podobné. Z tohoto důvodu docházelo k častým záměnám podobně vyjádřených emocí, například „zlost“ a „znechucení“ nebo „strach“ a „překvapení“. Pro kvalitnější rozpoznávání bude tedy potřeba získávat příznaky i jinde než z oblasti obličeje. K DeepFace a OpenFace bude nutné implementovat i OpenPose.

3.3.2 Realizace OpenPose v Pythonu

Pro implementaci OpenPose v Pythonu byl využit naučený model pro rozpoznávání pózy lidského skeletu. Tvůrci OpenPose nabízejí 3 různé modely, které jsou všechny ve formátu CAFFE modelu (Convolutional Architecture for Fast Feature Embedding). Modely jsou vlastně hluboké neuronové sítě, které jsou již naučeny a jsou zahrnuty v souboru s příponou *caffemodel*. Pro specifikaci architektury dané sítě se pak používá soubor s příponou *prototxt*. Takto připravenou síť lze v Pythonu jednoduše načíst pomocí OpenCV funkce *dnn.readNetFromCaffe(protoFile, weightsFile)*. Proměnná *protoFile*

představuje soubor s příponou *prototxt* a *weightsFile* pak představuje samotný CAFFE model.

Modely se jmenují podle datasetů, které byly použity pro učení neuronové sítě. Konkrétně se jedná o datasety BODY_25, COCO a MPII. Pro diplomovou práci byl zvolen model MPII, který je dle zdrojů 1,5násobně rychlejší než model COCO [36]. Model COCO umí detekovat 18 různých bodů lidského skeletu, zatímco model MPII detekuje pouze 15 bodů, což je možná i jeden z důvodů, proč vykazuje větší rychlost. Model BODY_25 nebyl pro použití v práci uvažován vůbec, hlavně z důvodu většího počtu detekovaných bodů a předpokládané vyšší výpočetní náročnosti. Model BODY_25 umí detekovat 25 bodů lidského skeletu. Formát výstupních dat MPII modelu je zobrazen na následujícím obrázku.



Obrázek 30 - Výstupní formát dat MPII modelu [36]

V rámci práce je uvažováno pouze využití detekce paží. Vzhledem k tomu, že bude systém zaznamenávat a vyhodnocovat sedícího člověka, nebudou dolní končetiny viditelné, a tudíž pro ně nebudou dané body detekovány. Pro OpenPose byl opět pro větší přehlednost vytvořen samostatný Python soubor, který obsahoval funkce, které jsou dále použity v hlavním zdrojovém souboru. Funkce, které stojí za zmínku jsou následující:

openPose.loadModel()

- funkce pro načtení modelu ze souboru *caffemodel* a *prototxt*
- rovněž slouží pro nastavení zařízení pro výpočetní operace, v tomto případě grafickou kartu
- funkce vrací instanci neuronové sítě *net*, která se pak používá pro další operace

openPose.getPoints(output, frame)

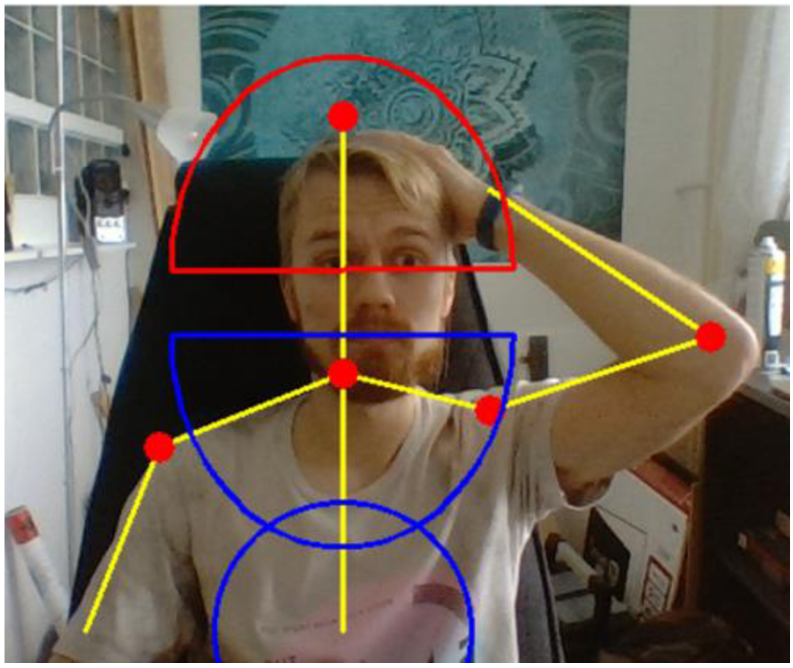
- funkce, která na základě proměnné *output*, která představuje výstup z MPII modelu a proměnné *frame,* která představuje vstupní obraz vrací detekované body skeletu
- body jsou vraceny jako dvourozměrné pole, v případě že není pro vstupní obraz daný bod detekován, je v poli reprezentován hodnotou *None*

openPose.DrawSkeleton(frame, points)

- funkce, která slouží pro grafické znázornění detekovaných bodů
- jednotlivé body jsou v případě jejich úspěšné detekce znázorněny červeným vyplněným kruhem, jednotlivé kosti, respektive spojnice mezi 2 body jsou vizualizovány žlutou přímkou

openPose.handsPos(frame, points, visButtState)

- funkce, která slouží pro detekci pozice paží na základě vstupního obrazu *frame* a detekovaných bodů skeletu *points*
- díky funkce je možné rozpoznat případ, kdy se ruka nachází v horní části obličeje, dolní části obličeje, v oblasti hrudi nebo je zvednutá
- v rámci funkce jsou definovány oblasti pro horní část obličeje – červená půl elipsa, spodní část obličeje – modrá půl elipsa a pro hrudník – modrý kruh
- jednotlivé oblasti jsou zobrazeny na následujícím obrázku



Obrázek 31 - Definované oblasti pro detekci pozice rukou

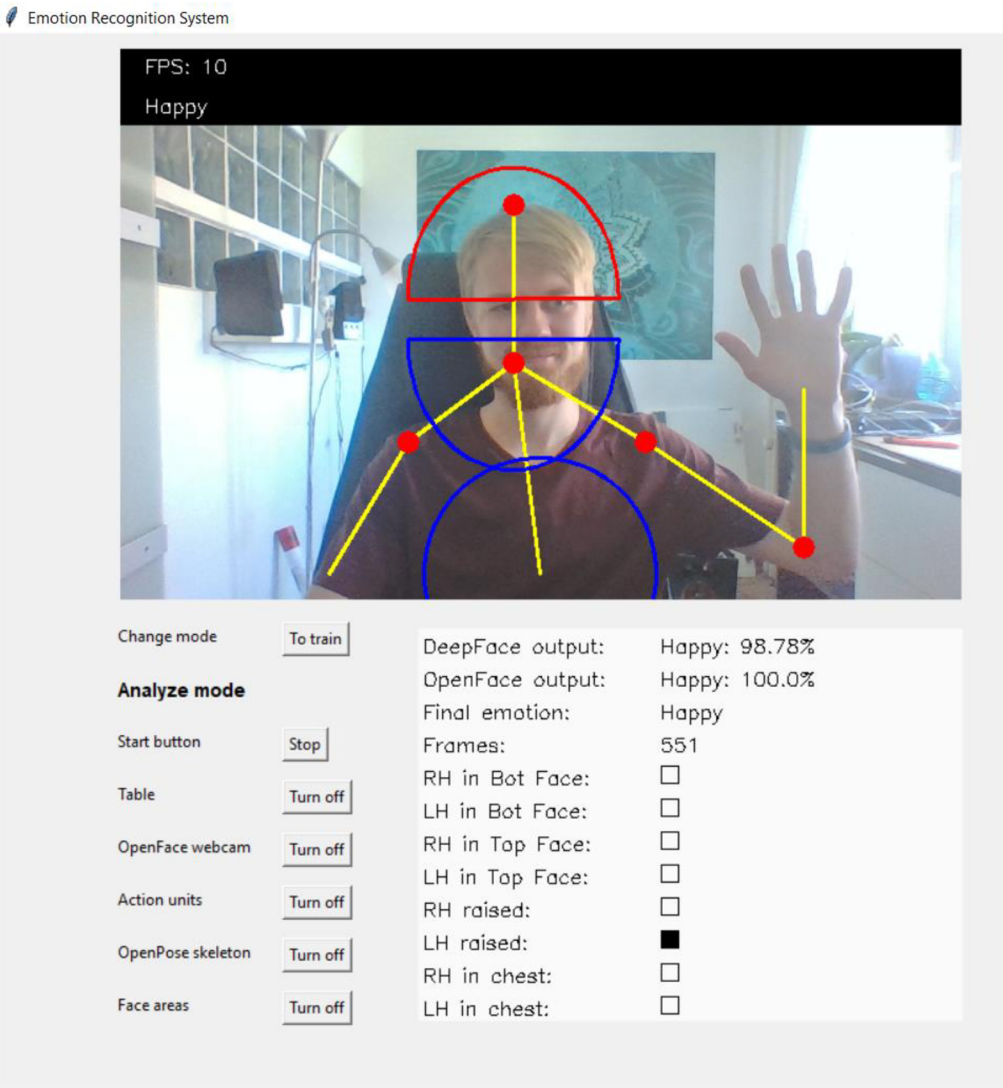
- funkce zjišťuje, zda se bod reprezentující zápěstí nachází v některé z definovaných oblastí, na výše uvedeném obrázku funkce vyhodnotí v oblasti horní části obličeje levé zápěstí
- definované oblasti jsou vypočítávány na základě výstupních bodů detekovaných pomocí MPII modelu
- oblasti mění své proporce, případně natočení na základě polohy jednotlivých bodů – například půl elipsy kopírují náklon hlavy, je-li nakloněna hlava, dojde i k náklonu obou půl elips
- výstupem funkce je pole binárních hodnot, které reprezentují jednotlivé pozice rukou, např. pravá ruka v horní části obličeji – ANO/NE
- proměnná *visButtState* slouží pro zapnutí či vypnutí vykreslování definovaných oblastí

Přestože byl pro diplomovou práci zvolen model MPII, který by měl být nejrychlejší, detekce bodů skeletu byla velice výpočetně náročná. Analýza dat z webkamery za použití pouze výpočetního výkonu procesoru byla téměř nemožná. Přestože byly parametry modelu nastaveny na menší přesnost detekce bodů, nebylo dosaženo vyšší rychlosti než 2 snímky za sekundu. Pro analýzu tedy bylo nutné využít výpočetní výkon grafické karty. S grafickou kartou už bylo dosaženo lepších výsledků a analýza byla o poznání plynulejší. Bohužel však výkon grafické karty nestačil k tomu, aby bylo možné detekovat jednotlivá gesta ruky. Pro detekování jednotlivých prstů nebyl výpočetní výkon karty dostačující. Navrhovaný emoční systém tedy pro zpřesnění detekované emoce využíval pouze pozice paží.

3.4 Návrh Uživatelského rozhraní

Pro snadné používání systému bylo nutné vytvořit jednoduché uživatelské rozhraní. Pro realizaci rozhraní v Pythonu byl zvolen modul Tkinter, který slouží jako rozhraní pro sadu nástrojů Tcl/Tk, které představují multiplatformní knihovnu základních prvků grafického uživatelského rozhraní. Díky modulu Tkinter je možné v Pythonu jednoduše vytvářet okna, přidávat elementy jako jsou tlačítka či výběrové listy, vykreslovat ve vrstvách různé grafické vstupy, dávat výzvy uživateli v podobě informačních oken atd.

Aplikace pro systém rozpoznávání emocí byla rozdělena na dva módy. Uživateli je umožněno si zvolit mód „Analýzy“, při kterém probíhá analýza dat z webkamery a uživateli jsou zobrazovány příslušné výstupy. Případně je možné se přepnout do „Trénovacího“ módu, kde lze emoční systém přeučit na výrazy konkrétní osoby. Podrobnější popis bude uveden dále v textu, na následujícím obrázku je zobrazena probíhající detekce emoce v režimu analýzy.

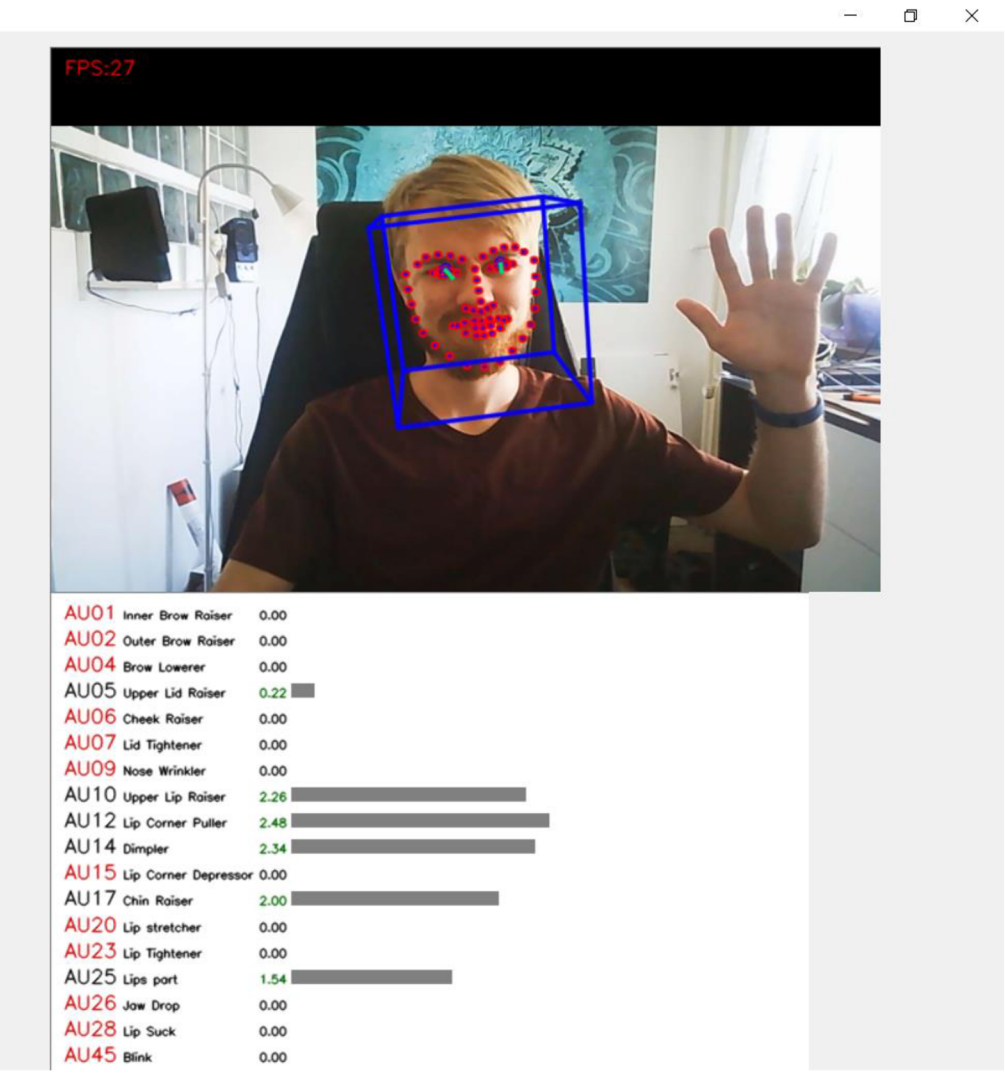


Obrázek 32 - Probíhající detekce emoce v režimu analýzy

V horní části obrázku je uživateli zobrazován výstup z hlavní webkamery. Tento záznam je používán pro analýzu pomocí DeepFace a OpenPose. Uživatel je informován o aktuálním počtu snímků za sekundu a výsledné emoci, která je dána kombinací výstupů všech tří metod, tedy DeepFace, OpenFace a OpenPose. V případě volby je výstup z webkamery doplněn o visualizaci skeletu a definovaných oblastí pro detekci pozice rukou.

Uživateli je umožněno aplikaci ovládat pomocí tlačítek, které se nacházejí v levé části pod výstupem z webkamery. Tlačítko „Change mode“, které se nachází nahoře slouží pro přepínání mezi oběma módy. Pod tlačítkem se nachází tučný text, informující uživatele o tom, v jakém módu se právě nachází. Následuje tlačítko „Start button“, které slouží pro spuštění a zastavení analýzy. Tlačítko „Table“ slouží pro zobrazení nebo skrytí výstupní tabulky, která se nachází vpravo od tlačítek. V tabulce jsou uživateli zobrazovány informace o dominantní emoci a její procentuální pravděpodobnosti modelu DeepFace

i OpenFace. Dále tabulka zobrazuje finální predikovanou emoci, která je dána kombinací všech 3 systémů, jedná se tedy o stejný parametr, který se vypisuje v černém rámečku na výstupu z webkamery. Dále je v tabulce zobrazen počet analyzovaných snímku od spuštění analýzy a pozice rukou. Systém umí rozpoznávat jsou-li ruce v oblasti hrudi, hlavy (horní nebo dolní část), případně je-li ruka zdvižena. Pokud je ruka v dané pozici, je daná situace indikována černým podbarvením čtverečku. V opačném případě zůstává čtvereček prázdný. Tlačítkem „OpenFace webcam“ si může uživatel zobrazit výstup z druhé webkamery, která se používá pro analýzu pomocí OpenFace. Tento výstup také zobrazuje sledované body obličeje a predikovaný směr pohledu očí či natočení obličeje. K zobrazení detekovaných akcí obličeje slouží tlačítko „Action units“. Aktivací tlačítka je uživateli zobrazena tabulka jednotlivých akcí obličeje AU01 až AU45. Černou barvou je podbarvená aktivní akce obličeje. Ta je doplněna i o informaci o její intenzitě, která je ve formě čísla a horizontálního sloupcového grafu. Červenou barvou je podbarvená neaktivní akce obličeje.

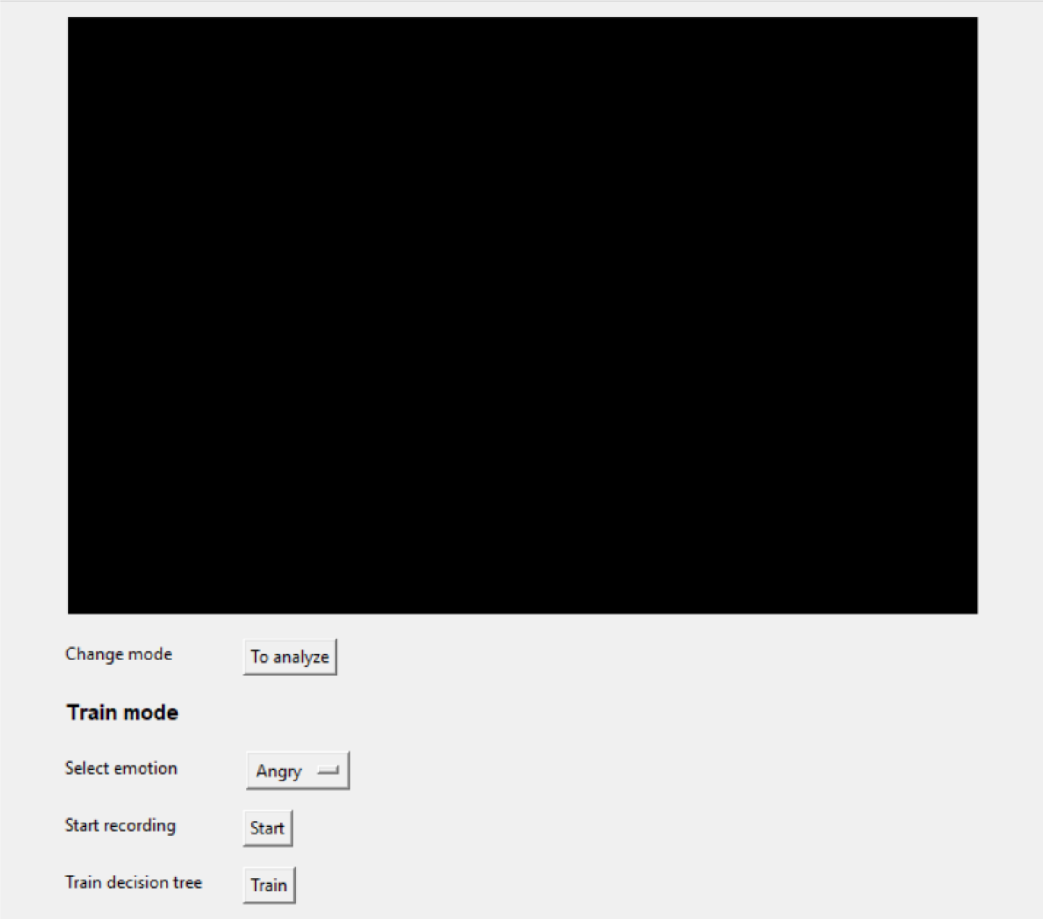


Obrázek 33 - Probíhající detekce emoce v režimu analýzy 2

Pomocí tlačítek „OpenPose skeleton“ a „Face areas“ je uživateli umožněno zobrazit nebo skrýt vizualizaci skeletu a detekovaných bodů člověka a definované oblasti hrudníku a hlavy.

Po přepnutí do trénovacího režimu se změní tlačítka, která jsou uživateli zobrazována, uspořádání oken pro zobrazování výstupů z webkamery zůstává stejné. Trénovací režim slouží k přeučení rozhodovacího stromu, který je využíván pro predikci emoce pomocí OpenFace. Vzhledem k časové náročnosti není realizováno přeučení neuronových sítí pro DeepFace ani OpenPose. Přeučení stromu nemá na DeepFace a OpenPose žádný vliv.

Přeučení na výrazy konkrétní osoby funguje na principu, že je předem vybrána emoce, kterou bude daná osoba předvádět. Pro snímky předváděné emoce je provedena pomocí OpenFace analýza detekovaných akcí obličeje a výstup OpenFace je zapsán do CSV souboru společně s textovou informací o předváděné emoci. Pro přetrénování stromu je nastaven sběr 250 snímků, kdy 200 snímků je použito pro trénování a 50 snímků pro testování naučeného stromu. Sběr dat trvá při snímkovací frekvenci 25-30 FPS přibližně 10 vteřin. Přeučení všech sedmi emocí na daného člověka je tedy proces trvající pouze pár minut, uvažuje-li se i obsluha tlačítek a čtení dialogových oken. Ukázku grafického rozhraní pro trénovací mód lze vidět na následujícím obrázku.



Obrázek 34 - Grafické rozhraní pro trénovací mód

Proces učení je velice jednoduchý, uživatel zvolí předváděnou emoci pomoci výběrového listu a stiskne tlačítko „Start recording“. V případě prvního nahrávání dané emoce je ještě dialogovým oknem informován, že musí vytvořit novou složku, kde bude nahrán trénovací a testovací CSV soubor. Poté je zahájen sběr snímků, který je ukončen po přibližně 10 vteřinách. V případě, že uživatel ukončí nahrávání výrazů, je možné pomocí tlačítka „Train decision tree“ přetrénovat rozhodovací strom. Po stisknutí tlačítka je nutné, aby uživatel zvolil CSV soubory s daty, na které chce strom přetrénovat. Po přetrénování je aplikace připravena na analýzu s nově naučeným rozhodovacím stromem.

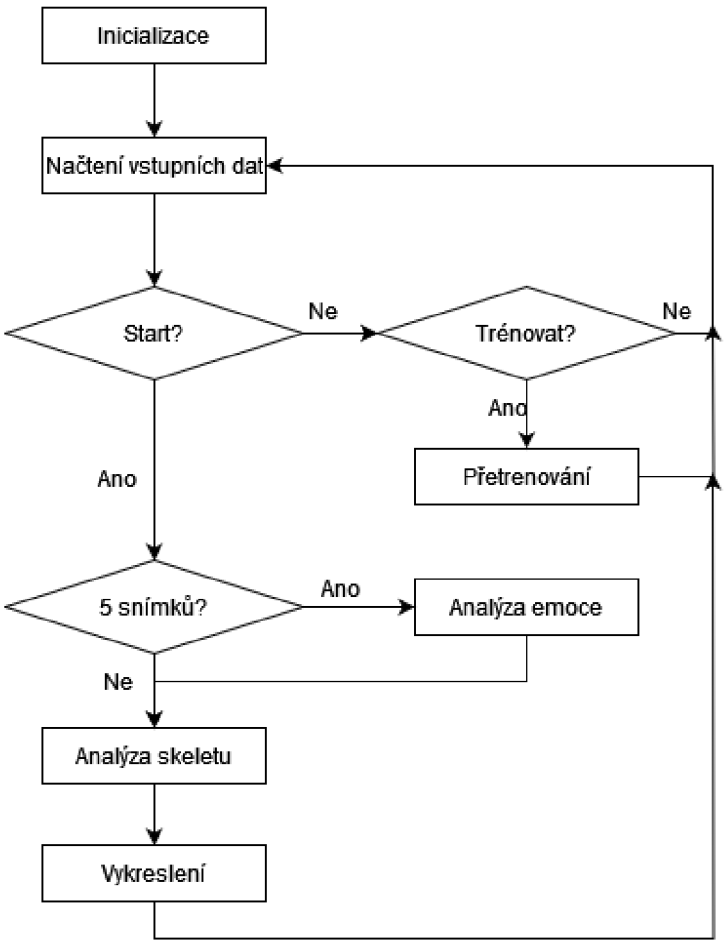
3.5 Zdrojový soubor

Zdrojový soubor obsahuje hlavní kód aplikace, která spojuje všechny výše zmíněné soubory. Zjednodušený diagram algoritmu je na následujícím obrázku. Prvním bodem algoritmu je inicializace systému. Při ní jsou definovány příslušné proměnné, realizován proces učení rozhodovacího stromu, načten model OpenPose, spuštěna hlavní webkamera a také spuštěna analýza pomocí druhé webkamery a OpenFace. Inicializace je ukončena v moment, kdy je detekován v příslušné složce CSV soubor s výstupními daty OpenFace. Inicializace je provedena vždy jen jednou při spuštění aplikace, poté je vykonáván kód v nekonečné smyčce.

Načtení vstupních dat reprezentuje načtení stavů tlačítek z grafického rozhraní a načtení snímku z webkamery. V případě, že není spuštěna analýza emoce, nejsou data z webkamery vykreslována. Je-li požadavek na přetrénování rozhodovacího stromu, je proces učení vykonán a objekt naučeného stromu je připraven pro emoční analýzu. V případě startu analýzy je spuštěna i analýza skeletu pomocí OpenPose. Jelikož jsou data z OpenPose vykreslovány přímo na výstup webkamery, provádí se analýza pomocí OpenPose v každé iteraci. Samotná analýza emoce pomocí OpenFace a DeepFace je prováděna jednou za 5 snímků. Z důvodu časové náročnosti je pomocí DeepFace vyhodnocován pouze každý pátý snímek. Analýza pomocí OpenFace je realizována pro všech 5 předchozích snímků. Emoce jsou predikovány ze všech 5 snímků a jako výsledná emoce je vybrána emoce s nejvyšším výskytem. Na základě detekovaných dat pomocí OpenPose a výstupů OpenFace a DeepFace je určena finální emoce, která je aktualizována společně s informací o počtu FPS, jednou za 5 snímků. Na rozdíl od vyhodnocování statických snímků datasetu AffectNet na Tabulka 15, je zde finální emoce vyhodnocena jako emoce s nejvyšším „skóre“. Jak již bylo několikrát zmíněno, DeepFace mimo emoci vyhodnocuje i její pravděpodobnost. Pravděpodobnost emoce lze teď vyhodnotit i pro OpenFace, jelikož je vyhodnocována sekvence několika snímků. Jednotlivé pravděpodobnosti jsou použity jako „skóre“, které se využívá pro určení výsledné emoce. V případě, že například DeepFace detekuje dominantní emoci „smutek“ s pravděpodobností 65 %, ale OpenFace bude vracet emoci neutrální s 80% pravděpodobností, bude jako výsledný výraz označena emoce „neutrální“. Do výsledného

„skóre“ se propisuje i výstup z OpenPose. Jsou-li například detekovány ruce v dolní oblasti obličeje, je velice pravděpodobná emoce „překvapení“ nebo „strach“. OpenPose tedy zvýší „skóre“ pro obě zmíněné emoční třídy a ovlivní tím výslednou emoci systému. Podobným způsobem jsou řešeny i jiné případy detekce rukou.

Poslední fází smyčky je vykreslení, která vizualizuje grafické výstupy na základě načtených vstupních dat. Zobrazení nebo skrytí grafických výstupů si uživatel může volit pomocí stisku příslušných tlačítek na grafickém uživatelském rozhraní.



Obrázek 35 - Blokový diagram zdrojového kódu

Je-li stisknut křížek pro zavření hlavní aplikace, nekonečná smyčka je ukončena, ať se nachází v kterémkoliv stavu. V případě ukončení je bezpečně ukončen proces OpenFace, který běžel na pozadí, stejně tak je i uvolněna webkamera a jsou zavřeny všechny okna aplikace.

3.6 Dosažené výsledky

Pro otestování sytému nebyl nalezen vhodný dataset, který by obsahoval zároveň obličej s projevovanou emocí a tělo člověka pro analýzu skeletu. Navíc je systém navržen pro

vyhodnocovaní emoce z webkamery, tedy ze sekvence snímků, nikoliv pro statické obrázky. Pro finální otestování byly tedy osloveni konkrétní lidé, na kterých bylo testováno rozpoznávání všech sedmi emocí. Každá osoba byla vždy vyzvána, aby projevila danou emoci výsledek systému byl zaznamenán. Je nutné podotknout, že oslovené osoby nejsou profesionální herci, projevované emoce tedy nejsou 100% autentické, což se může projevit na finálních výsledcích. Každá osoba byla hodnocena 2x. Poprvé bylo testování provedeno s použitím obecného rozhodovacího stromu naučeného na datasetu AffectNet. U neúspěšně detekovaných emocí bylo provedeno přeučení a druhé testování bylo provedeno s využitím přeučeného rozhodovacího stromu. Výslednému testování bylo podrobeno 23 osob. Nastávaly situace, kdy systém nemohl spolehlivě rozpoznat konkrétní emoci, ale neustále svůj výstup měnil, ačkoliv analyzovaný výraz byl stále stejný. V tomto případě nebyla zaznamenána nesprávně predikovaná emoce, ale výsledek byl vyhodnocen pouze jako nesprávná klasifikace. Podrobné výsledky prvního testování jsou na následující matici záměn.

Tabulka 16 - Výsledky na reálných výrazech osob (před učením)

Celková správnost 42,06 %		**Skutečnost**								
		Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení	Celkem	Přesnost
Predikce	Zlost	2	0	2	0	0	0	0	4	50 %
	Znechucení	3	3	1	0	0	0	0	7	43 %
	Strach	4	5	8	1	1	4	6	29	28 %
	Štěstí	3	3	1	21	1	1	4	34	62 %
	Neutrální	5	4	3	1	18	10	2	43	42 %
	Smutek	5	7	4	0	3	6	2	27	22 %
	Překvapení	0	0	4	0	0	1	8	13	62 %
	Celkem	22	22	23	23	23	22	22		
	Senzitivita	9 %	14 %	35 %	91 %	78 %	27 %	36 %		
	Správnost	86 %	85 %	77 %	90 %	81 %	76 %	88 %		
	F1 míra	15 %	21 %	31 %	74 %	55 %	24 %	46 %		

Celková správnost klasifikace oproti Tabulka 14 poklesla přibližně o 8,5 %. Co se týče F1 míry pro jednotlivé emoční třídy, nejvyšší hodnoty bylo dosaženo pro emoci „štěstí“.

U 21 případů z 23 se podařilo emoci „štěstí“ správně klasifikovat a hodnota F1 míry dosahovala 74 %. Oproti hodnoty F1 míry na Tabulka 14 je to nižší hodnota, a to o 7 %.

Systém byl poměrně senzitivní na emoci „štěstí“, nicméně hodnotu F1 míry negativně ovlivnila nízká hodnota přesnosti. Ta dosahovala pouhých 62 %. Za slušný výsledek je považováno také rozpoznání emoce „neutrální“ s hodnotou F1 míry 55 % a emoce „překvapení“ s hodnotou 46 %. Ostatní emoce byly pro systém obtížně rozpoznatelné, například emoce „zlost“ a „znechucení“ systém nedokázal rozpoznat téměř vůbec. Obtížně rozpoznával i emoce „strach“ a „smutek“.

Důvodem pro nefungování systému pro výše zmíněné emoce je pravděpodobně ten, že rozhodovací strom byl naučen na výrazy datasetu AffectNet. Ty zachycují poměrně intenzivně projevované emoce, které lidé v dané situaci lidé vyjadřují. Testované osoby však ve většině případů projevily danou emoci jen s velmi nízkou intenzitou. Výsledek tedy byl, že výrazy pro různé emoce byly téměř shodné, nebo se lišily pouze v drobných nuancích, které naučený rozhodovací strom nebyl schopný rozeznat. Problém byl také ten, že každá osoba projevuje emoce jiným způsobem. Univerzální systém se tedy nejevil jako nejlepší řešení.

Pro neúspěšně klasifikované emoce bylo tedy provedeno přeučení. Výsledky systému po přeučení jsou shrnuty v následující matici záměn.

Tabulka 17 - Výsledky na reálných výrazech osob (po učení)

Celková správnost 62,99 %		**Skutečnost**								
		Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení	Celkem	Přesnost
Predikce	Zlost	13	1	2	0	0	1	0	17	76 %
	Znechucení	1	12	0	2	1	1	0	17	71 %
	Strach	5	5	17	1	2	5	4	39	44 %
	Štěstí	1	0	0	20	2	1	2	26	77 %
	Neutrální	2	1	0	0	14	8	1	26	54 %
	Smutek	0	1	0	0	2	7	1	11	64 %
	Překvapení	0	1	3	0	0	0	14	18	78 %
	Celkem	22	21	22	23	21	23	22		
	Senzitivita	59 %	57 %	77 %	87 %	67 %	30 %	64 %		
	Správnost	92 %	91 %	82 %	94 %	88 %	87 %	92 %		
	F1 míra	67 %	63 %	56 %	82 %	60 %	41 %	70 %		

Naučením bylo dosaženo zlepšení celkové správnosti klasifikace téměř o 21 %. Systém nyní mnohem lépe rozpoznával emoce „zlost“ i „znechucení“, které se v předešlém případě nedařilo klasifikovat. V obou případech byla hodnota F1 míry vyšší

než 60 %. Zlepšení bylo dosaženo i pro ostatní emoční třídy, u všech byla hodnota F1 míry vyšší než v předcházejícím případě. Mimo emoci „smutek“ bylo dosaženo hodnoty F1 míry vyšší než 55 %. Rozpoznávání emoce „smutek“ bylo poměrně obtížné z důvodu, že byla velice podobně projevována jako neutrální výraz. OpenFace extrahovalo velice podobné příznaky a rozhodovací strom tedy nemohl být pro rozlišování těchto dvou emocí dobře naučen.

Téměř u všech případů však došlo ke zlepšení klasifikace po přeučení. Míra zlepšení byla tím vyšší, čím odlišnější byly projevované emoce. Nejlepších výsledků bylo tedy dosaženo u osob, které emoce vyjadřovali intenzivně a jednotlivé výrazy se od sebe lišily. Nasnímaná data pak vykazovala různé hodnoty pro detekované akce obličeje, a to vedlo ke spolehlivějšímu naučení rozhodovacího stromu. U osob, které emoce vyjadřovaly podobným způsobem, k výraznému zlepšení klasifikace po přeučení nedocházelo. Podrobné vyhodnocení klasifikace emocí pro jednotlivé osoby je uvedeno v Příloha A - Vyhodnocení funkčnosti systému.

4. ZÁVĚR

V rámci diplomové práce byla provedena rešerše v oblasti rozpoznávání emocí. Rešerše se zaměřovala především na využití rozpoznávání emocí v praxi, metody získávání příznaků, knihovní funkce a algoritmy pro emoční rozpoznávání a v neposlední řadě na hledání vhodného datasetu. Pro emoční rozpoznávání bylo nalezeno několik možných řešení, jako nejlepší metody byly vyhodnoceny DeepFace a OpenFace, které obě pracují na odlišném principu. DeepFace obsahuje předučený emoční model, který pomocí konvoluční neuronové sítě vyhodnocuje nejpravděpodobnější emoci. Naopak OpenFace pouze detekuje akce obličeje, dle systému FACS. Vyhodnocení probíhalo na datasetech FER2013 a AffectNet a konkrétní výsledky jsou uvedeny v kapitolách 2.2 a 2.3. Jako vhodnou metodou pro zpřesnění klasifikace emoce byla určena i knihovna OpenPose, která se dá použít pro detekci lidského skeletu.

Pro řešení s OpenFace byl navržen klasifikátor v podobě rozhodovacího stromu fungujícím na optimalizovaném CART algoritmu. Rozhodovací strom byl naučen na trénovací množině datasetu AffectNet. Pro testování kvality klasifikace byla z důvodu malé testovací množiny datasetu AffectNet provedena její augmentace. Na augementovaných datech byl otestován OpenFace i DeepFace a výsledky jsou uvedeny v Tabulka 13 a Tabulka 14. Bylo zjištěno, že spojením výstupů obou metod lze dosáhnout lepších výsledků. Konkrétně bylo dosaženo celkové správnosti 58,79 % a výsledky jsou uvedeny v Tabulka 15. Pro ještě větší zpřesnění klasifikace systému byla použita detekce pozic rukou pomocí OpenPose.

Jako vyhodnocovací prostředek byl zvolen notebook s dedikovanou grafickou kartou. Spojení metod bylo realizováno v programovacím jazyce Python a pro systém bylo navrženo i jednoduché uživatelské rozhraní. Uživatelské rozhraní umožňovalo jednoduše spustit záznam z webkamery a analyzovat emoci snímané osoby. V případě neúspěšně klasifikované emoce bylo možné systém adaptovat na konkrétní výrazy osoby. Tato funkce byla realizována přeučením rozhodovacího stromu.

Výsledky systému byly otestovány na reálných výrazech 23 osob, které vždy na výzvu předvedli požadovanou emoci. Pro naučený rozhodovacího strom ve výchozím nastavení bylo dosaženo celkové správnosti klasifikace 42,06 %, podrobné výsledky jsou v Tabulka 16. Díky přeučení rozhodovacího stromu na konkrétní osobu bylo dosaženo výrazného zlepšení a celková správnost se zlepšila na 62,99 %. Podrobné výsledky jsou v Tabulka 17. Proces získávání dat pro učení a samotný proces učení byl velice rychlý a uživatelsky přívětivý.

LITERATURA

[1] EKMAN, Paul, Wallace V. FRIESEN a Joseph C. HAGER. *Facial Action Coding System: The manual*. United States of America: Research Nexus division of Network Information Research Corporation, 2002. ISBN 0-931835-01-1.

[2] FARNSWORTH, Bryn, 2022. Facial Action Coding System (FACS) – A Visual Guidebook. *Imotions* [online]. Denmark: iMotions [cit. 2022-10-26]. Dostupné z: https://imotions.com/blog/facial-action-coding-system/#emotions-action-units

[3] A. CLARK, Elizabeth, 2020. The Facial Action Coding System for Characterization of Human Affective Response to Consumer Product-Based Stimuli: A Systematic Review. *Frontiersin* [online]. United States, 2020, 16 [cit. 2022-10-26]. Dostupné z: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00920/full#SM1

[4] PARVEZ, Hanan. Fear facial expression analyzed. *Psychmechanics* [online]. 2020 [cit. 2022-10-26]. Dostupné z: https://www.psychmechanics.com/facial-expressions-fear/

[5] COLUMBUS, Chris a John HUGHES. *Home Alone*. United States of America: 20th Century Studios, 1990.

[6] SERENGIL, Sefik Ilkin a Alper OZIPINAR. A Hybrid Deep Face Recognition Framework. *Innovations in Intelligent Systems and Applications Conference* [online]. 2020, 1-5 [cit. 2022-10-27]. Dostupné z: doi:10.1109/ASYU50717.2020.9259802

[7] SHERRAH, J. a Shaogang GONG. *Fusion of 2D face alignment and 3D head pose estimation for robust and real-time performance* [online]. 1999, 1-5 [cit. 2022-10-27]. Dostupné z: doi:10.1109/RATFG.1999.799219

[8] SERENGIL, Sefik Ilkin. *Face Alignment for Face Recognition in Python within OpenCV* [online]. 2020 [cit. 2022-10-31]. Dostupné z: https://sefiks.com/2020/02/23/face-alignment-for-face-recognition-in-python-within-opencv/

[9] TAIGMAN, Yaniv, Ming YANG, Marc'Aurelio RANZATO a Lior WOLF. *Closing the Gap to Human-Level Performance in Face Verification* [online]. IEEE, 2014 [cit. 2022-10-31]. Dostupné z: doi:10.1109/CVPR.2014.220

[10] ÁLVAREZ, Tino. Faster, smoother, smaller, more accurate and more robust face alignment models on CPU. *Towards Data Science* [online]. 2021 [cit. 2022-11-01]. Dostupné z: https://towardsdatascience.com/faster-smoother-smaller-more-accurate-and-more-robust-face-alignment-models-d8cc867efc5

[11] *MorphCast: Showcase* [online]. [cit. 2022-11-02]. Dostupné z: https://www.morphcast.com/showcase/

[12] SAXENA, Parul. *Real-time emotion recognition: Potential use cases and challenges* [online]. 2021 [cit. 2022-11-02]. Dostupné z: https://indiaai.gov.in/article/real-time-emotion-recognition-potential-use-cases-and-challenges

[13] *Movel* [online]. [cit. 2022-11-02]. Dostupné z: https://movel-service.webflow.io/
[14] *Emotional recognition technology enters recruitment* [online]. 2018 [cit. 2022-11-09]. Dostupné z: https://privacyinternational.org/examples/1971/emotional-recognition-technology-enters-recruitment
[15] LIU, Xin, Henglin SHI, Haoyu CHEN, Zitong YU, Xiaobai LI a Guoying ZHAO. *IMiGUE: An Identity-free Video Dataset for Micro-Gesture Understanding and Emotion Analysis* [online]. 1-8 [cit. 2022-11-14]. Dostupné z: https://openaccess.thecvf.com/content/CVPR2021/papers/Liu_iMiGUE_An_Identity-Free_Video_Dataset_for_Micro-Gesture_Understanding_and_Emotion_CVPR_2021_paper.pdf
[16] LIU, Xin, Henglin SHI, Haoyu CHEN, Zitong YU, Xiaobai LI a Guoying ZHAO. *IMiGUE database* [online]. Finsko: Univerzita v Oulu [cit. 2022-11-14]. Dostupné z: https://www.oulu.fi/en/university/faculties-and-units/faculty-information-technology-and-electrical-engineering/center-machine-vision-and-signal-analysis#accordion-control-imigue-database
[17] LIU, Xin, Henglin SHI, Haoyu CHEN, Zitong YU, Xiaobai LI a Guoying ZHAO. *IMiGUE: GitHub* [online]. Finsko: Univerzita v Oulu [cit. 2022-11-14]. Dostupné z: https://github.com/linuxsino/iMiGUE
[18] BALTRUSAITIS, Tadas, Amir ZADEH, Yao Chong LIM a Louis-Philippe MORENCY. *OpenFace 2.0: Facial Behavior Analysis Toolkit* [online]. Xi'an, Čína, 2018 [cit. 2022-11-29]. Dostupné z: doi:10.1109/FG.2018.00019
[19] BALTRUSAITIS, Tadas, Amir ZADEH, Yao Chong LIM a Louis-Philippe MORENCY. *OpenFace: GitHub* [online]. 2021 [cit. 2022-11-29]. Dostupné z: https://github.com/TadasBaltrusaitis/OpenFace
[20] SERENGIL, Sefik Ilkin. *DeepFace: GitHub* [online]. 2022 [cit. 2022-11-29]. Dostupné z: https://github.com/serengil/DeepFace
[21] *Challenges in Representation Learning: Facial Expression Recognition Challenge* [online]. 2013 [cit. 2022-11-29]. Dostupné z: https://www.kaggle.com/competitions/challenges-in-representation-learning-facial-expression-recognition-challenge/data
[22] MOLLAHOSSEINI, Ali, Behzad HASANI a Mohammad H. MAHOOR. *AffectNet: A New Database for Facial Expression, Valence, and Arousal Computation in the Wild* [online]. In: . IEEE Transactions on Affective Computing, 2017 [cit. 2022-12-06].
[23] *AffectNet-HQ* [online]. [cit. 2022-12-06]. Dostupné z: https://www.kaggle.com/datasets/tom99763/affectnethq
[24] *Scikit-learn: Decision trees* [online]. [cit. 2022-12-29]. Dostupné z: https://scikit-learn.org/stable/modules/tree.html
[25] *Python - Decision tree implementation* [online]. [cit. 2022-12-29]. Dostupné z: https://www.geeksforgeeks.org/decision-tree-implementation-python/

[26] ZEPF, Sebastian, 2020. Driver Emotion Recognition for Intelligent Vehicles. *ACM Computing Surveys* [online]. **2021**(64), 11 [cit. 2022-12-30]. Dostupné z: doi:10.1145/3388790

[27] *Facial Expression Recognition with Keras* [online]. 2018 [cit. 2022-12-30]. Dostupné z: https://sefiks.com/2018/01/01/facial-expression-recognition-with-keras/

[28] *MorphCast demo* [online]. [cit. 2022-12-30]. Dostupné z: https://demo.morphcast.com/sdk-features/index.html?video=https%3A%2F%2Fdemo.morphcast.com%2Fsdk-features%2FBreeze_Woodson.mp4&sv=false&cta=vp

[29] *OpenPose* [online]. [cit. 2022-12-30]. Dostupné z: https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_02_output.html

[30] MANDIC, Vladimir. *Human* [online]. [cit. 2023-02-09]. Dostupné z: https://github.com/vladmandic/human

[31] BALAJI, Atul. *Emotion Detection* [online]. [cit. 2023-02-09]. Dostupné z: https://github.com/atulapra/Emotion-detection

[32] FABIEN, Maël, Anatoli DE BRADKE a Ayan SAHA. *Multimodal Emotion Recognition* [online]. [cit. 2023-02-09]. Dostupné z: https://github.com/maelfabien/Multimodal-Emotion-Recognition

[33] AGARWAL, Vardan. Complete Image Augmentation in OpenCV. *Towardsdatascience* [online]. [cit. 2023-04-09]. Dostupné z: https://towardsdatascience.com/complete-image-augmentation-in-opencv-31a6b02694f5

[34] *Python documentation: Subprocess management* [online]. [cit. 2023-04-09]. Dostupné z: https://docs.python.org/3/library/subprocess.html

[35] *Python documentation: CSV File Reading and Writing* [online]. [cit. 2023-04-09]. Dostupné z: https://docs.python.org/3/library/csv.html

[36] GUPTA, Vikas. Deep Learning based Human Pose Estimation using OpenCV. *LearnOpenCV* [online]. [cit. 2023-04-18]. Dostupné z: https://learnopencv.com/deep-learning-based-human-pose-estimation-using-opencv-cpp-python/

[37] MUNEA, Tewodros Legesse, Yalew Zelalem JEMBRE, Halefom Tekle WELDEGEBRIEL, Longbiao CHEN, Chenxi HUANG a Chenhui YANG. *The Progress of Human Pose Estimation* [online]. [cit. 2023-04-19]. Dostupné z: doi:10.1109/ACCESS.2020.3010248

[38] PEASE, Allan a Barbara PEASE, 2004. *The definitive book of body languague*. Austrálie: Pease international. ISBN 1-9208160-7-0.

[39] CHELLIAH, Indhumathy. *Confusion Matrix for Multiclass Classification* [online]. MLearning.ai [cit. 2023-05-03]. Dostupné z: https://medium.com/mlearning-ai/confusion-matrix-for-multiclass-classification-f25ed7173e66

[40] *Microsoft stock images: Cutout people* [online]. Microsoft 365 [cit. 2023-05-13]. Dostupné z: https://support.microsoft.com/en-us/office/insert-images-icons-and-more-in-microsoft-365-c7b78cdf-2503-4993-8664-851085c30fce

SEZNAM PŘÍLOH

Příloha A - Vyhodnocení funkčnosti systému

A.1 Výsledky systému před trénováním

Osoba	Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení
Kuba	Strach	-	Strach	Štěstí	Neutrální	Smutek	-
Niki M.	Strach	Smutek	Smutek	Štěstí	Smutek	Strach	Překvapení
Péťa	Znechucení	Znechucení	Smutek	Štěstí	Smutek	Smutek	Neutrální
Ondra	Zlost	Smutek	Strach	Štěstí	Neutrální	Neutrální	Strach
Zuzka	Smutek	Smutek	Strach	Štěstí	Neutrální	-	Štěstí
Jája	Smutek	Smutek	Strach	Štěstí	Smutek	Smutek	Strach
Jakub	Smutek	Strach	Strach	Štěstí	Neutrální	Neutrální	Strach
Marek	-	Neutrální	Neutrální	Štěstí	Neutrální	Neutrální	Štěstí
Babička	Neutrální	Strach	Neutrální	Štěstí	Strach	Smutek	Překvapení
Benko	Štěstí	Znechucení	Smutek	Štěstí	Štěstí	Smutek	Překvapení
Taťka	Strach	Neutrální	Překvapení	Štěstí	Neutrální	Neutrální	Smutek
Julča	Znechucení	Neutrální	Neutrální	Štěstí	Neutrální	Strach	Štěstí
Já	Neutrální	Znechucení	Překvapení	Štěstí	Neutrální	Strach	Překvapení
Niki	Neutrální	Štěstí	Překvapení	Štěstí	Neutrální	Neutrální	Překvapení
Janka	Neutrální	Smutek	Strach	Štěstí	Neutrální	Neutrální	Neutrální
Honza	Smutek	Strach	Překvapení	Štěstí	Neutrální	Překvapení	Překvapení
Klárka	Smutek	Smutek	Štěstí	Neutrální	Neutrální	Strach	Smutek
Filip	Zlost	Smutek	Strach	Štěstí	Neutrální	Neutrální	Překvapení
Adam	Štěstí	Štěstí	Znechucení	Štěstí	Neutrální	Smutek	Štěstí
Sofie	Strach	Štěstí	Zlost	Strach	Neutrální	Neutrální	Strach
Péťa	Znechucení	Strach	Zlost	Štěstí	Neutrální	Neutrální	Strach
Florian	Štěstí	Strach	Strach	Štěstí	Neutrální	Štěstí	Strach
Alenka	Neutrální	Neutrální	Smutek	Štěstí	Neutrální	Neutrální	Překvapení

A.2 Výsledky systému po trénování

Osoba	Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Překvapení
Kuba	Strach	-	Strach	Štěstí	Neutrální	Smutek	-
Niki M.	Zlost	Strach	Překvapení	Štěstí	Smutek	Smutek	Překvapení
Péťa	Zlost	Znechucení	Strach	Štěstí	Smutek	Smutek	Překvapení
Ondra	Zlost	Znechucení	Strach	Štěstí	Neutrální	Neutrální	Překvapení
Zuzka	Strach	Strach	Strach	Štěstí	Strach	Strach	Překvapení
Jája	Znechucení	Znechucení	Překvapení	Štěstí	Znechucení	Znechucení	Překvapení
Jakub	Zlost	Znechucení	Strach	Štěstí	Neutrální	Smutek	Strach
Marek	-	Překvapení	Strach	Štěstí	Neutrální	Strach	Překvapení
Babička	Zlost	Znechucení	Strach	Znechucení	Neutrální	Neutrální	Strach
Benko	Neutrální	Znechucení	Strach	Štěstí	Štěstí	Neutrální	Překvapení
Taťka	Zlost	Znechucení	Strach	Štěstí	Neutrální	Neutrální	Smutek
Julča	Neutrální	Neutrální	Zlost	Štěstí	Neutrální	Strach	Štěstí
Já	Zlost	Zlost	Strach	Štěstí	Neutrální	Neutrální	Překvapení
Niki	Zlost	Znechucení	Překvapení	Štěstí	Neutrální	Neutrální	Překvapení
Janka	Zlost	Znechucení	Strach	Štěstí	Neutrální	Strach	Překvapení
Honza	Strach	Strach	Strach	Štěstí	Neutrální	Smutek	Překvapení
Klárka	Zlost	-	Strach	Strach	-	Strach	Strach
Filip	Zlost	Znechucení	Strach	Štěstí	Štěstí	Smutek	Překvapení
Adam	Štěstí	Znechucení	Zlost	Znechucení	Strach	Neutrální	Neutrální
Sofie	Strach	Znechucení	-	Štěstí	Neutrální	Štěstí	Překvapení
Péťa	Strach	Strach	Strach	Štěstí	Neutrální	Smutek	Strach
Florian	Zlost	Strach	Strach	Štěstí	-	Zlost	Štěstí
Alenka	Zlost	Smutek	Strach	Štěstí	Neutrální	Neutrální	Překvapení

Příloha B - Elektronická příloha

Hlavní složka

- Dataset_scripts
  - augmentation.py
  - saveImgFromPPT.py
  - validationDeepface.py
  - validationDeepfaceTop2.py
- Emotion_Recognition_System
  - Csv
    - base
    - default
    - testing
  - Processed
  - decTree.py
  - gui.py
  - main.py
  - openFace.py
  - openPose.py
- Pictures
  - Decision_tree
  - Emotion_system

Podsložka „Dataset_scripts“ obsahuje Python skripty pro práci s datasety, jako je augmentace nebo validace. Složka „Emotion_Recognition_System“ obsahuje veškeré Python soubory nutné pro fungování systému. Dále obsahuje podsložku „csv“, která obsahuje trénovací a testovací CSV soubory. Podsložka „processed“ se používá pro uložení CSV souboru s výstupními daty OpenFace. Obrázky naučených rozhodovacích stromů a ukázky aplikace systému je možné najít ve složce „Pictures“.

Pro úspěšné spuštění systému je nutné mít k dispozici i modely a soubory OpenFace a OpenPose, které však překračují limit elektronických příloh. Nutné je také mít nainstalované příslušné Python knihovny a OpenCV DNN (deep neural network) modul pro využití grafické karty na výpočty OpenPose.