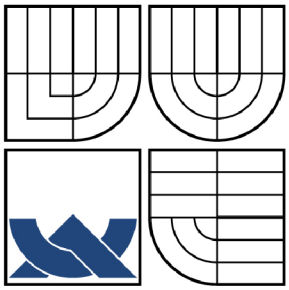
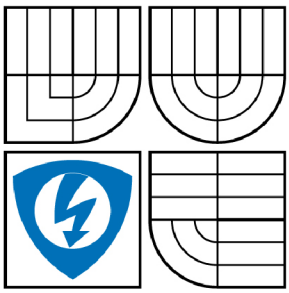


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

DEPARTMENT OF TELECOMMUNICATIONS

ANALÝZA EMOCIONÁLNÍCH STAVŮ NA ZÁKLADĚ OBRAZOVÝCH PŘEDLOH

EMOTIONAL STATE ANALYSIS UPON IMAGE PATTERNS

DOKTORSKÁ PRÁCE

DOCTORAL THESIS

AUTOR PRÁCE — Ing. JIŘÍ PŘINOSIL

AUTHOR

VEDOUCÍ PRÁCE — Mgr. PAVEL RAJMIC, Ph.D.

SUPERVISOR

BRNO 2008

Anotace

Tato disertační práce se zabývá návrhem automatického systému pro rozpoznávání základních emocionálních výrazů ve tváři ze statických obrazů. Celkově je tento systém rozdělen do tří nezávislých částí, které jsou ovšem jistým způsobem propojeny. Jedná se o automatickou detekci tváře v barevném obraze, kde byl navržen vlastní detektor pracující na základě barvy lidské kůže a metody pro lokalizaci pozic očí a rtů v již nalezených tvářích pomocí barevných map. Součástí tohoto celku je modifikovaný algoritmus detekce tváří *Viola-Jones*, který byl také experimentálně využit pro detekci očí. Spolehlivost těchto detektorů byla testována pomocí obrazové databáze obličejů *Georgia Tech Face Database*. Další částí automatického systému je extrakce příznaků skládající se ze dvou statistických metod a jedné metody založené na filtraci obrazu pomocí sady Gaborových filtrů. Pro účely této práce byly taky experimentálně vyzkoušeny různé kombinace příznaků extrahovaných pomocí těchto metod. Poslední částí systému pak je matematický klasifikátor reprezentován dopřednou neuronovou sítí. Celý systém je navíc doplněn o přesné určení pozic jednotlivých částí lidského obličeje pomocí aktivního modelu tvaru. Spolehlivost celého automatického systému byla testována na rozpoznání základních emocionálních výrazů ve tváři pomocí databáze *Japanese Female Facial Expression*.

Klíčová slova: emocionální výraz, detekce tváří, lokalizace očí, extrakce příznaků, klasifikace, aktivní model tvaru.

Abstract

This dissertation thesis deals with the automatic system for basic emotional facial expressions recognition from static images. Generally the system is divided into the three independent parts, which are linked together in some way. The first part deals with automatic face detection from color images. In this part they were proposed the face detector based on skin color and the methods for eyes and lips position localization from detected faces using color maps. A part of this is modified *Viola-Jones* face detector, which was even experimentally used for eyes detection. The both face detectors were tested on the *Georgia Tech Face Database*. Another part of the automatic system is features extraction process, which consists of two statistical methods and of one method based on image filtering using set of Gabor's filters. For purposes of this thesis they were experimentally used some combinations of features extracted using these methods. The last part of the automatic system is mathematical classifier, which is represented by feed-forward neural network. The automatic system is utilized by adding an accurate particular facial features localization using active shape model. The whole automatic system was benchmarked on recognizing of basic emotional facial expressions using the *Japanese Female Facial Expression* database.

Key words: emotional facial expression, face detection, eyes localization, features extraction, classification, active shape model.

Bibliografická citace této práce:

PŘINOSIL, J. *Analýza emocionálních stavů na základě obrazových předloh.* Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008. Vedoucí dizertační práce Mgr. Pavel Rajmic, Ph.D.

Prohlášení

Prohlašuji, že jsem svoji disertační práci na téma Analýza emocionálních stavů na základě obrazových předloh vypracoval samostatně pod vedením svého školitele s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v rámci práce a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené disertační práce dále prohlašuji, že v souvislosti s vytvořením této práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nepovoleným 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 §152 trestního zákona č. 140/1961 Sb.

V Brně dne …… ……

Podpis autora

Poděkování

Děkuji vedoucímu mé disertační práce Mgr. Pavlu Rajmicovi, Ph.D, za konstruktivní pomoc a cenné rady při zpracování disertační práce.

Seznam použitých zkratek

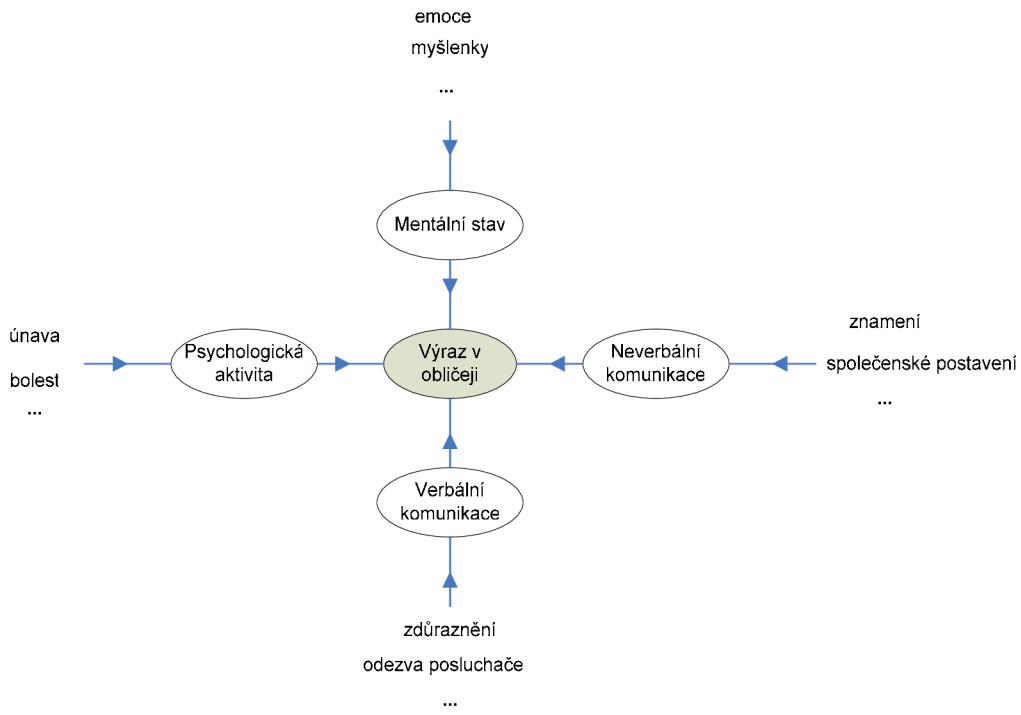
AAM	aktivní model vzhledu *(Active Appearance Model)*
ASM	aktivní model tvaru *(Active Shape Model)*
DWT	diskrétní waveletová transformace (*Discrete Wavelet Transform*)
FN	nesprávně určený negativní vzorek (*false negative*)
FP	nesprávně určený pozitivní vzorek (*false positive*)
FLDA	Fischerova lineární diskriminační analýza (*Fischer Linear Discriminant Analysis*)
GTFD	obrazová databáze tváří (*Georgia Tech Face Database*)
GW	Gaborovy wavelety (*Gabor Wavelets*)
HSV	barevný model skládající se barevného tónu (*Hue*) H, sytosti barvy S (*Saturation*) a hodnoty jasu V (*Value*)
JAFFE	obrazová databáze emocionálních výrazů (*Japanese Female Facial Expression*)
JPEG	komprimační formát pro digitální obrazy (*Joint Photographic Experts Group*)
k-NN	algoritmus k-nejbližších sousedů (*k-Nearest Neighbors*)
LDA	lineární diskriminační analýza (*Linear Discriminant Analysis*)
NN	neuronová síť (*Neural Network*)
PAL	televizní přenosový standard (*Phase Alternating Line*)
PCA	analýza hlavních komponent (*Principal Component Analysis*)
RGB	barevný model reprezentující aditivní míchání jednotlivých složek červené-zelené-modré barvy (*Red-Green-Blue*)
rgb	normalizovaný barevný model RGB
SVM	podpůrné vektory (*Support Vector Machine*)
TN	správně určený negativní vzorek (*true negative*)
TP	správně určený pozitivní vzorek (*true positive*)
YCbCr	barevný model skládající se z jasové složky Y, modrého a červeného chrominančního komponentu Cb a Cr

OBSAH

1 Úvod

V posledních desetiletích se značná část výzkumu v oblasti číslicového zpracování obrazových signálů zaměřila na detekci a rozpoznávání jednotlivých objektů ve statických obrazech s komplexním pozadím. Takovými objekty mohou být například i lidské tváře. Detekce a analýza lidských obličejů bývá především využívána v biometrických aplikacích a v bezpečnostních systémech zaměřených na verifikaci či identifikaci identity jednotlivých subjektů. V současné době je této oblasti věnována významná pozornost, a přestože problematika rozpoznávání obličejů při změnách vnějšího prostředí zůstává převážně nedořešena, byla navržena řada postupů pro vývoj robustního systému pro rozpoznávání obličejů [50]. Jednotlivé přístupy lze ovšem využít i pro jiné způsoby zpracování informací získaných z lidského obličeje. Jedním z nich může být i analýza emocionálních stavů vyjádřených určitým výrazem v obličeji.

Počátky analýzy emocionálních stavů podle výrazu lidského obličeje byly položeny Ch. Darwinem v roce 1872, který prokázal univerzálnost jednotlivých výrazů obličeje a jejich spojitost mezi lidmi a zvířaty [10]. Tyto základy byly přibližně o století později přepracovány a podepřeny o nové poznatky S. Tomkisem [61], který postuloval základy výzkumu obličejových výrazů jako pomocného prostředku při vyjadřování emocí. Podle Tomkinse jsou emoce hybnou silou našeho motivačního systému ve funkci zesílení odrazů okolních vlivů a náš obličej je nejdůležitějším prostředkem k jejich vyjádření. A to z toho důvodu, že lidské emoce jsou univerzálně sdíleny. Lidé různých kultur mohou snadno rozeznat veselý či smutný výraz v obličeji díky vrozenému emocionálnímu cítění pro vše, co je považováno za základní emoci, která působí na obličejové svalstvo tak, že vytváří dané výrazy v obličeji. V roce 1971 pak P. Ekmen a V. Friesen stanovili šest těchto základních emocionálních stavů, které mohou být vyjádřeny pomocí jedinečných výrazů v obličeji [14]. Těmito základními emocemi jsou: *radost, smutek, strach, odpor, překvapení a zlost*. Tyto emoce jsou však jen jedním z několika podnětů k vyjádření určitého výrazu v obličeji (viz obr. 1.1).



Obr. 1.1. Příčiny výrazů v obličeji [14].

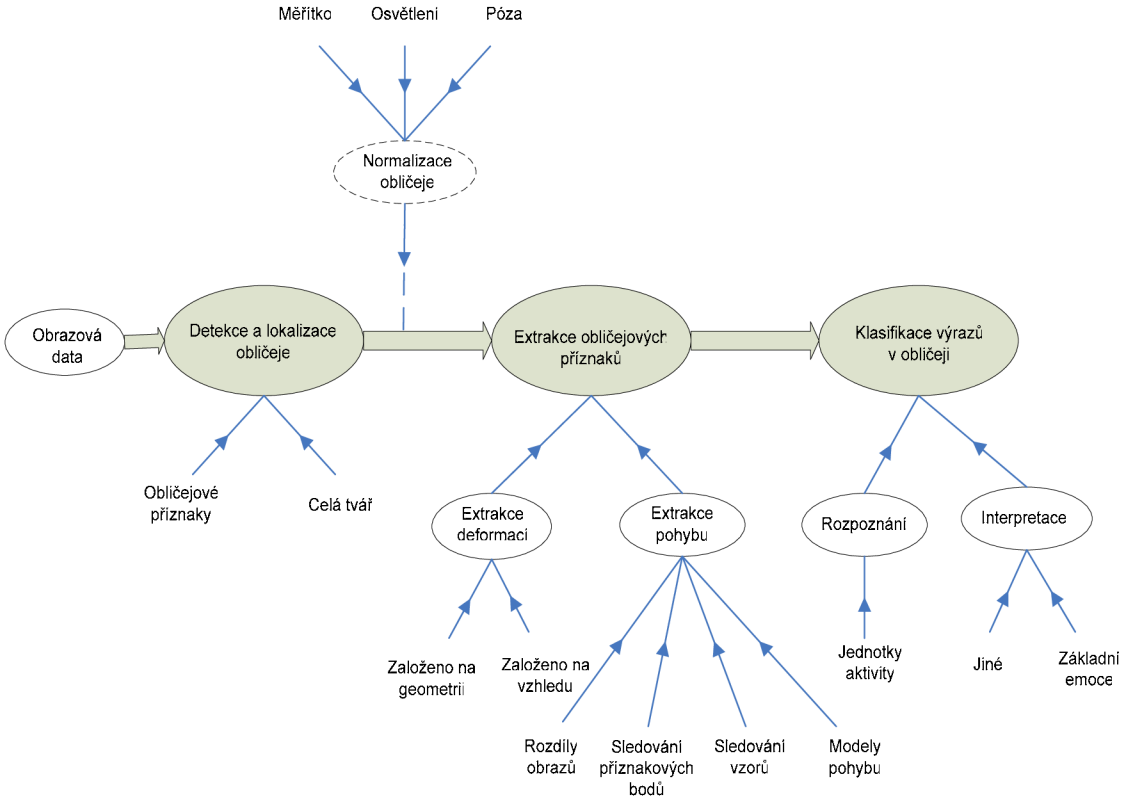
Analýza výrazů v obličeji a na jejich základě i analýza emocionálních stavů je důležitou součástí společenského života a mezilidských vztahů. Zároveň se však jedná o velmi obtížný proces, kdy pro správnou interpretaci může být potřeba školených specialistů (sociologové, psychologové apod.). Bylo by tedy vhodné využít již dříve zmíněných poznatků získaných při rozpoznávání tváří a pokusit se navrhnout automatický analyzátor emocionálních stavů na základě obrazové předlohy (buď ze statického obrazu nebo video-sekvence), který by tuto obtížnou úlohu zjednodušoval. Takový analyzátor by mohl najít uplatnění v různých oblastech:

- interaktivní rozhraní mezi člověkem a počítačem,
- neverbální komunikace,
- psychologie, psychiatrie,
- vyhodnocování bolesti,
- detektor lži,
- a jiné.

Součástí této práce je kapitola věnovaná přehledu současného stavu problematiky, který byl vypracován pomocí dostupné literatury, a na jehož základě byly stanoveny cíle disertační práce. V následujících kapitolách jsou poté popsány jednotlivé části systému pro automatické rozpoznávání emocionálních výrazů ve tváři, který se sestává z detekce obličeje, extrakce příznaků a zvoleného klasifikátoru. Výsledky práce následně jsou shrnuty v poslední kapitole.

2 Současný stav problematiky

Jak bylo uvedeno, principy analýzy emocionálních stavů vyjadřovaných prostřednictvím výrazů v obličeji mají společný základ s principy používanými při rozpoznávání obličejů. To znamená, že budou použity podobné metody, ale místo příznaků popisujících jedinečnosti tváře při různých výrazech se budeme orientovat na příznaky popisující jedinečnost výrazů v různých tvářích. Celý proces analýzy se dá rozdělit do tří po sobě jdoucích fází [18] viz obr. 2.1. Jedná se o detekci a lokalizaci obličeje v obraze s komplexním pozadím, včetně jeho normalizace, dále pak o extrakci vhodných příznaků popisujících daný výraz v obličeji a nakonec určení odpovídajícího výrazu na základě extrahovaných příznaků pomocí zvoleného klasifikátoru. Volba sledovaných příznaků, klasifikátoru apod. je vždy závislá na použitém přístupu a požadovaných výstupech. Z důvodu požadavku na minimální nutnou obsluhu takového analyzátoru je nezbytné, aby veškeré kroky a úpravy probíhaly zcela automaticky podle předem stanovených pravidel, vzorů či modelů.



Obr. 2.1. Celkový přehled automatického analyzátoru emocionálních stavů [18].

2.1 Metody detekce a lokalizace obličeje

Úkolem této oblasti vědy je nalezení přesné pozice obličeje/obličejů v obraze a obličej vhodným způsobem z celého obrazu extrahovat (např. pomocí tvaru pravidelného čtyřúhelníku či elipsy) pro další zpracování. Řešení této problematiky lze rozdělit na dvě základní skupiny: detekci obličejů ze statických obrazů a detekci obličejů z video-sekvencí. Obecně lze říci, že metody pro detekci obličejů ze statických obrazů lze použít i pro detekci tváří z video-sekvencí (samozřejmě s nutnou redukcí datového toku a snížení výpočetního výkonu, aby daná metoda mohla pracovat v reálné čase). Avšak opačný postup není možný, jelikož metody pro detekci obličeje ve video-sekvenci jsou založeny na informaci o pohybu a tato informace není ve statických obrazech obsažena. Z toho důvodu budou v celé další práci uvažovány pouze metody použitelné pro statické obrazy. Pro korektní práci obličejového detektoru je nezbytné, aby zvolená metoda byla imunní vůči následujícím vlivům:

- Vzájemná poloha obličeje a snímacího zařízení, což má za následek různé úhly natočení hlavy (frontální pohled, pohled z 45°, pohled z profilu, pohled shora či ze spodu) a také mohou být určité části obličeje mimo záběr (např. část úst či nosu, oko a podobně).
- Vzájemné vzdálenosti obličeje a snímacího zařízení a tím ovlivněná různé velikosti obličejů a jejich jednotlivých částí, což při velké vzdálenosti mimo jiné způsobuje ztrátu informací o detailech obličeje.
- Osvětlení snímané scény, kde záleží nejen na intenzitě jasu, která ovlivňuje všechny prvky dané scény stejně, ale především na jeho distribuci, což má za následek nestejnoměrné ovlivnění všech prvků snímané scény.
- Menší překážky, které dočasně zakrývají určitou část obličeje (např. ruka před ústy) a mohou tak způsobit ztrátu části informace, ale také působí jako rušivý prvek.
- Obličejové doplňky (brýle, vousy, apod.) mohou taktéž způsobit ztrátu informaci či působit jako rušivé prvky, ale na rozdíl od výše zmíněných menších překážek, jsou tyto doplňky standardním a často i trvalým prvkem, který daný obličej charakterizuje a je tedy nezbytné tyto prvky do případné detekce zahrnout.
- Národnost, věk, pohlaví, vzhled a jiné atributy, kterými se liší tváře jednotlivých lidí.
- Výrazy v obličeji, které jsou nestálé a mají vliv na celkový vzhled obličeje v daný okamžik.

Na základě přístupu lze rozdělit jednotlivé metody zabývající se detekcí a lokalizací obličeje do čtyř kategorií [65]:

Znalostní metody

Tyto metody detekce byly vyvinuty na základě vědeckých poznatků o lidské tváři. Popisují základní rysy tváře (ústa, oči, nos) a jejich vzájemné vztahy. Příkladem zde mohou být oči, které jsou ve tváři symetricky umístěny. Vztah mezi nimi pak může být definován určitými pravidly, vycházejícími z jejich umístění vzhledem k tváři nebo relativní vzdáleností vůči sobě [65]. Prvním krokem v tomto přístupu je extrakce rysů obličeje z obrazu, po které následuje vlastní detekce obličeje na základě znalosti vztahů mezi jednotlivými rysy. Problémem tohoto přístupu je ovšem jednak nutnost precizní lokalizace jednotlivých rysů a také složitý převod znalostí lidského obličeje do správně definovaných pravidel, kdy za předpokladu velmi přísných pravidel dochází k neúspěšné detekci řady tváří (tváře nejsou nalezeny) a za předpokladu volnějších pravidel umístění jednotlivých rysů dochází k falešné detekci objektů, které tvářemi nejsou. Dále pak je složité tyto metody rozšířit pro detekci tváří s různým úhlem natočení, neboť pro každý úhel natočení by musela existovat přesně definovaná pravidla. Tento přístup se využívá zejména k detekci tváří s frontálním pohledem v jednoduchých scénách. Příkladem takového přístupu může být publikovaná práce G. Yanga a T. Huanga [66], kteří použili více rozlišení jednoho obrázku, která získali podvzorkováním a průměrováním tohoto obrázku, a tříúrovňovou hierarchickou strukturu pravidel. Na nejnižší úrovni se v obraze s nejnižším rozlišením hledá tvář pomocí rozdílu intenzit jednotlivých pixelů. Postupem k vyšší úrovni se rozlišení zvyšuje tak, že při přestupu na vyšší úroveň jsou detekované oblasti nižší úrovně podrobeny hranové detekci. Na nejvyšší úrovni jsou pak zbylé segmenty ověřovány na přítomnost očí a úst. Díky použití hierarchické struktury došlo k omezení počtu nutných operací, avšak úspěšnost detekce je nízká.

Invariantní rysy

Na základě pozorování, že lidé dokáží detekovat obličej i jiné objekty v různých pózách a při různých světelných podmínkách, vzniká předpoklad, že musí existovat vlastnosti nebo příznaky, které jsou vůči okolním podmínkám invariantní [65]. Odtud jsou odvozeny metody založené na obecně platných rysech lidské tváře, které nepodléhají změnám osvětlení či natočení obličeje. Těmito rysy mohou být buď obličejové příznaky (oči, nos ústa, apod.), obličejová textura a nebo také barva obličeje. Obličejové příznaky jsou obvykle extrahovány za použití hranových detektorů a na jejich základě je vytvořen statistický model, který popisuje jejich vzájemný vztah a ověřuje existenci tváře. Tento postup může být ovšem ovlivněn velkou mírou šumu v obraze nebo nevhodným osvětlením, kdy stín způsobí množství ostrých hran v obraze a tím znehodnotí celou detekci. Příkladem takového přístupu může být morfologicky založený postup navržen C. Hanem [24], který vychází z tvrzení, že oči a obočí jsou nejvýznačnější rysy lidské tváře. Definují segmenty podobné očím jako hrany v kontuře očí. Nejprve je provedena řada morfologických operací pro získání pixelů, jejichž intenzita se významně mění, tyto pixely pak tvoří segment podobný očím. Pozice těchto segmentů pak slouží jako vodítko při hledání možných oblastí obsahující obličej, společně s geometrickou kombinací očí, nosu, obočí a rtů. Postup při detekci na základě textury

obličeje a barvy kůže bude uveden později. Výhodou těchto metod je snadná a rychlá implementace s kvalitními výsledky.

Srovnávání šablon

Tyto metody jsou založeny na korelaci obrazu se standardním vzorem tváře, šablonou, která je ručně vytvořena nebo parametrizována funkcí. Korelace s přednastavenými šablonami probíhá buď pro celý obličej nebo jen jeho jednotlivé části (např. oči, ústa, nos, apod.). Předpoklad existence tváře v daném obraze je pak založen na hodnotách korelační funkce, tedy do jaké míry odpovídá vstupní obraz standardní vzorové tváři, či jeho části [65]. Výhodou těchto metod je jejich snadná implementace, která je založena pouze na korelační analýze a jejím vyhodnocení. Přesto však tento přístup k detekci tváře není dostačující a je neefektivní, zejména z důvodu různých rozměrů tváře, orientaci či tvaru jednotlivých částí obličeje. V praxi se tedy používá více šablon tváře (různém rozměry, deformované tvary a jiné), avšak nevýhodou tohoto přístupu je nutnost vytvořit a mít uloženy v paměti jednotlivé šablony, což je velmi pracné a časově náročné. Příkladem tohoto přístupu je metoda popsaná A. Lanitisem a C. Taylorem [37], kteří předpokládají reprezentaci tváře pomocí informace o tvaru a intenzitě obrazu. Nejprve byly vytvořen soubor testovacích obrazů, kde byly manuálně označeny vzorové kontury (např. oblast očí, brada, nos) a vektor takto získaných bodů byl použit pro reprezentaci tvaru obličeje. Různorodost jednotlivých tvarů byla vyjádřena pomocí modelu rozdělení bodů a ten byl použit při detekci tváře pomocí aktivního modelu tvaru.

Metody založené na vzhledu

Tyto metody jsou odvozeny od metod využívajících srovnání s předdefinovanou šablonou. V tomto případě však šablony nejsou vytvářeny manuálně podle vlastností lidského obličeje, ale modely obličeje jsou získány strojovým učením z trénovací množiny, která obsahuje různé vzory tváří [65]. Detekce se opírá o technologii statistické analýzy a již zmíněného strojového učení, pomocí nichž hledá charakteristiky, které odlišují obrazy obsahující tváře a obrazy, ve kterých se tváře nevyskytují. Naučené charakteristiky mají tvar distribučních modelů nebo rozlišující funkce a jsou použity pro detekci obličeje. Metod tohoto typu existuje velké množství, v mnoha z nich je vzhledem k velkému množství dat prováděna redukce dimenzionality, čímž je dosažena lepší výpočetní efektivita a zároveň vyšší účinnost detekce. Příkladkem takového přístupu může být metoda použitá H. Rowleyem a S. Balujaou [53] zahrnující dvě základní komponenty. První komponenta je vícevrstevná síť sloužící k detekci vzoru tváří. Tato neuronová síť je trénována skupinou obrazů, které obsahují a neobsahují tváře,. Jejím vstupem je výřez obrazu o velikosti 20x20 pixelů a výstupem hodnota 1 a mínus1, kde hodnota 1 indikuje, že se ve výřezu nachází tvář a hodnota mínus 1 opak. Aby bylo možno detekovat tvář kdekoliv v obraze, vybírá se jako vstup neuronové sítě postupně každá oblast obrazu (pomocí tzv. okénkování). Aby bylo možno detekovat tváře větší než 20 x 20 pixelů, je obraz podvzorkován a předchozí postup se

opakuje. Druhou komponentou tohoto přístupu je modul rozhodování, který provádí konečné rozhodnutí z více detekcí. Tento výběr může být založen například na logických operacích.

Ne vždy lze říci, že určitá metoda zcela spadá do jedné z výše uvedených skupin, většina současně používaných metod přebírá některé poznatky z různých skupin, které následně kombinuje tak, aby dosáhla co nejlepších výsledků.

Pro lokalizaci jednotlivých částí obličeje jako nos, oči a podobně se používají stejné metody jako pro detekci celého obličeje, pouze jsou použity odlišné příznaky popř. vzory. Ovšem pouze v tom případě, pokud přesnou lokalizaci těchto obličejových částí nevyžaduje již samotná detekce obličeje. Výhodou při lokalizaci jednotlivých částí obličeje je fakt, že tyto části nejsou vyhledávány v celém vstupním obraze, ale pouze v segmentech, které odpovídají jednotlivým tvářím. Dále pak zde lze využít již zmíněných pravidel o znalosti lidské tváře, a tedy i přibližné pozici a orientaci jednotlivých částí obličeje. Na základě znalosti přibližné pozice a orientace lze potom prohledávaný prostor daného obličejového segmentu lépe specifikovat a snížit tak výpočetní náročnost procesu lokalizace a zároveň i zvýšit jeho přesnost.

Výstupem části detekce obličeje jsou segmenty detekovaných obličejů o různé velikosti, s různou orientací a podobně, avšak vstupem další části analýzy emocionálních stavů dle obr. 2.1, by měl být detekovaný obličeje s jeho přesně lokalizovanými částmi v jednotném předem definovaném formátu. Z tohoto důvodu je tedy nezbytné provést normalizaci detekovaného obličeje z následujících hledisek:

- Velikost obličeje lze normalizovat pomocí změny měřítka daného segmentu tj. nadvzorkováním či podvzorkováním.
- Úhel natočení obličeje (póza) lze rozdělit do dvou skupin. Při natočení obličeje kolem osy z v třírozměrném prostoru (x,y,z), kdy nedochází k zakrytí žádné části obličeje, lze natočení jednoduše kompenzovat změnou úhlu natočení. Pokud však došlo k natočení obličeje kolem os x nebo y (tj. kolem horizontální nebo vertikální osy), je tato kompenzace znesnadněna překrytím části obličeje sebou samým či případně prostorovým zkreslením. Tato problematika bývá většinou řešena pomocí obličejových modelů [16], kdy je chybějící část obličeje doplněna dle daného modelu a prostorové zkreslení je na základě proporcí modelů redukováno. Výsledky těchto metod jsou vždy závislé na velikosti úhlu otočení.
- Osvětlení lze normalizovat na základě jeho intenzity nebo jeho distribuce. Normalizace intenzity osvětlení se řeší pomocí úpravy histogramu jasové složky obrazu tak, aby rozložením odpovídala předdefinovanému vzoru. Řešení problematiky normalizace distribuce osvětlení bývá prováděno pomocí tzv. jasových modelů [5], kterým je vstupní obraz přizpůsobován.

2.2 Metody extrakce obličejových příznaků

Cílem této části je zvolit takovou sadu příznaků, které jsou jedinečné pro každý specifický výraz v obličeji. Výběr vhodných příznaků patří v celé analýze k nejdůležitějším, protože pouze ze správně zvolených příznaků lze analyzovat odpovídající emocionální stav. Zvolené příznaky musí být pro daný výraz zcela invariantní vůči pohlaví, věku, rasy a vzhledu analyzovaného subjektu. Podle [47] můžeme rozlišovat dva základní typy obličejových příznaků:

a) Trvalé obličejové příznaky – jsou vždy na tváři přítomny, ale mohou být deformovány na základě výskytu určitého výrazu v obličeji. Jsou to zejména oči, rty, obočí, ale také permanentí vrásky či celková textura obličeje.

b) Přechodné obličejové příznaky – vrásky a vybouleniny objevující se pouze v případě určitého výrazu v obličeji. Nejčastěji na čele, v oblastech kolem očí, koutcích úst a podobně.

Podle formátu vstupních obrazových dat (statické obrazy nebo video-sekvence) rozlišujeme extrakci příznaků na:

- Extrakci příznaků deformací – používá se u statických obrazů i u video-sekvencí, příznaky se extrahují nezávisle z každého obrazu nebo snímku. Jedná se o příznaky typické pro vyjádření různých deformací jednotlivých částí lidského obličeje (přivřené oko, otevřená ústa a podobně), které se podílí na daném celkovém emocionálním výrazu.

- Extrakci příznaků pohybu – používá se pouze u video-sekvencí z důvodu požadavku na časovou variabilitu, příznaky jsou extrahovány na základě rozdílů pozic (pohybu) sledovaných objektů v jednotlivých snímcích. Tyto příznaky slouží k popisu dynamiky jednotlivých částí lidského obličeje (mrkání, pohyb úst a podobně), které se podílí na daném celkovém emocionálním výrazu. Jelikož se tyto příznaky mění s časem, musí být tomuto přístupu přizpůsoben i výsledný klasifikátor.

Při použití video-sekvencí je sice dosahováno lepších výsledků, avšak za cenu návrhu a použití velmi komplikovaných klasifikátorů a velkých výpočetních nároků. V této práci budou dále uvažovány pouze přístupy zabývající se zpracováním statických obrazů. Základní metody extrakce obličejových příznaků ze statických obrazů lze rozdělit podle jednotlivých přístupů na [47]:

- Holistické metody – zpracovává se celá tvář najednou, tyto metody jsou vhodné zejména při rozpoznávání nejběžnějších výrazů.

- Lokální metody – zpracovávají se pouze jednotlivé předem zvolené oblasti obličeje, vhodné pro rozpoznávání jemných změn v malých oblastech.

- Metody založené na obrazu – není zde předpokládána žádná předchozí znalost o zpracovávaném objektu, jsou rychlé a jednoduché.
- Metody založené na modelech – 2D nebo 3D modely průměrných lidských tváří, poměrně přesné, avšak pracné na vytvoření a výpočetně náročnější.

Přehled jednotlivých metod a jim odpovídajících representativních prací je uveden v tab. 2.1.

Tab. 2.1. Rozdělení reprezentativních prací dle použitých metod.

	Holistické metody	**Lokální metody**
Metody založené na obrazu	Metoda založená na pixelech	Profily intenzity jasu
	Gaborovy wavelety	Analýza hlavních komponent
Metody založené na modelech	Model aktivního vzhledu	Geometrické modely obličeje
	Značené grafy	Modely založené na dvou různých pohledech

Metoda založená na pixelech

Tato jednoduchá metoda v podstatě ani do kategorie extrakce obličejových příznaků nespadá, jelikož jako jednotlivé příznaky jsou brány přímo hodnoty jednotlivých pixelů popisující originální tvář [38]. Hlavní úlohu zde pak zastává zvolený klasifikátor. Výhodou této metody je její jednoduchá implementace, nevýhodou pak velká nadbytečnost vstupních dat a s tím spojené velké výpočetní nároky při procesu klasifikace, dále pak poměrně nízká přesnost.

Gaborovy wavelety

Podobně jako v předchozí metodě odpovídá počet extrahovaných příznaků počtu pixelů odpovídajících obrazu tváře, avšak jednotlivé příznaky neudávají přímo hodnotu daných pixelů, ale hodnotu odezvy těchto obrazových bodů na filtraci Gaborovým waveletem [21]. Hlavní úlohu zde pak opět zastává vhodně zvolený klasifikátor. Výhodou této metody oproti metodě předchozí je potlačení vlivu osvětlení a zvýšení celkové přesnosti, avšak zároveň dochází ke zvýšení výpočetních nároků při samotném procesu extrakce.

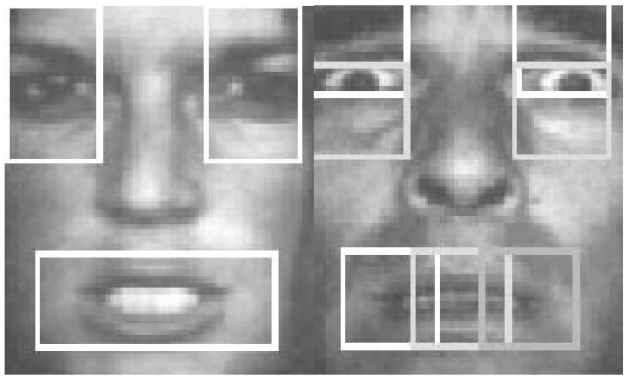
Profily intenzity jasu

Princip této metody spočívá v měření profilů intenzit jasu *(Intensity Profiles)* podél zvolených segmentů [3]. Jednotlivé segmenty pak obsahují ty části obličeje, které se na utváření příslušného emocionálního výrazu ve tváři podílí největší měrou (zejména oči a rty).

Tato metoda se zejména používá pro popis přechodových obličejových příznaků, je ale velmi náchylná na změnu světelných podmínek.

Analýza hlavních komponent

Analýza hlavních komponent *PCA (Principal Component Analysis)* je statistická metoda sloužící k redukci vstupních dat při zachování maximální možné míry variability a byla použita C. Padggetem a W Cottrelem pro lokální extrakci obličejových příznaků z významných segmentů obrazu lidského obličeje [46]. Z trénovaní množiny obrazových dat byly vybrány bloky pixelů kolem obou očí a úst obr. 2.2 a), které byly pomocí analýzy hlavních komponent promítnuty na 15 vlastních vektorů s největší mírou variability pro jednotlivé vstupní obrazy. Výhodou této metody jsou nízké výpočetní nároky a dobré výsledky pro nejčastěji používané výrazy ve tváři, které se navzájem výrazně liší, nevýhodou pak závislost na světelných podmínkách. Pro zvýšení přesnosti byla výše zmíněnými autory navržena modifikace spočívající v rozdělení jednotlivých bloků do tří a čtyř navzájem se překrývajících pod-bloků (tři pod-bloky pro oblasti kolem očí a čtyři pod-bloky pro oblast kolem úst), které jsou zpracovány samostatně viz. obr. 2.2 b).



a) b)

Obr. 2.2. Příklad výběrů oblastí zájmů [46].

Aktivní model vzhledu

Aktivní modely vzhledu *AAM (Active Appearance Model)* vychází z aktivních modelů tvaru *ASM (Active Shape Model)*, jež byly navrženy T. Cootsem a C. Taylorem [9]. Modely tvaru jsou vytvářeny pomocí manuálně umístěných bodů, které se nacházejí na významných pozicích lidského obličeje, zejména v místech ostrých přechodů intenzity jasu (oblasti okolo očí, rtů, nosních dírek a celého obličeje na kontrastním pozadí). Z jednotlivých modelů tvaru různých obličejů je poté vytvořen normalizovaný model tvaru obličeje. Při extrakci příznaků se pak snažíme rozmístit jednotlivé body normalizovaného modelu tvaru obličeje tak, aby co nejlépe odpovídaly skutečným pozicím těchto bodů v analyzovaném obličeji. Vlastní rozmísťování jednotlivých bodů se skládá ze dvou fází. V první fázi jsou nalezeny nejpravděpodobnější pozice těchto bodů v jejich blízkém okolí a vytvořen model tvaru

daného obličeje, ve druhé fázi je pak tento model statisticky porovnán s normalizovaným model tvaru obličeje (z hlediska posunutí, zvětšení a rotace) a pozice jednotlivý bodů modelu tvaru daného obličeje aktualizovány. Tyto dvě fáze se opakují do té doby dokud nedojde ke konvergenci k požadovanému výsledku. Aktivní model vzhledu pak tento přístup rozšiřuje o statistický model textury v okolí bodů modelu tvaru a jeho porovnání s normalizovaným modelem textury [13]. Tato metoda se tedy používá k vytváření realisticky vypadajících modelů tváří při různých výrazech v obličeji a jejich porovnávání s testovanou tváří na základě tvaru a textury. Výhodou jsou velmi dobré výsledky pro nejčastěji používané výrazy v obličeji a možné použití i při lokalizaci jednotlivých částí lidského obličeje (oči, nos, ústa), avšak manuální vytváření jednotlivých modelů je velmi pracné a časově náročné, dále pak úspěšnost použití této metody velmi záleží na co nejpřesnějším umístění počátečního modelu tvaru daného obličej. Při nedostatečně přesném umístění tato metoda selhává.

Značené grafy

Značené grafy využívají filtraci vstupního obrazu sadou Gaborových waveletů (různé velikosti měřítka a různé prostorové orientace). Vychází se zde z předpokladu, že odezvy na jednotlivé filtry jsou v sousedních bodech vzájemně silně korelované, a je tedy možné uvažovat pouze několik klíčových bodů [40]. Z tohoto důvodu byla navržena pevně stanovená 50-ti bodová obličejová maska (*Labeled Graph*) viz obr. 2.3, kde každý bod této masky obsahuje uzel tzv. *jet* skládající se z pole komponent (každá komponenta je odezva vstupního obrazu na daný Gaborův wavelet v příslušném bodě obrazu). Jednotlivé uzly jsou váhovány dle své významnosti při vytváření příslušných emocionálních výrazů ve tváři (tj. váhy nízké hodnoty v okolí vlasů a podobně, vyšší hodnoty v okolí rtů, očí) a porovnávány s uzly modelových obrazů, odvozených z trénovaní množiny obličejů. Výhodami této metody jsou velká odolnost proti změně světelných podmínek a při správném nastavení obličejové masky také do jisté míry odolnost proti úhlu natočení hlavy. Nevýhodou jsou ovšem vyšší výpočetní nároky kladené na proces vlastní extrakce příznaků a také požadavek na přesné přiřazení jednotlivých bodů obličejové masky analyzovanému obličeji.

Obr. 2.3. 50-bodová maska obličeje [40].

Geometrické modely obličeje

Tato metoda je založena na myšlence manuálního vytvoření geometrického modelu jednotlivých oblastí obličeje, normalizace příslušných oblastí analyzovaného obličeje vzhledem k těmto geometrickým modelům a srovnávání těchto oblastí na základě distribuce osvětlení [34]. Vytváření takovýchto modelů je pracné, časově náročné a vyžaduje poměrně přesnou lokalizaci jednotlivých oblastí obličeje, porovnání jednotlivých oblastí je náchylné na změnu světelných podmínek.

Modely založené na dvou různých pohledech

Modely založené na dvou různých pohledech *(2 View Point-Based Models)* navrženy M. Panticovou a L. Rothkrantzem na principu vytváření modelů jednotlivých částí obličeje na základě jejich kontury z dvou různých pohledů (frontální a z profilu) [48]. Požadavek dvou fixně postavených snímacích zařízeních pracujících současně a větší výpočetní náročnost patří mezi hlavní nevýhody tohoto přístupu.

2.3 Metody klasifikace výrazů v obličeji

Poslední částí automatické analýzy emocionálních stavů z obličejové předlohy je klasifikace jednotlivých výrazů na základě zvolených příznaků. Jedná se v podstatě o přiřazení analyzovaného emocionálního výrazu v obličeji do některé z požadovaných kategorií na základě hodnot extrahovaných příznaků. V této části lze zvolit řadu různých přístupů a nelze všeobecně říci, který z nich je nejvhodnější. Záleží vždy na typu zvolených příznaků, požadavcích na výstup a vhodném nastavení. Obecně lze rozdělit tyto metody do tří základních skupin.

Klasifikátory založené na pravidlech

Tento typ klasifikátorů patří mezi nejjednodušší, jedná se v podstatě o soubor pravidel nastavených uživatelem. Emocionální výraz v obličeji spadá do určité kategorie, pokud jeho příznaky splňují pravidla pro tuto kategorii. Klasifikátory založené na pravidlech je vhodné použít pro malý počet příznaků s přesně definovatelnými hranicemi jejich hodnot (např. pro rozhodnutí zda je oko zavřené či otevřené na základě vzájemné vzdálenosti spodní části oka a víčka). Tento typ klasifikátorů lze také rozšířit o pravděpodobnostní rozdělení, tj. přiřazení do příslušné kategorie na základě četnosti příznaků či skupin příznaků splňujících pravidla dané kategorie.

Klasifikátory založené na porovnávání s modely

Tento typ klasifikátorů lze použít v případě, kdy máme k dispozici vzorový model nebo šablonu pro každou požadovanou kategorii. Využívá se zde porovnání extrahovaných příznaků modelu analyzovaného výrazu ve tváři s příznaky vzorových modelů. Toto porovnání se provádí pomocí vzdáleností tzv. *metrik*. Mezi nejznámější metriky patří Euklidova, Čebyševova, Mahalanobisova a další. Jednotlivé příznaky mohou být uživatelem při výpočtu zvolené metriky vhodně váhovány dle své významnosti na určení příslušné kategorie.

Klasifikátory využívající strojového učení

Tento typ klasifikátorů je založen na myšlence, že algoritmy a systémy mohou zdokonalovat svůj výkon v závislosti na čase, tj. mohou se učit. Učením se zde rozumí definice pravidel pro jednotlivé kategorie na základě matematických funkcí. Na základě přístupu k procesu učení lze algoritmy strojového učení rozdělit do dvou základních skupin [17]:

- *Učení s učitelem* (*Supervised learning*) - na základě trénovacích dat, což je množina vstupních příznaků a jim odpovídající požadované výstupní hodnoty klasifikátoru (tj. přiřazení odpovídající kategorie), vytváří klasifikátor

přenosovou funkci, která s největší možnou měrou mapuje vstupní příznaky do příslušné výstupní kategorie.

- U*čení bez učitele* (*Unsupervised learning*) – v tomto případě chybí trénovací množina dat a algoritmus tak nezná požadované výstupní hodnoty. Vstupní data jsou shromažďována a považována za množinu náhodných proměnných . K této množině pak existují dva přístupy. Jednak je možné vytvořit statistický model (např. tzv. *Bayesova síť*) pomocí odhadu hustoty pravděpodobnosti nebo pomocí techniky extrakce charakteristických vlastností získat ze vstupu statistické zákonitosti [11].

Pro klasifikaci emocionálních výrazů ve tváři bývá většinou použit první zmíněný typ. Výhodou klasifikátorů využívajících strojové učení je poměrně vysoká klasifikační schopnost bez nutnosti manuálního zásahu do procesu učení, což je výhodné zejména pro větší počet vstupních příznaků. Nevýhodou pak jsou vyšší výpočetní nároky kladené na proces učení a nutnost vytvoření trénování množiny dat.

Algoritmů strojového učení bývá v praxi používána velká řada s různým počtem jejich modifikací. Níže jsou uvedeny tři nejčastěji používané základní algoritmy:

Algoritmus *k*-nejbližších sousedů *(k-Nearest Neighbors)*

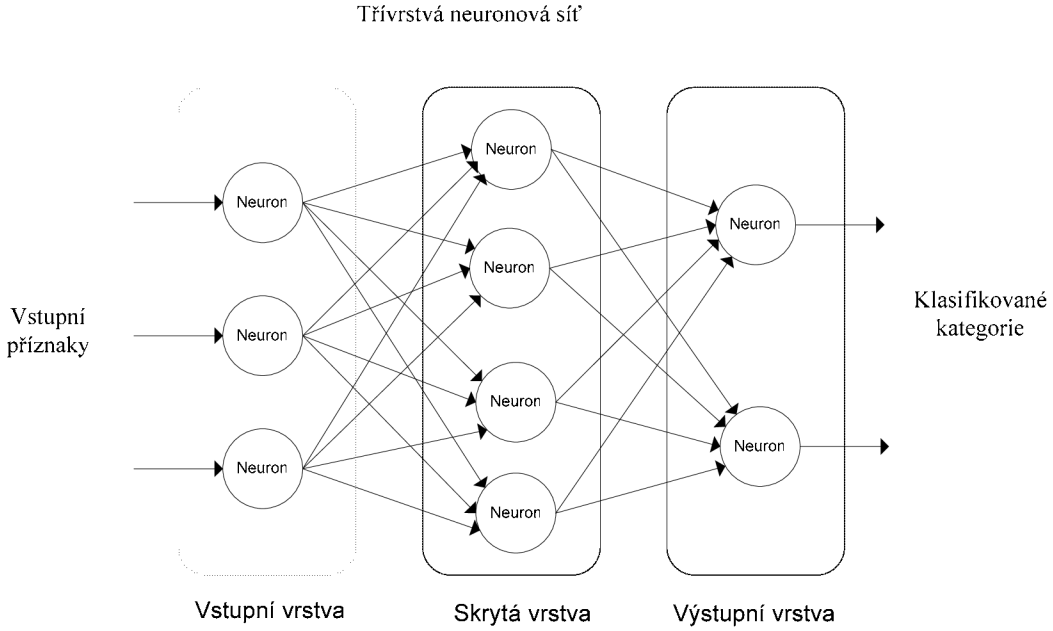
Metoda *k*-nejbližších sousedů *k-NN* patří k nejjednodušším klasifikačním algoritmům založených na strojovém učení. Vychází z metody nejbližších sousedů (*Nearest Neighbors*), kdy je sledovaný objekt klasifikován do některé z kategorií na základě nevětší míry podobnosti extrahovaných příznaků s příznaky trénovací množiny dat dané kategorie. Míra podobnosti je obvykle vyhodnocována pomocí metrik (např. Euklidovy vzdálenosti). Na rozdíl od metody nejbližších sousedů, kdy je sledovaný objekt přiřazen přímo do kategorie do níž spadá příslušný vektor příznaků z trénovací množiny dat s nejvyšší mírou podobnosti k příznakům sledovaného objektu, je v metodě *k*-nejbližších sousedů vybráno *k* vektorů příznaků z trénovací množiny dat s nejvyšší mírou podobnosti k příznakům sledovaného objektu. Objekt je následně zařazen do té kategorie, do které spadá většina těchto vektorů [4]. Výběr parametru *k* závisí na charakteru vstupních dat. Větší hodnota *k* zmenšuje vliv šumu na klasifikaci, ale více rozostřuje hranici mezi jednotlivými kategoriemi.

Neuronové sítě (*Neural Networks*)

Jedná se o matematický model založený na pozorování biologické neuronové sítě. Ta se skládá z neuronů, což jsou elektricky excitované buňky nervového systému, které dokáží zpracovávat a přenášet informaci. Podobně je pak umělá neuronová síť tvořena z mezi sebou vzájemně spojených skupin umělých neuronů a zpracovává informace pomocí takového spojení. První matematický model neuronu byl navržen W. McCullochem a W. Pittsem a dodnes se běžně používá [41]. Matematický model neuronu je tvořen nastavitelnými váhami jednotlivých vstupů, prahovou hodnotou neuronu a nelineární *aktivační* (přenosovou) funkcí, která se může u jednotlivých neuronů lišit v závislosti na konkrétní typu řešené úlohy.

Hodnoty jednotlivých vah pro každý neuron jsou nastavovány v procesu učení na základě daného algoritmu učení.

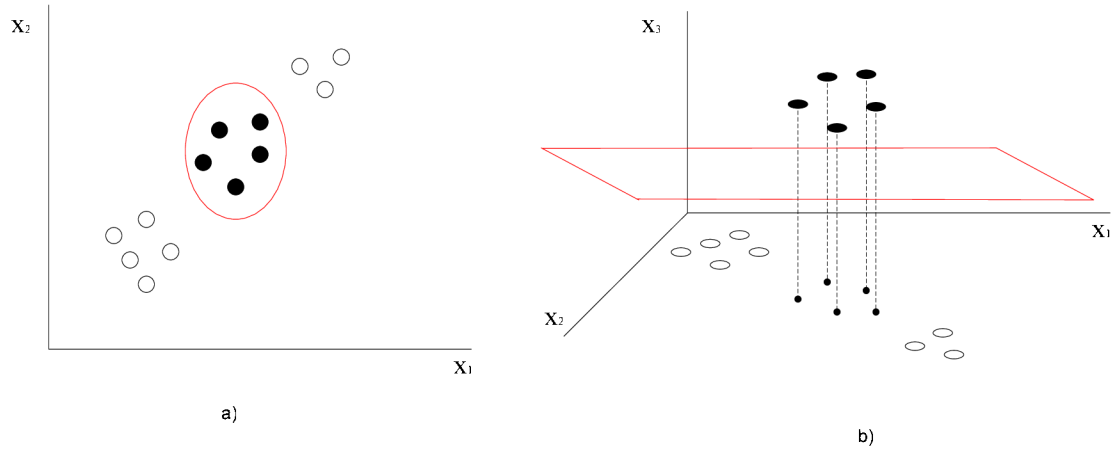
Neuronová síť se dělí na vrstvy, kde každá vrstva obsahuje určitý počet umělých neuronů, které paralelně zpracovávají vstupní informace dané vrstvy na informace výstupní. Každá neuronová síť obsahuje jednu vstupní, kde je počet neuronů dán počtem vstupních příznaků, a jednu výstupní vrstvu, kde je počet neuronů dán počtem klasifikovaných kategorií, a dále může obsahovat několik vrstev skrytých s libovolným počtem neuronů. Podle struktury zapojení lze rozdělit neuronové sítě do dvou hlavních skupin: na sítě s dopředným šířením a na sítě se zpětnou vazbou. U sítě s dopředným šířením jsou výstupy všech neuronů jedné vrstvy přivedeny na vstup všech neuronů vrstvy následující (v případě vstupní vrstvy se jedná o vstupní příznaky) viz obr. 2.4. Tento typ sítě bývá používán nejčastěji. Sítě se zpětnou vazbou se liší od předchozích v tom, že výstupy jedné vrstvy jsou zároveň vedeny zpět na vstup dané vrstvy.



Obr. 2.4. Příklad třívrstvé neuronové sítě s dopředným šířením.

Algoritmy podpůrných vektorů (*Support Vector Machine*)

Jedná se o lineární klasifikační algoritmus, jehož geometrickou interpretaci lze chápat jako hledání ideální nadroviny, která od sebe odděluje příznaky různých kategorií. Tato klasifikační metoda je tedy z principu binární, tj. lze pomocí ní rozlišit pouze dvě různé kategorie. Popsaným způsobem ovšem můžeme klasifikovat pouze lineárně separovatelné vstupní příznaky. V případě, že tyto příznaky nejsou lineárně separovatelné, lze zde využít transformace jádra, která převádí příznaky do vícedimensionálního prostoru, kde již separovatelné jsou [7] viz obr. 2.5. Tímto způsobem lze tedy representovat i vysoce složité nelineární funkce.



Obr. 2.5. Klasifikace příznaků dvou kategorií a) nelineární funkcí a b) lineární nadrovinou pomocí převodu do vyšší dimenze.

3 Cíle disertace

Jak bylo uvedeno v předchozí kapitole, většina současných systémů zabývajících se automatickou analýzou emocionálních stavů z výrazu v obličeji věnuje svoji pozornost pouze samotné analýze ze zvolených příznaků, avšak nikoliv již vlastní detekci, lokalizaci a normalizaci tváře, a přesné lokalizaci jeho jednotlivých částí. Z tohoto důvodu dochází ke zkreslení konečných výsledků, jelikož je třeba si uvědomit, že do celkového výsledku musí být také zahrnuta chyba nalezení správné pozice obličeje, jeho jednotlivých částí a další nezbytné výše zmíněné postupy. Navíc lze využít poznatků získaných v jednotlivých částech analýzy i pro části ostatní a zvýšit tak spolehlivost celého systému.

Hlavním cílem disertační práce bude návrh a realizace kompletního automatického systému určeného k rozpoznávání emocionálních stavů na základě výrazů v obličeji a vyhodnocení jeho celkové úspěšnosti na vhodné databázi obličejových výrazů.

Nejprve je tedy nezbytné navrhnout efektivní metodu detekce a lokalizace obličeje v obraze nezávisle na snímaném pozadí a množství obličejů ve snímané scéně. Dále pak metodu vhodnou lokalizaci jednotlivých obličejových částí, jež budou použity k extrakci zvolených příznaků.

V další fázi bude třeba optimalizovat dosud publikované metody extrakce příznaků, a to vhodnou kombinací statistických metod a metod založené na dvourozměrné filtraci obrazů. Optimalizace bude probíhat nejen z hlediska úspěšnosti, ale i z hlediska robustnosti, výpočetní náročnosti a efektivnosti celého procesu.

V poslední fázi práce bude kladen důraz na zvolení vhodného typu klasifikátoru a jeho nejlepšího možného nastavení tak, aby bylo dosahováno co nejspolehlivějších výsledků.

4 Detekce obličeje a lokalizace obličejových částí

Tato část práce se věnuje návrhu a realizaci efektivní metody detekce a lokalizace obličeje a jednotlivých obličejových rysů. Z důvodů uvedených v kapitole 2 byly pro vlastní detekci obličeje vybrány metody založené na invariantních rysech. Tento typ metod lze dále rozdělit podle sledovaných rysů na:

- *Obličejové rysy* − tyto techniky se zaměřují přímo na hledání obličejových rysů jako jsou oči, ústa, nos apod. Využívají se různé matematické transformace k zvýraznění těchto rysů vůči zbytku obličeje či celému obrazu. Mezi tyto techniky patří např. detekce hran a další filtrační metody [68]. Pokročilejší algoritmy využívají znalosti, že jedna část obličeje má určitou vzdálenost od jiné popř. ostatních částí, dále se využívá pravděpodobnosti, kde se daná část má nacházet, ale také pravděpodobnost, kde daná část být nemůže. Příkladem může být prostý fakt, že klasifikátor by neměl vyhodnotit testovaný objekt, kde rty se nacházejí mezi očima a nosem, jako obličej. Tato metoda selhává zejména pokud se testovaný obličej nachází v obraze s komplexním prostředím (velká míra hran a jiných strukturálních elementů).
- *Obličejové textury* − lidské obličeje mají zřetelnou strukturu, která se liší od ostatních objektů v obraze. Obvykle se oddělují od sebe kůže, vlasy a ostatní objekty v obraze. Využívá se barevné informace obrazu pro hledání obličejové textury [1]. Jedná se o složitější a výpočtově náročnou metodu.
- *Barva kůže* − podstatou této metody je hledání oblastí barevně odpovídajících barvě kůže [55]. Výhodou použití této metody je především rychlost, velmi malá citlivost na změnu světelných podmínek, necitlivost na velikost a úhel natočení tváře, věk, či pohlaví dané osoby. Dále pak ve velké míře umožňuje detekovat obličej s různými strukturálními komponenty obličeje (jako např. vousy, brýle, vlasy padající do čela apod.) a to i na velmi komplexním pozadí. Její největší slabinou ovšem je, pokud se hledaný obličej nachází v obraze, jehož pozadí je barevně shodné s předpokládanou barvou kůže. Proto je vhodné tuto metodu doplnit o rozšíření tak, aby tato nežádoucí vlastnost byla potlačena.

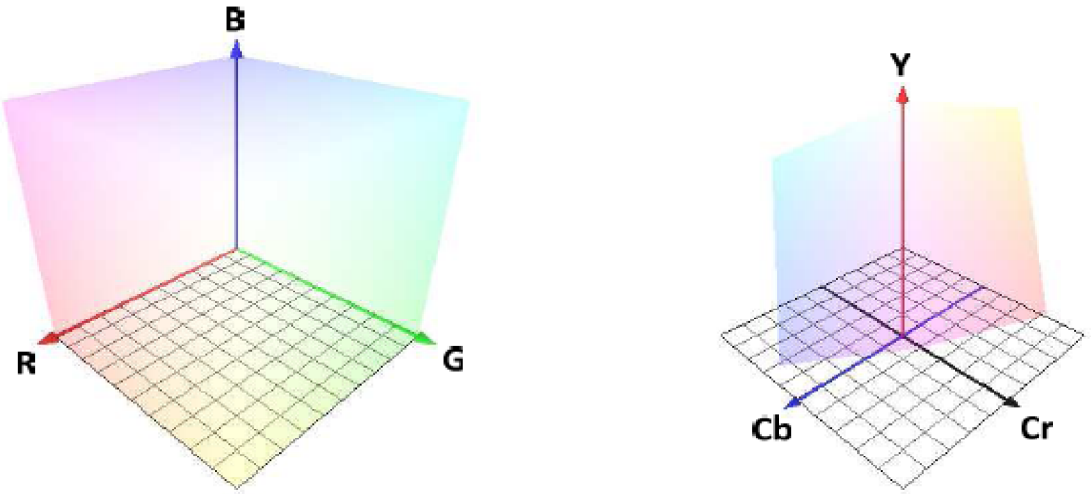
Při návrhu algoritmu automatického systému detekce obličeje jsme vycházeli z předpokladu, že se v analyzovaném obraze může vyskytovat neomezené množství tváří v různém úhlu natočení a v různé vzdálenosti od snímací kamery. Avšak z důvodu požadavku na další zpracování detekovaného obličeje budou pro lokalizaci jednotlivých částí obličeje uvažovány pouze ty obličeje, u kterých budou viditelné obě oči i ústa a které dosahují alespoň minimální velikosti 20×20 pixelů.

4.1 Detekce obličeje na základě barvy kůže

Pro realizaci automatického systému detekce a lokalizace obličeje [79] byla z důvodu požadavků na minimální výpočetní náročnost a maximální přesnost nejprve vybrána metoda pracující na základě barvy kůže. Základní princip vychází z pozorování, že barva lidské kůže může být v daném barevném prostoru vyjádřena pomocí určitého spojitého podprostoru. Ohraničením tohoto podprostoru lze modelovat barvu lidské kůže a následně pak klasifikovat barvu jednotlivých pixelů obrazu. Výzkum v oblasti kolorimetrie, zobrazování a zpracování obrazů přinesl mnoho barevných prostorů s rozdílnými vlastnostmi. Tvar a kompaktnost podprostoru barev kůže se v jednotlivých barevných prostorech liší [33]. Výběr vhodného barevného modelu je tedy zásadní pro mnoho metod klasifikace.

Barevné prostory a podprostor barvy kůže

Za základní barevný prostor používaný ve zpracování obrazu lze označit barevný prostor *RGB,* kde je barva jednotlivých pixelů definována aditivním mícháním tří základních barevných složek (červené, modré a zelené) o různé intenzitě. Vzhledem k tomu, že v tomto barevném prostoru není explicitně oddělena jasová a barevná složka, není při zpracování obrazů příliš využívaný. Místo toho se častěji využívá jiných barevných prostorů jako např. *HSV, YCbCr* a podobně. Televizní standard *PAL* i obrazový kompresní algoritmus *JPEG* používají barevný prostor *YCbCr*, kde *Y* reprezentuje jasovou složku, *Cb* a *Cr* barevné chrominanční složky. Výhodou použití tohoto barevného prostou je důsledná separace jasové složky a jednotlivých chrominačních složek a také jednoduchý transformační vztah pro převod ze základního barevného prostoru *RGB* (4.1) [65]. Na obr. 4.1 je ukázán grafický převod z barevného prostoru RGB do barevného prostoru *YCbCr*.

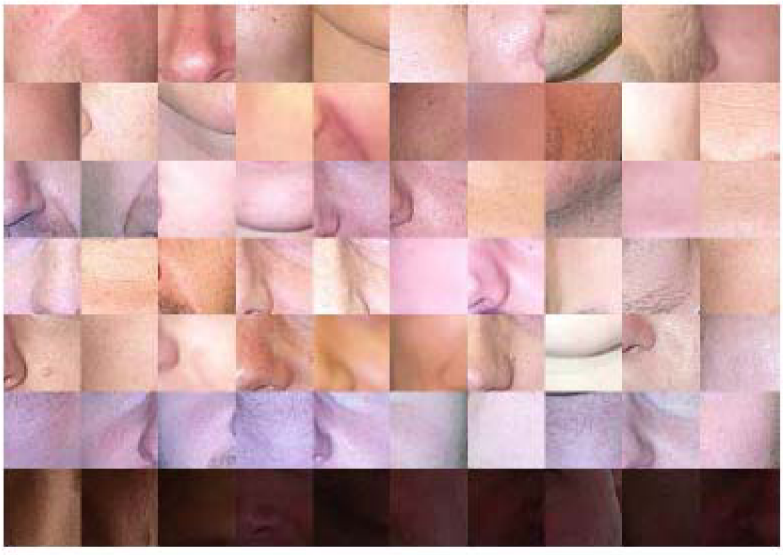


Obr. 4.1. Transformace barevného prostoru *RGB* do barevného prostoru *YCbCr* (vytvořeno v programu *ColorSpace* [8]).

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65,481 & 128,553 & 24,966 \\ -37,797 & -74,203 & 112,000 \\ 112,000 & -93,786 & -18,214 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4.1}$$

Podstatou barevné segmentace při detekci obličeje ve scéně je rozdělení jednotlivých pixelů podle jejich barevné hodnoty na pixely barevně spadající do oblasti kůže a pixely, jejichž barevná složka barvě kůže neodpovídá [33]. Nejprve je ovšem nezbytné tuto barevnou oblast kůže definovat. Zde můžeme využít dva různé způsoby definice barevného prostoru odpovídajícího barvě kůže:

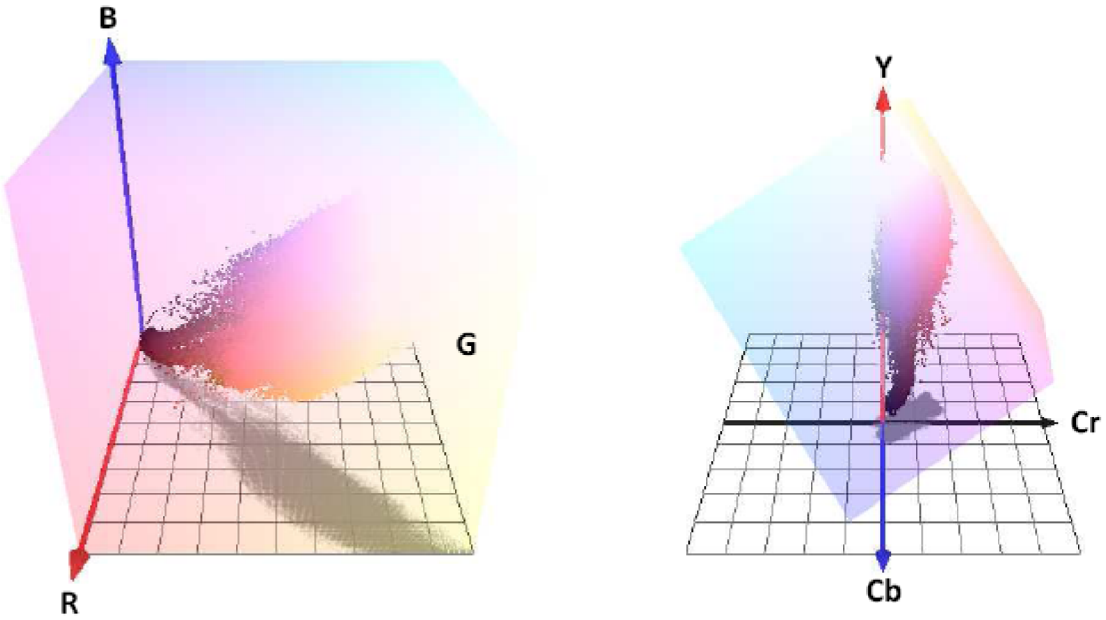
- Neparametrické modely rozložení barvy kůže - klíčová myšlenka tohoto přístupu je zjistit rozložení barvy kůže z trénovací množiny bez explicitního vyjádření barevného podprostoru pomocí pravidel. Namísto toho je většinou konstruována takzvaná pravděpodobnostní mapa, tj. jednotlivým bodům diskretizovaného barevného prostoru je přiřazena pravděpodobnost, s jakou má daný bod barvu kůže.
- Parametrické modely - parametrické modely se snaží popsat tvar rozložení barvy kůže v rámci barevného modelu vhodným matematickým modelem. Parametry tohoto modelu jsou pak získány z trénovací množiny. Podprostor barvy kůže má ve většině používaných modelů eliptický tvar.



Obr. 4.2. Příklad vzorů lidské obličejové kůže.

Nejprve byl tedy v rámci práce shromážděn dostatečný počet různých obrazových vzorků kůže (celkově přes 400 vzorků), které byly pořízeny třemi odlišnými snímacími zařízeními z různých obličejových částí přibližného počtu 50 lidí (viz. obr. 4.2). Odtud byla vytvořena množina vzorů jednotlivých barev, která je reprezentována třírozměrnou maticí o velikost 256×256×256, kde index každého prvku této matice zastupuje odpovídající barevnou hodnotu z barevného prostoru *YCbCr* a je nastaven buď na hodnotu 1 (daná barevná

hodnota spadá do barevné oblasti kůže) nebo na hodnotu 0 (daná barevná hodnota do barevného prostoru kůže nespadá). Rozdělení prvků spadajících do barevného podprostoru odpovídacího barvě kůže v různých barevných prostorech lze vidět na obr. 4.3.



Obr. 4.3. Rozdělení prvků barevně odpovídajících barvě kůže v barevných prostorech *RGB* a *YCbCr* (vytvořeno v programu *ColorSpace* [8]).

Pomocí této *binární matice* jsme nyní schopni na základě srovnávání jednotlivých pixelů testovaného obrazu odpovídajícím indexům této matice vytvořit tzv. *binární mapu*, informující nás, které pixely z testovaného obrazu spadají do barevné oblasti kůže a které nikoli. Výhodou použití množiny vzorů jednotlivých barev, popisující barevnou oblast kůže, oproti matematickému modelu [23] je rychlost zpracování obrazu, ovšem za cenu větších nároků na paměť (je nutno mít v paměti uloženu databázi o velikosti přibližně 16 Mb).

Tuto binární matici lze však použít jen za předpokladu, že mimo obličejovou oblast se budou nacházet pixely barevně odpovídající barvě kůže jen velmi zřídka a navíc se nebudou shlukovat ve větších významnějších celcích. Tyto podmínky lze splnit jen při umělém vytvoření snímané scény, což by v praxi nebylo příliš využitelné. Z tohoto důvodu byl počet obrazových vzorků kůže rozšířen o stejný počet obrazových vzorků různých částí pozadí a vytvořena *pravděpodobnostní matice,* kde index každého prvku této matice zastupuje odpovídající barevnou hodnotu z barevného prostoru *YCbCr* a hodnota prvku odpovídá pravděpodobnosti výskytu příslušné barevné kombinace v množině vzorků kůže. Pravděpodobnost jednotlivých prvků byla získána dle Bayesova teorému (4.2):

$$P(1\,|\,M_S(x,y,z)) = \frac{P(M_S(x,y,z)\,|\,1)\cdot P(1)}{P(M_S(x,y,z)\,|\,1)\cdot P(1) + P(M_S(x,y,z)\,|\,0)\cdot P(0)}, \tag{4.2}$$

kde *x,y,z* jsou indexy matice M_S, které odpovídají jednotlivým složkám daného barevného prostoru, $P(1|\ M_S(x,y,z))$ je pravděpodobnost, že prvek s indexy jednotlivých složek daného

barevného prostoru *x,y,z* spadá do barevného podprostoru barevně odpovídajícímu barvě kůže, $P(M_S(x,y,z)|1)$ a $P(M_S(x,y,z)|0)$ odpovídá pravděpodobnosti výskytu prvku s indexy *x,y,z* v množině vzorků kůže a množině vzorků, které kůži neodpovídají a pravděpodobnost $P(1)$ a $P(0)$ dává pravděpodobnost výskytu vzorků barvy kůže a vzorků jiných, v našem případě $P(0) = P(1) = 0{,}5$.

Detekce barvy kůže a lokalizace obličeje

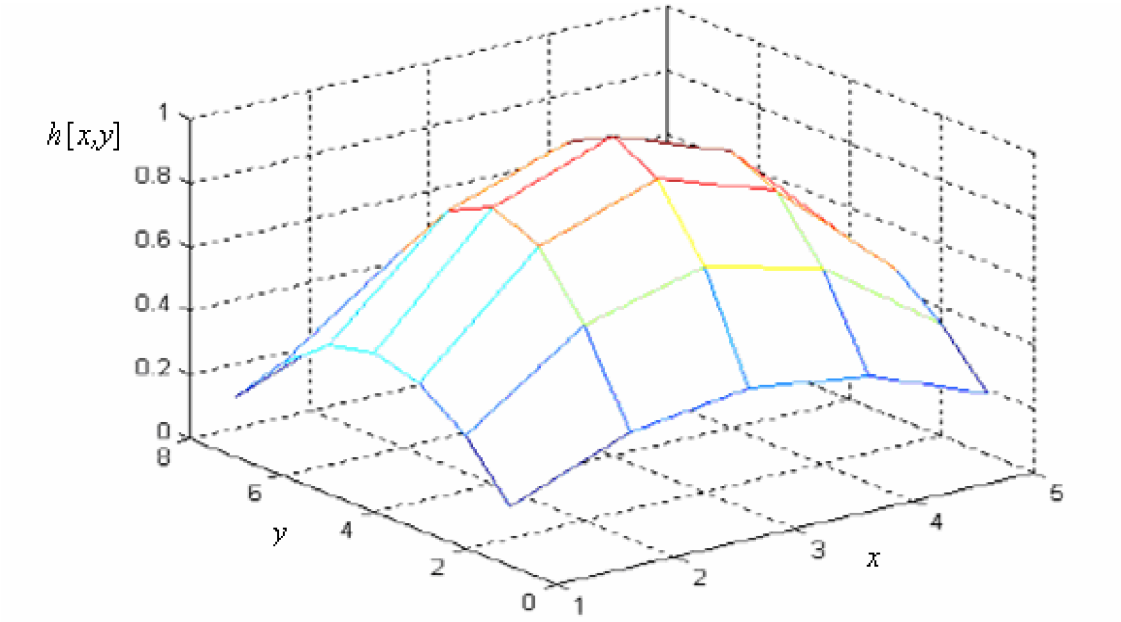
Na rozdíl od práce s binární maticí, kdy jsou jednotlivé barevné kombinace striktně rozděleny na barvy spadající do barevného prostoru kůže a barvy do tohoto prostoru nespadající, tak při detekci obličeje pomocí pravděpodobnostní matice se vychází z předpokladu, že pixely spadající do obličejové oblasti budou mít v pravděpodobnostní matici vysokou hodnotu pravděpodobnosti a dále pak se tato hodnota nebude v okolních pixelech příliš lišit (z důvodu barevné homogenity kůže). Postup je tedy následující: obraz je rozdělen na bloky o velikosti 3×3 pixelů, pro každý pixel v příslušném bloku je určena podle pravděpodobnostní matice hodnota pravděpodobnosti, že se svou barevnou hodnotou spadá do barevné oblasti kůže, a na základě porovnání prvního a druhého statistického momentu pravděpodobnosti všech pixelů v daném bloku s adaptivním prahem přiřadíme tomuto bloku buď jedničku (vysoká střední hodnota, nízký rozptyl – blok spadá do barevné oblasti kůže) nebo nulu (nízká střední hodnota, vysoký rozptyl – blok nespadá do barevné oblasti kůže). Tímto způsobem je získána binární mapa odpovídající výskytu oblastí barevně spadající do barevné oblasti kůže, která má oproti původnímu obrazu třetinovou velikost, čímž se sníží výpočetní nároky při lokalizaci obličeje [73]. Na obr. 4.4 je uveden příklad použití binární a pravděpodobnostní matice při vytvoření binární mapy obličeje na základě barvy kůže.

Obr. 4.4. Příklad vytvoření binární masky obličeje na základě barvy kůže (a) originální vstupní obraz, b) binární mapa za použití binární matice, c) binární mapa za použití pravděpodobnostní matice).

Jak je vidět z obr. 4.4 c), tak i přes použití pravděpodobnostní matice se v obraze nachází významné množství bodů nespadajících do obličejové oblasti. Aby byl počet těchto bodů zredukován, je jasová složka *Y* testovaného obrazu podrobena waveletové transformaci [80]. Změny jasu v obličeji jsou ve vertikálním směru v oblasti očí, rtů, apod. velmi výrazné, zatímco v horizontálním směru jsou zanedbatelné. Vycházíme přitom z předpokladu, že v obličeji lze ve vertikálním směru detekovat velmi výrazné přechody jasu

v oblasti očí a rtů, zatímco v horizontálním směru jsou tyto přechody zanedbatelné (zvláště pokud obličej splývá s pozadím). Z tohoto důvodu jsou tedy použity pouze detailní koeficienty waveletové transformace ve vertikálním směru prvního stupně rozkladu za použití vlnky Daubechies druhého řádu, kde vysoké hodnoty koeficientů určují přítomnost ostrého přechodu (očí, rtů, atd.). Aby tyto vysoké hodnoty pokryly větší plochu obličeje, je použita dvourozměrná filtrace obrazu dolní propustí s impulsní odezvou definovanou funkcí (4.3) viz. obr. 4.5. Takto získané pole koeficientů je upraveno na velikost binární mapy a s touto mapou vynásobíme.

$$h[x,y] = \sin\left(2 \cdot \pi \cdot \frac{x}{6}\right) \cdot \sin\left(2 \cdot \pi \cdot \frac{y}{8}\right), \text{ pro } x = 1,..6 \text{ a } y = 1,..8. \tag{4.3}$$



Obr. 4.5. Impulsní charakteristika filtru typu dolní propust.

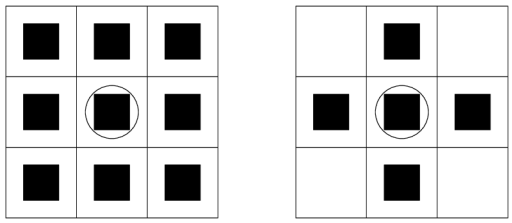
Pomocí waveletové transformace byly redukovány oblasti bodů, které byly ve vertikálním směru homogenní. Ovšem i přesto v binární mapě mohou zůstat oblasti bodů, které nepatří do obličejové oblasti a doposud splňovaly všechna stanovená kritéria. Z tohoto důvodu je provedena série morfologických operací uzavření a otevření (4.4) a (4.5).

$$X \bullet B = (X \oplus B) \Theta B, \tag{4.4}$$

$$X \circ B = (X \Theta B) \oplus B, \tag{4.5}$$

kde $\oplus$ odpovídá morfologické dilataci a Θ morfologické erozi, X je vstupní obraz a B je strukturní element [6].

Matematická morfologie poskytuje nástroje pro extrakci požadovaných částí obrazu a je založena na nelineárních operacích v obrazu. Každá operace je vnímána jako transformace mezi obrazem a tzv. strukturním elementem, viz obr. 4.6. Strukturní element představuje, stejně jako obraz, množinu bodů. Od vstupního obrazu se však liší svou výrazně menší velikostí. Mezi základní morfologické oprace patří dilatace a eroze.

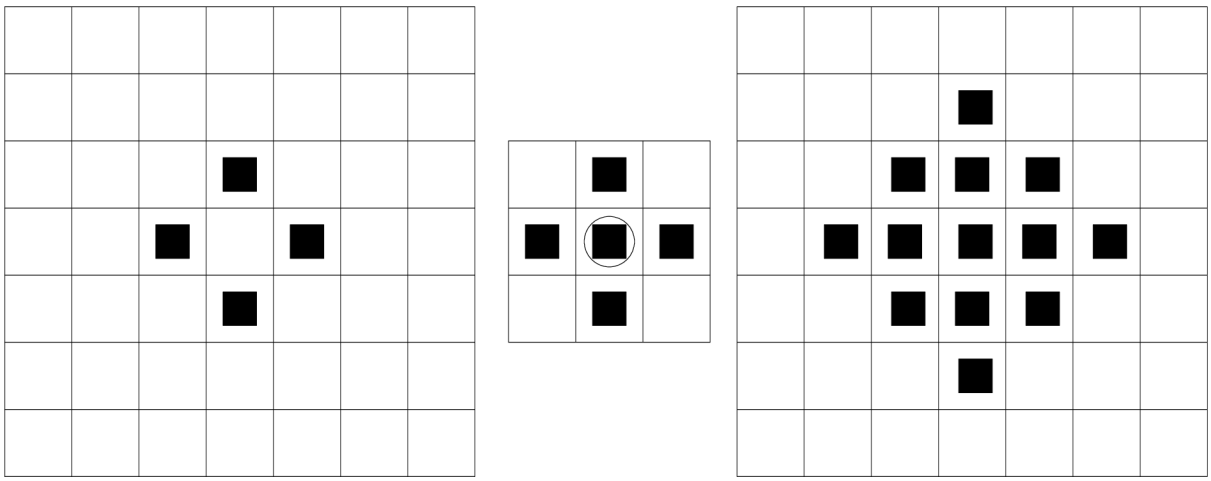


Obr. 4.6. Příklad dvou nejčastěji používaných strukturních elementů (velikost strukturních elementů je 3×3, černé čtverce reprezentují hodnotu 1, bílá místa pak hodnotu 0, střed strukturního elementu je vyznačen kruhem).

Dilatace

Tato morfologická operace sčítá dvě množiny - strukturní element *B* a vstupní obraz *X*. Tento množinový součet je definován rovnicí (4.6) [6] a lze jej zjednodušeně popsat jako porovnání vstupního obrazu se strukturním elementem, který je v obraze posouván. V případě, že je hodnota středu strukturního elementu rovna odpovídajícímu bodu vstupního obrazu, je tento obraz sečten se strukturním elementem viz obr. 4.7. Efektem binární dilatace je zvětšení objektu, zatímco díry v objektu se zaplňují.

$$X \oplus B = \{p \in \varepsilon^2 : p = x + b, x \in X, b \in B\}. \tag{4.6}$$



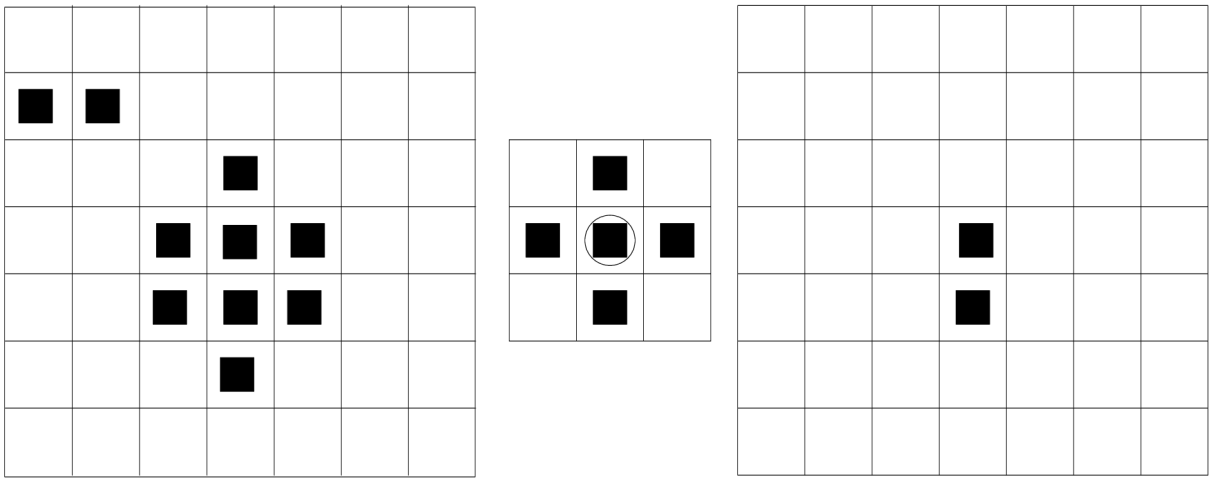
a) b) c)

Obr. 4.7. Příklad použití morfologické operace dilatace a) originální obraz, b) strukturní element, c) výsledek operace dilatace (černé čtverce reprezentují hodnotu 1, bílá místa pak hodnotu 0, střed strukturního elementu je vyznačen kruhem).

Eroze

Eroze je definována jako průnik všech posunů strukturního elementu B se vstupním obrazem X. Tato operace je definována rovnicí (4.7) [6] a její základní funkcí je ořezání hranic objektu. Objekt se touto operací zmenšuje, díry v objektech se naopak zvětšují a objekty menší než strukturní element jsou odstraněny viz obr. 4.8. Eroze se často využívá k rozložení objektu na jednodušší části a tím k zjednodušení struktury.

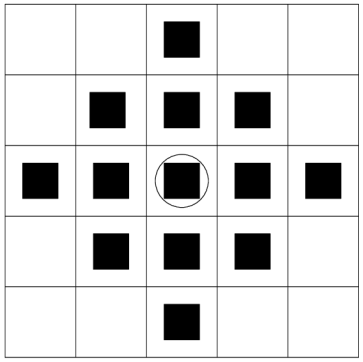
$$X \Theta B = \{p \in \varepsilon^2 : p + b \in X, \forall b \in B\}. \tag{4.7}$$



a) b) c)

Obr. 4.8. Příklad použití morfologické operace eroze a) originální obraz, b) strukturní element, c) výsledek operace eroze (černé čtverce reprezentují hodnotu 1, bílá místa pak hodnotu 0, střed strukturního elementu je vyznačen kruhem).

V našem případě byl použit strukturní element o velikosti 5×5 viz obr. 4.9. Použití po sobě jdoucích operací otevření a uzavření má za následek odstranění samostatně stojících bodů a seskupení ostatních bodů do kompaktních celků.



Obr. 4.9. Ukázka použitého strukturního elementu (černé čtverce reprezentují hodnotu 1, bílá místa pak hodnotu 0, střed strukturního elementu je vyznačen kruhem).

Tyto celky jsou poté pomocí jednoduchého hranového detektoru prostorově popsány pomocí pravoúhlých čtyřúhelníků. Na základě poměru velikostí jejich jednotlivých stěn jsou vyřazeny oblasti, které tvarově nevyhovují reprezentaci antropologickému modelu obličeje [67], a dále pak oblasti, jež jsou příliš malé na to, aby poskytly o případném obličeji více informací, tj. s oblasti s menším rozlišením než 20×20 pixelů (viz výše stanovená kriteria). Na obr. 4.10. jsou znázorněny jednotlivé binární mapy po průměrování, násobení waveletovými koeficienty a po morfologických operacích s výběrem vhodných obličejových kandidátů.

Obr. 4.10. a) zprůměrňovaná binární mapa, b) vynásobená zprůměrovanými waveletovými koeficienty, c) po provedení morfologických operací s výběrem kandidátských oblastí.

Jak je patrno z obr. 4.10 c) tak stále ještě nemusely být vyloučeny všechny oblasti, které nejsou oblastmi obličejovými. Na závěr je tedy proveden ověřovací test pro vybrané kandidáty, a to porovnání nalezené pozice úst a očí s jejich umístěním v antropologickém modelu obličeje. Nejprve však musí být pozice těchto obličejových částí v dané kandidátské oblasti přesně lokalizovány. K tomu problému je přistupováno pomocí tzv. barevných map jednotlivých kandidátských oblastí [74].

Lokalizace rtů a očí

Při lokalizaci úst je opět využito barevné informace obrazu a je provedena barevná transformace do normalizovaného barevného prostoru *rgb* dle transformačního vztahu (4.8), kde vycházíme z předpokladu, že barevná skladba rtů se sestává z vysoké hodnoty červené a velmi nízké modré barvy barevného prostoru *RBG*. Ze vztahu (4.8) lze vidět, že jednotlivé složky tohoto barevného prostoru jsou navzájem korelované, z toho důvodu stačí využít pouze normalizované složky *g*. Jelikož se barevná skladba rtů u různých lidí liší, nejsme tedy schopni postihnout všechny barevné kombinace pouze jednou složkou. Z tohoto důvodu je zavedena další barevná transformace do jednorozměrného barevného prostoru *FLD* (4.9) [27].

$$\begin{aligned} r[x,y] &= \frac{R[x,y]}{(R[x,y]+G[x,y]+B[x,y])}, \\ g[x,y] &= \frac{G[x,y]}{(R[x,y]+G[x,y]+B[x,y])}, \\ b[x,y] &= \frac{B[x,y]}{(R[x,y]+G[x,y]+B[x,y])}. \end{aligned} \tag{4.8}$$

$$FLD[x,y] = [-0.289 \quad 0.379 \quad 0.038] \cdot \begin{bmatrix} R[x,y] \\ G[x,y] \\ B[x,y] \end{bmatrix}. \tag{4.9}$$

Výsledkem jsou dvě podobné barevné mapy úst dosahující v oblasti odpovídající rtům velmi nízkých hodnot viz. obr. 4.11. Nevýhoda použití těchto barevných transformací spočívá v možnosti výskytu části pozadí s velmi vysokou hodnotou červené a velmi nízkou hodnotou modré barvy (např. červený šátek kolem krku). Využijeme-li však faktu, že máme k dispozici binární masku předpokládaného obličeje a že ústa leží vždy celé uvnitř obličeje, můžeme tedy zpracovávat pouze ty části obrazu, které se nacházejí uvnitř binární masky obličeje. Dalším problémem při lokalizaci rtů může být případ, kdy se červená barva ve větším množství vyskytuje na samotném obličeji (např. silně načervenalé tváře). Využijeme-li však již provedené waveletové transformace ve vertikálním směru (rty se ve vertikálním směru jeví jako ostrý přechod, zejména pokud jsou ústa otevřená), můžeme waveletovými koeficienty vynásobit obě barevné mapy a tím tyto nepříznivé vlivy potlačit [81].

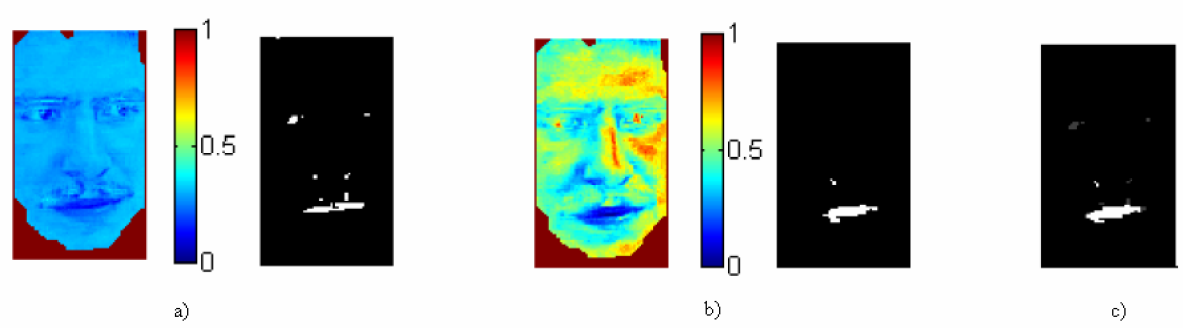
Následně jsou jednotlivé mapy převedeny na binární obrazy X_1 a X_2 pomocí prahování, kde hodnota prahu pro jednotlivé barevné mapy je stanovena pomocí histogramu na základě 10% hranice četnosti výskytu (tj. hodnota prahu je rovna indexu histogramu při němž součet všech prvků od počátku až po tento index odpovídá 10% celkového součtu všech prvků daného histogramu) viz (4.10).

$$\sum_{n=0}^{p} h[n] = 0{,}1 \cdot \sum_{n=0}^{N} h[n], \tag{4.10}$$

kde p je hledaný práh, $h[n]$ je hodnota histogramu v bodě n (tj. četnost výskytu hodnoty n v dané barevné mapě) a N je celkový počet prvků histogramu (tj. maximální hodnota v dané barevné mapě).

Tyto dva binární obrazy jsou poté sloučeny pomocí vztahu (4.11) na výslednou binární mapu X viz obr. 4.11. Pozice rtů pak odpovídá pozici maximální četnosti pozitivního předpokladu výskytu rtů v celkové binární mapě, která může být určena pomocí vzájemné korelace celkové binární mapy s obdélníkovým segmentem odpovídajícímu předpokládanému tvaru rtů (odvozeno z antropologického modelu obličeje [67]).

$$X[x,y] = \frac{1}{4} \cdot \left(X_1[x,y] + X_2[x,y]\right) + \frac{3}{4} \cdot X_1[x,y] \cdot X_2[x,y]. \tag{4.11}$$

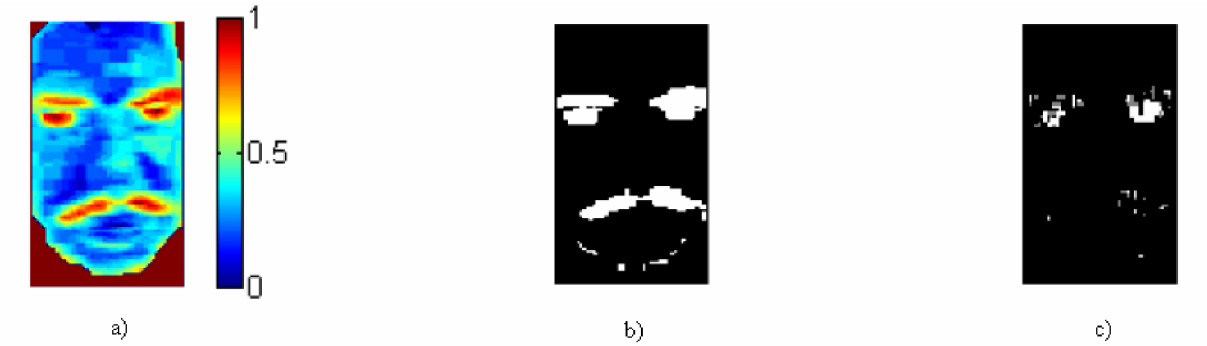


Obr. 4.11. a) barevná mapa složky g a její binární obraz, b) barevná mapa složky FLD a její binární obraz, c) výsledný binární obraz dle rovnice (4.11).

Postup při lokalizaci očí je podobný. Avšak jelikož je barva zornice i bělma u většiny lidí velmi podobná, je zde použita pouze jedna barevná mapa M odvozená z barevného modelu *YCbCr* (4.12) [25], kde oblasti odpovídající očím dosahují vysokých hodnot viz obr. 4.12. I zde je využita binární maska obličeje (rušivé vlivy pozadí) a waveletové koeficienty (pro případ přivřených očí, kdy se uplatní vliv obočí a očních řas). Barevná mapa je opět převedena na binární obraz pomocí prahování (hodnota prahu je získána stejným způsoben jako v případě rtů, pouze hledáme horní hranici histogramu viz (4.13)) a určeny pozice obou očí pomocí vzájemné korelace se čtvercovým segmentem (odvozen z antropologického modelu obličeje). Je nutné brát v potaz také pozici rtů, protože zuby dosahují také velmi vysokých hodnot v barevné mapě očí, z toho důvodu není oblast kolem již lokalizovaných rtů prohledávána.

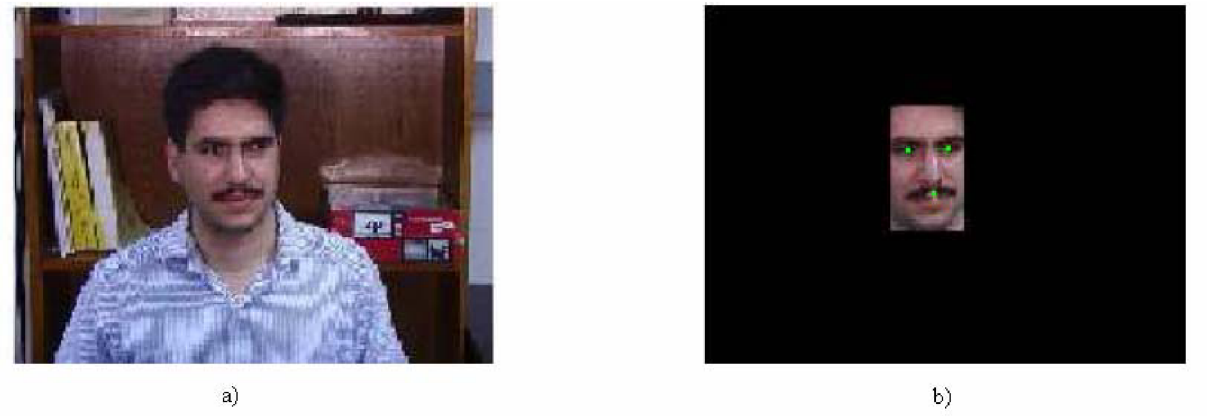
$$M[x,y] = \frac{1}{3} \cdot Cb^2[x,y] + (256 - Cr[x,y])^2 + \frac{Cb[x,y]}{Cr[x,y]}. \tag{4.12}$$

$$\sum_{n=N}^{p} h[n] = 0{,}1 \cdot \sum_{n=0}^{N} h[n]. \tag{4.13}$$



Obr. 4.12. a) barevná mapa očí, b) její binární obraz, c) výsledný binární obraz po vynásobení waveletovými koeficienty.

Na základě porovnání vzájemných pozic a vzdáleností mezi rty, oběma očima a s přihlédnutím k vertikálním a horizontálním rozměrům předpokládaného obličeje s antropologickým modelem obličeje je stanoveno, zda se o obličej jedná či nikoli. V případě souhlasu jsou souřadnice obličeje v obraze včetně jeho rozměrů a pozic očí a rtů uchovány pro další zpracování a obličej je zahrnut do výsledného obrazu viz obr 4.13. V opačném případě je kandidátská oblast zavržena. Touto detekcí jsou postupně podrobeny všechny kandidátské oblasti.



Obr. 4.13. a) originální obraz, b) detekovaná tvář s vyznačenou přibližnou pozicí očí a rtů (zelené tečky).

Testování a výsledky

Pro účely testování byla zvolena obrazová databáze obličejů *GTFD Georgia Tech Face Database* [44], která obsahuje 750 snímků 50 subjektů (muži i ženy různé národnosti a věku). Rozlišení jednotlivých snímků je 640×480 pixelů, přičemž průměrná velikost obličejů je přibližně 150×150 pixelů. Tato databáze byla zvolena z toho důvodu, že obsahuje komplexní pozadí kombinované s pozadím barevně odpovídajícím barvě kůže, což představuje pro algoritmy lokalizace obličejů založených na detekci barvy kůže značný problém. Dále pak databáze obsahuje značnou variabilitu jednotlivých subjektů (různé úhly natočení hlavy, emocionální výrazy, nasazení brýlí, různé osvětlení, atd.) viz obr. 4.14.

Obr. 4.14. Ukázka z obrazové databáze GTFD (nahoře variabilita jednoho subjektu, dole variabilita subjektů).

Aby bylo možno ohodnotit kvalitu detektoru, je nutné zavést několik označení. Budeme zde přistupovat k detekci jako k úloze binární klasifikace, kdy dochází k rozhodnutí, zda určitá oblast obsahuje obličej či nikoli. Pak tedy můžeme zavést množiny C_P a C_N, které obsahují oblasti obličejů (pozitivní množina C_P) a oblasti pozadí (negativní množina C_N). Velikost množiny C_P je v našem případě rovna počtu tváří tj. celkovému počtu testovaných obrazů, P=750. Budeme-li uvažovat jednotlivé části pozadí každého obrazu jako jeden celek, pak velikost množiny C_N bude také rovna celkovému počtu testovaných obrazů N=750. Dále počet případů kdy došlo ke správné detekci obličeje udává *TP* (*true positive*) a počet případů kdy došlo ke správné detekci okolního pozadí (tj. žádná část pozadí nebyla detekována jako obličej) *TN* (*true negative*). Odtud můžeme odvodit dva možné typy chyb detekce [57]:

- Chyba I. typu – označuje se *FP* (*false positive*) a nastává v případech, kdy detektor chybně určí třídu C_P (určí pozadí jako tvář).
- Chyba II. typu – označuje se *FN* (*false negative*) a dochází k ní tehdy, pokud detektor chybně určí třídu C_N (určí tvář jako pozadí).

Tyto základní ukazatele lze zapsat do tzv. čtyřpolní tabulky, viz tab. 4.1. Čísla na hlavní diagonále reprezentují správnou detekci, čísla na vedlejší diagonále vyjadřují chybu mezi dvěma třídami (detekcí tváře a pozadí).

Tab. 4.1. Kontingenční tabulka – čtyřpolní tabulka.

		Skutečná třída	
		Pozitivní	Negativní
Klasifikovaná třída	Pozitivní	TP	FP
	Negativní	FN	TN

Důležitou veličinou je senzitivita nebo také detekční poměr (*detection rate, true positive rate, TPR*). Senzitivita je procentuálním vyjádřením počtu tváří v obrazech, které byly detektorem správně detekovány [57].

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN}, \quad TPR \in \langle 0,1 \rangle . \tag{4.14}$$

Dále pak specificita (*true negative rate, TNR*), která vyjadřuje poměr chybně určených tříd C_P k celkové velikosti množiny C_N [57].

$$TNR = \frac{TN}{N} = \frac{TN}{TN + FP}, \quad TNR \in \langle 0,1 \rangle . \tag{4.15}$$

Přesnost (*accuracy*, *ACC*) udává poměr správně určených tříd k celkovému počtu tříd [57].

$$ACC = \frac{TP + TN}{N + P} = \frac{TP + TN}{TP + TN + FP + FN}, \quad \mathrm{ACC} \in \langle 0,1 \rangle . \tag{4.16}$$

Pozitivní prediktivní hodnota (*positive predictive value, PPV*) určuje poměr správně pozitivně detekovaných ku všem pozitivně detekovaným [57].

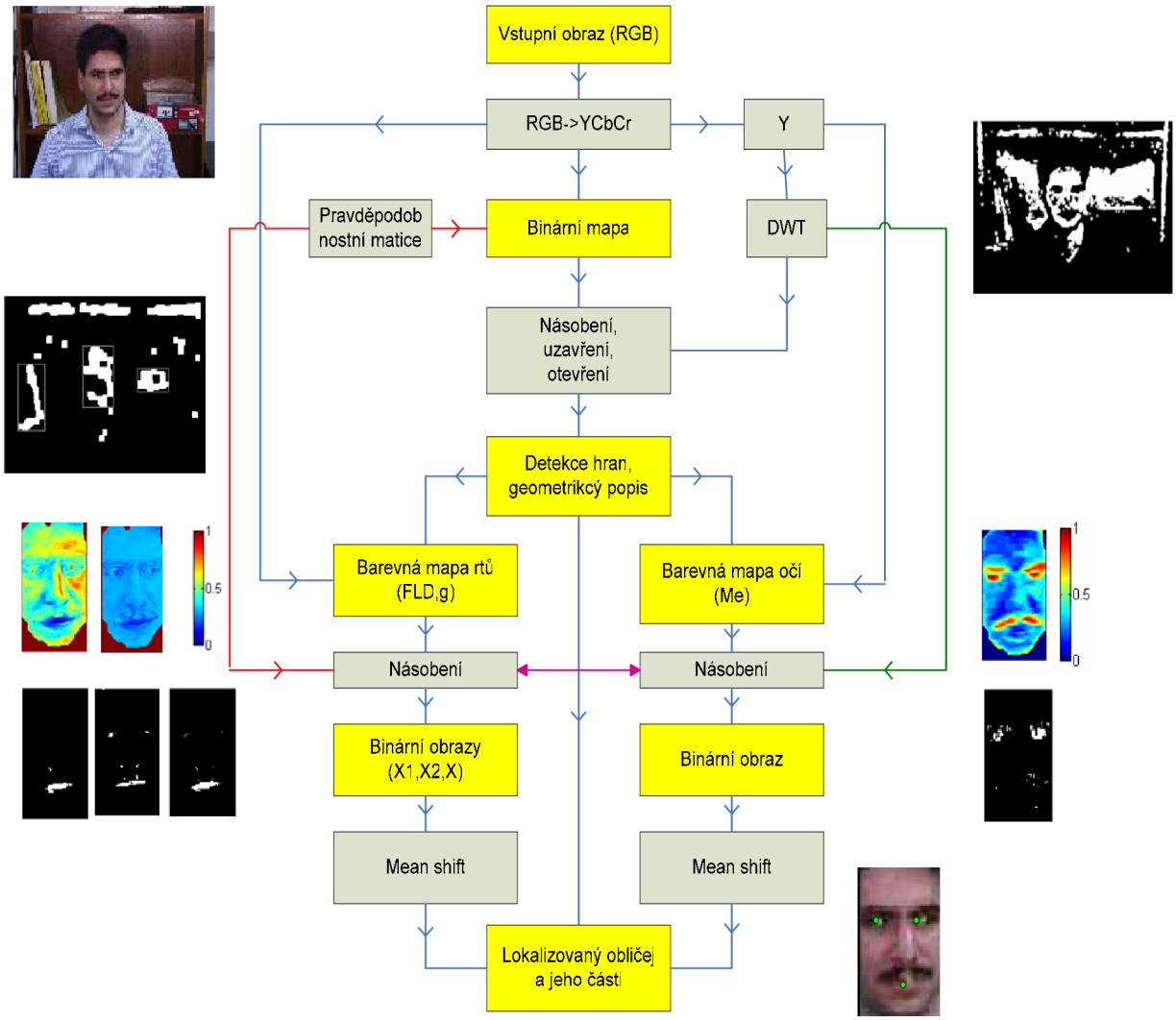
$$PPV = \frac{TP}{TP + FP}, \quad PPV \in \langle 0,1 \rangle . \tag{4.17}$$

V ideálním případě bude mít detektor hodnoty *TPR, TNR, ACC, PPV* rovny maximální hodnotě tj.1. V reálných aplikacích není většinou možné dosáhnout těchto hodnot, proto se hledá vhodný kompromis mezi senzitivitou *TPR* a specificitou *TNR*. V našem případě byla preferovaná vyšší hodnota specificitity *TNR* na úkor senzitivity *TPR*, tj. aby nedocházelo k dalšímu zpracování částí obrazu, které ve skutečnosti nejsou obličejem. Dosažené výsledky jsou vedeny tab. 4.2.

Tab. 4.2. Výsledky testování detektoru obličejů založeného na detekci barvy kůže na obrazové databázi GTFD.

TP = 669	*TPR* = 89,2 %
TN = 732	*TNR* = 97,6 %
FP = 18	*ACC* = 93,4 %
FN = 81	*PPV* = 97,4 %

Při testování tedy bylo dosaženo celkové přesnosti 93,4 %. Jak bylo výše zmíněno, projevil se zde vyšší podíl nesprávné detekce obličeje než nesprávné detekce pozadí. Hlavní příčinou neúspěšné detekce obličeje bylo zejména nedostatečné potlačení barevného vlivu pozadí na obličej pomocí waveletových koeficientů, kdy nebylo možno separovat obličej od pozadí. Dále pak se negativně projevila přítomnost hustých vousů, kdy rty již nebyly zařazeny do obličejové oblasti a obličejový kandidát musel být tedy zavržen. Celkový přehled algoritmu je zobrazen na obr. 4.15.



Obr. 4.15. Přehled algoritmu pro lokalizaci obličeje a jeho částí (obrázky přísluší žlutým blokům).

4.2 Detekce obličeje pomocí objektového detektoru Viola-Jones

V předchozí podkapitole byl představen návrh a realizace detektoru obličejů na základě detekce barvy kůže. Tento typ detekce se používá zejména v obrazech obsahujících pozadí, které se od barvy kůže výrazně odlišuje. V tomto případě jsou výpočetní nároky poměrně nízké a celková detekce dostatečně rychlá. Avšak pro účely této práce barva pozadí nebyla striktně definována, takže možný výskyt pozadí barvě odpovídající barvě kůže musel být předpokládán. Z tohoto důvodu byl vlastní algoritmus detekce barvy kůže rozšířen a doplněn o další metody vedoucí k vyšší robustnosti celkového detektoru, což ovšem mělo za následek zvýšení výpočetních nároků a s tím související vyšší čas potřebný k samotné detekci. V případě vstupního obrazu o velikosti 640×480 pixelů byl čas detekce na počítači P4 2,8 GHz průměrně 1,2 s. Dále pak byly dosud uvažovány pouze barevné složky obrazu, zatímco jasová složka měla pouze doplňující funkci, ačkoliv i na základě výhradně jasové složky jsme snadno schopni detekovat v obraze obličej.

Tato část se zabývá implementací detektoru obličejů založeném na objektovém detektoru *Viola-Jones*, s možností využití barevné informace obrazu a shlukové analýzy ke zvýšení robustnosti tohoto detektoru. Dále pak je objektový detektor *Viola-Jones* experimentálně využit pro lokalizaci očí v nalezených obličejích s využitím barevné oční mapy popsané v předchozí podkapitole.

Objektový detektor Viola-Jones

Objektový detektor *Viola-Jones* byl poprvé představen P. Violou a M. Jonesem v roce 2001 [63]. Jedná se o detektor objektů pracující s šedo-tónovými obrazy (popř. jasovou složkou obrazu), který se skládá ze tří základních částí: integrálního obrazu, Haarova waveletu a klasifikačního algoritmu *AdaBoost*. Výhodou tohoto detektoru je rychlost, dostatečná spolehlivost a značná nezávislost na osvětlení a velikosti sledovaného objektu. Z těchto důvodů je v praxi tento detektor často používán například při detekci obličejů a zároveň vzniká velká řada jeho modifikací (např. využití jiného než Haarova waveletu, nahrazení algoritmu *AdaBoost* algoritmem *GentleBoost* a jiné) [22].

AdaBoost

AdaBoost (název je zkratkou pro *Adaptive Boosting*) je klasifikační algoritmus [59], který vychází z metody strojového učení zvaného *boosting*. Cílem metody *boosting* je zlepšení klasifikační přesnosti libovolného algoritmu strojového učení. Základem je vytvoření více klasifikátorů označovaných jako slabí žáci (*weak learners*). Tyto klasifikátory vznikají pomocí výběru vzorků ze základní trénovací množiny. První klasifikátor má přesnost jen o málo větší než je přesnost odhadu (tj. přes 50 % v případě dvoustavového klasifikátoru). Postupně jsou pak přidávány další klasifikátory s obdobnou mírou přesnosti, čímž je vygenerován soubor klasifikátorů označovaný jako silný žák (*strong learner*), jehož celková

klasifikační přesnost je libovolně vysoká vzhledem ke vzorkům v trénovaní množině, tj. klasifikace byla zesílena (*boosted*) [17].

Klasifikační algoritmus *AdaBoost* tedy pro učení využívá slabých klasifikátorů $h_t(x)$, které jsou vybírány z množiny klasifikátorů H, a jejichž lineární kombinací vzniká nelineární silný klasifikátor $H(x)$. Vstupem algoritmu je trénovací množina S, která je složena z dvojic (x_i,y_i), kde x_i je získaná hodnota příznaku a y_i je třída odpovídající příznaku i, $y_i \in \{-1,1\}$ pro $i=1,\ldots M$, kde M je velikost trénovací množiny. *AdaBoost* na rozdíl od základního algoritmu *boostingu* používá vážení trénovací množiny váhami D_t, které jsou ze začátku nastaveny rovnoměrně a v každé smyčce algoritmu se pak provádí následující:

- Výběr slabého klasifikátor s nejmenší chybou klasifikace při daných váhách D_t,
- ověření, že chyba klasifikátoru nepřekročila hodnotu 0,5,
- výpočet koeficientu slabého klasifikátoru v lineární kombinaci $H(x)$,
- aktualizace jednotlivých vah D_t trénovací množiny.

Jak již bylo řečeno je slabý klasifikátor vybírán tak, aby jeho přesnost klasifikace byla o něco lepší, než kdyby prováděl klasifikaci náhodně. Pokud tedy chyba ε_t překročí hodnotu 0,5, pak tento požadavek není splněn a není tedy zaručeno, že algoritmus bude konvergovat. Aktualizace vah zapříčiní, že váha špatně klasifikovaných měření se zvětší a váha dobře klasifikovaných se zmenší. To znamená, že v následujícím kroku bude hledán slabý klasifikátor, který bude lépe klasifikovat doposud chybně provedená měření [63]. *Adaboost* tedy redukuje trénování chybu exponenciálně v závislosti na rostoucím počtu klasifikátorů. Zvyšování počtu klasifikátorů ve skupině ale může vést k tzv. *přetrénování* (tj. ztrátě schopnosti generalizovat vlivem přílišného zaměření klasifikátorů na rozeznávání pouze konkrétních trénovacích dat). Simulační experimenty však ukázaly, že k tomu relativně zřídka dochází i pro extrémně vysoké hodnoty počtu klasifikátorů [52]. Na obr. 4.16 je uveden stručný popis procesu učení algoritmu *AdaBoost*.

1. Vstup:

$$S = \{(x_1, y_1), \ldots, (x_m, y_m)\}, \text{počet iterací T}$$

2. Inicializace vah:

$$D_1(i) = \frac{1}{m}$$

3. Cyklus pro t = 1, …, T:

   a. Výběr klasifikátoru na základě vážené trénovací chyby

$$\varepsilon_j = \sum_{i=1}^{m} D_t(i) I\left[y_i \neq h_j(x_i)\right]$$

$$h_t = \arg\min_{h_j \in H} \varepsilon_j$$

   b. Pokud

$$\varepsilon_t = 0 \text{ nebo } \varepsilon_t \geq \frac{1}{2}, \text{pak konec cyklu}$$

   c. Nastavení

$$\alpha_t = \frac{1}{2} \log\left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right)$$

   d. Úprava vah

$$D_{t+1}(i) = \frac{D_t(i) e^{-\alpha_t y_i h_t(x_i)}}{Z_t}$$

$$\text{kde } Z_t = \sum_{i=1}^{m} D_t(i) e^{-\alpha_t y_i h_t(x_i)}$$

4. Výstup

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

Obr. 4.16. Proces učení klasifikačního algoritmu AdaBoost [63].

Haarův wavelet

Jak bylo zmíněno v předchozí části, vstupem procesu učení klasifikačního algoritmu je množina příznaků *S*. Čím je tato množina příznaků obsáhlejší, tím existuje větší pravděpodobnost výběru slabého klasifikátoru s vyšší mírou přesnosti (samozřejmě s ohledem na vzájemnou korelaci jednotlivých příznaků). Snahou detektoru *Viola-Jones* je získat velkou řadu jednoduchých příznaků s minimálními výpočetními nároky. Takovým typem příznaků jsou příznaky založené na principu podobnému definici *Haarova* waveletu (tzv. *Haar-like features*), viz obr. 4.17. Hodnota takového příznaku se tak vypočítá jako suma pixelů obrazu odpovídajících světlé části příznaku, od které je odečtena suma pixelů odpovídacích tmavé části. Tyto wavelety mohou být tvořeny dvěma (hranový příznak), třemi (čárový příznak) či čtyřmi (diagonální příznak) obdélníkovými oblastmi. Jednotlivé příznaky jsou použity na celý vstupní obraz, přičemž zároveň dochází ke změně velikostí jednotlivých příznaků (tj. velikosti jednotlivých obdélníků) z velikosti 1×1 až na velikost odpovídající vstupnímu obrazu. To znamená, že pro vstupní obraz o rozměrech 19×19 dostáváme přibližně 64 tisíc hodnot příznaků, které jsou vstupem učícího procesu klasifikačního algoritmu *AdaBoost*. Ten z nich potom vybere jen určité malé množství příznaků společně s se stejným počtem natrénovaných slabých klasifikátorů, pomocí nichž lze vstupní obraz vhodně klasifikovat (stejný příznak se ve výsledném silném klasifikátoru může vyskytovat i několikrát, ovšem pokaždé s jiným nastavením slabého klasifikátoru). Pouze toto malé množství příznaků je pak použito při samotné detekci. Uvedené typy příznaků patří k tzv. *základním* příznakům, v současnosti se používají i další typy ze *základních* příznaků odvozené [39].

Obr. 4.17. Příznaky podobné *Haarově* waveletu [63].

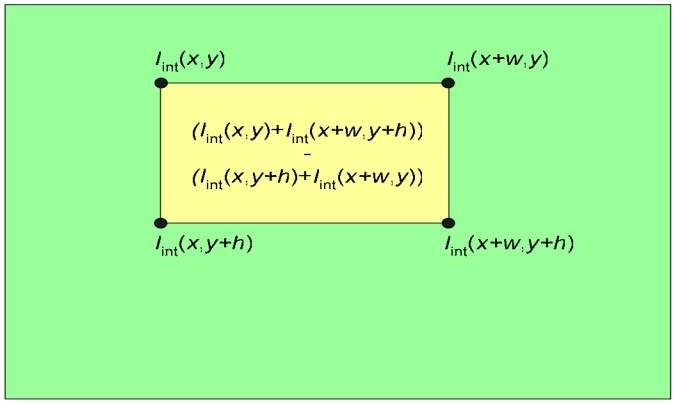
Integrální obraz

Integrální obraz slouží k rychlému a efektivnímu výpočtu hodnot jednotlivých příznaků ze vstupního obrazu. Aby pro každou hodnotu příznaku vstupního obrazu nemusela být počítána suma hodnot pixelů odpovídacích danému příznaku, je obraz převeden do reprezentace integrálního obrazu, kde každý bod tohoto obrazu odpovídá součtu hodnot všech předcházeních bodů dle [63]:

$$s(x,y) = s(x,y-1) + I(x,y),$$
$$I_{\text{int}}(x,y) = I_{\text{int}}(x-1,y) + s(x,y), \tag{4.18}$$

kde $s(x,y)$ je kumulovaný součet hodnot v řádku obrazu, $I(x,y)$ představuje hodnoty intenzit jednotlivých pixelů vstupního obrazu a $I_{int}(x,y)$ jsou jednotlivé hodnoty integrálního obrazu. Dále pak platí: $s(x,0) = 0$, $I_{int}(0,y)=0$.

Výpočet hodnot jednotlivých příznaků se pak výrazně zjednoduší, protože na výpočet sumy libovolného obdélníku v obraze postačí dvě operace sčítání a jedna operace odčítání viz obr. 4.18, kde x, y jsou počáteční souřadnice a w, h jsou šířka a výška požadovaného obdélníku.



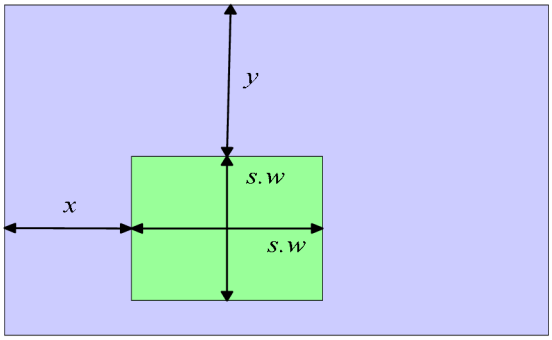
Obr. 4.18. Příklad výpočtu sumy libovolného obdélníku pomocí integrálního obrazu.

Detekce objektů v obraze

V předchozích částech byl popsán proces efektivní extrakce vhodných příznaků a způsob trénování klasifikátoru. Aby bylo možno detekovat určitý objekt v obraze je nejprve nutno vytvořit trénovací množinu obrazů obsahující pozitivní vzory objektu a negativní vzory pozadí (tj. vše co není sledovaným objektem). Všechny vzory by měly mít stejnou velikost/rozlišení $w \times h$ (pro zjednodušení dále budeme uvažovat, že $h=w$), zároveň by velikost w měla dosahovat co nejmenších hodnot (ovšem za předpokladu dostatečného optického rozlišení sledovaného objektu), a to z důvodu extrémní výpočetní náročnosti v procesu trénování. Pomocí těchto vzorů je natrénován klasifikační algoritmus *AdaBoost* (tj. získána množina optimálních příznaků jim odpovídající natrénovaná množina slabých klasifikátorů společně s jejich váhami α_t).

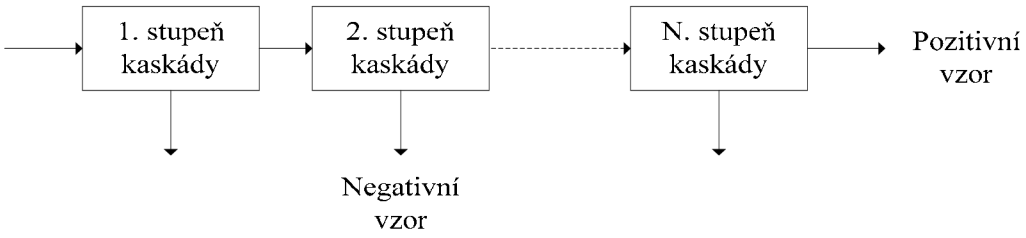
Při vlastní detekci není zpracováván vstupní obraz jako celek, ale po částech. Tyto části jsou vybírány pomocí posuvného pod-okna, které mění svoji pozici x, y napříč obrazem a zároveň mění i svoji velikost $s \times w$ v závislosti na předpokládané velikost detekovaného objektu viz. obr. 4.19, kde x a y jsou počáteční souřadnice daného pod-okna w je základní velikost okna odpovídající velikosti trénovacích vzorů a s je faktor zvětšení daného pod-okna. Vstupem klasifikátoru jsou vybrané příznaky tohoto pod-okna a klasifikátor rozhodne, zda toto pod-okno sledovaný objekt obsahuje či nikoli (resp. zda se vybrané

příznaky tohoto pod-okna shodují s příznaky pozitivních vzorů objektu trénovaní množiny obrazů).



Obr. 4.19. Příklad umístění pod-okna v testovaném obraze.

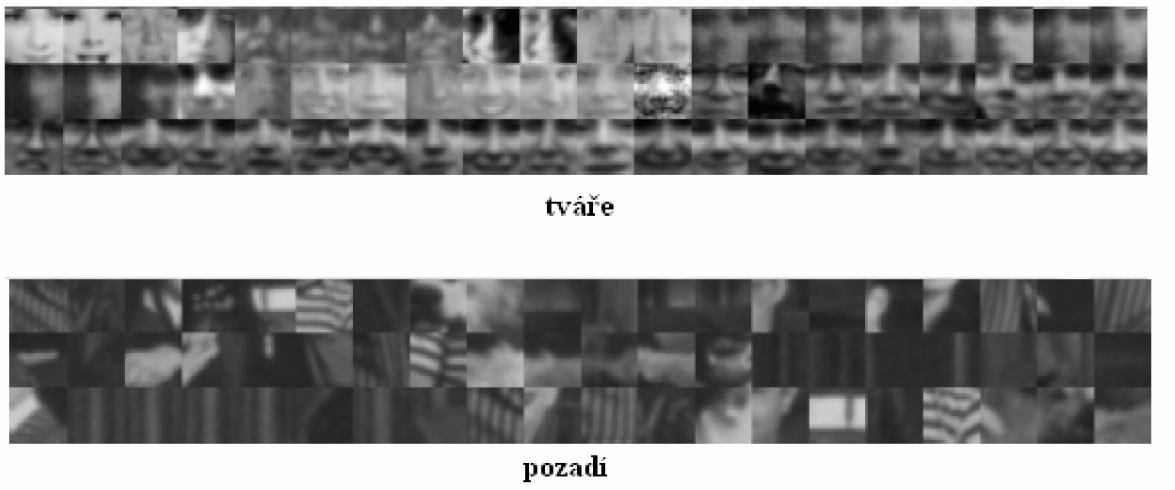
Tím, že je vstupní obraz prohledáván velkou řadou pod-oken, stává se detekce výpočetně náročná (např. pro obraz o rozměrech 320 × 240 pixelů je nutno zpracovat přibližně 500 000 pod-oken s počáteční velikostí w=19). Snaha o redukci výpočetní doby celkové detekce je řešena snížením průměrné doby, kterou detektor věnuje prohledávání každého pod-okna. Toho je dosaženo *kaskádovým* zapojením klasifikátorů [19]. Myšlenka tohoto řešení je založena na pozorování, že k vytvoření klasifikátoru, který dokáže vybrat téměř všechny pozitivní případy a zároveň množství (20 – 50%) negativních případů, stačí jen několik málo příznaků. Dalším důležitým postřehem pro vytvoření kaskády je skutečnost, že většina pod-oken v obraze jsou negativní, tedy neobsahují sledovaný objekt. Proto se v každém stupni kaskády snaží zamítnout co nejvíce negativních pod-oken, zatímco do dalších stupňů kaskády přechází jen pod-okna označena za pozitivní, viz obr. 4.20. Většina pod-oken obsahujících pozadí je tedy vyřazena již v několika prvních stupních kaskády a není tedy nezbytné extrahovat a klasifikovat hodnoty příznaků pro další stupně kaskády. Způsob trénování *kaskádního* detektoru se od klasického *monolitického* detektoru liší (viz níže).



Obr. 4.20. Příklad zapojení klasifikátorů do kaskády.

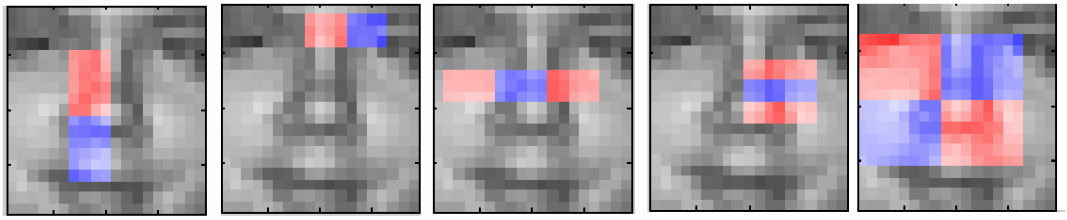
Realizace detektoru obličejů

Jak bylo již výše řečeno objektový detektor *Viola-Jones* se často používá k detekci tváří v šedo-tónových obrazech. Aby bylo možno takovýto detektor realizovat je nejprve nezbytné vytvořit soubor trénovacích vzorů obsahující různé pozitivní vzory objektů (obličeje) a negativní vzory pozadí. Při naší realizaci bylo využito trénovaní množiny z databáze MIT CVCL Face Database [64], která obsahuje 2429 pozitivních vzorů tváří a 4548 negativních vzorů pozadí v rozlišení 19 × 19 pixelů. Ukázka obrázků z této databáze je na obr. 4.21.



Obr. 4.21. Ukázka obrázků z databáze MIT CVLC.

Dále byl vytvořen soubor příznaků, jejichž hodnoty se použijí ve fázi trénovaní. V tomto případě jsme použili pouze *základní* typy příznaků, takže pro každý vzor extrahujeme 63 960 hodnot příznaků viz obr. 4.22.



Obr. 4.22. Příklad extrakce jednotlivých příznaků (daný příznak je roven rozdílu součtů hodnot pixelů označených červeně a modře).

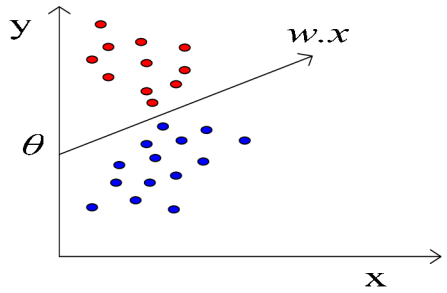
Jako slabý klasifikátor byl zvolen lineární perceptron s přenosovou funkcí definovanou dle [51]:

$$h(x_i) = sign(w \cdot x_i - \theta), \tag{4.19}$$

kde $h(x_i)$ je binární výstup perceptronu, x_i hodnota příznaku i-tého vzoru, θ práh a w váha peceptronu. Počáteční hodnota váhy w a hodnota prahu θ je stanovena náhodně. Proces trénování spočívá v úpravě váhy w, a to na základě dvojic z trénovací množiny $S = \{(x_1,y_1), (x_2,y_2), \ldots, (x_M,y_M)\}$, $y_i \in \{-1,1\}$ pro $i=1,\ldots M$, kde M je počet vzorů v trénovací množině, dle následující rovnice [52]:

$$w = w + y_i \cdot x_i, \quad \text{pokud } h(x_i) \neq y_i,$$
$$w = w, \quad \text{pokud } h(x_i) = y_i. \tag{4.20}$$

Toto pravidlo se postupně opakuje pro všechny dvojice z S. Pokud jsou prvky trénovací množiny lineárně separovatelné, tak po konečném počtu opakování je nalezena taková váha w, která pro daný práh θ správně klasifikuje každý vzorek z trénovací množiny. Chyba na trénovací množině bude tedy nulová. Na obr. 4.23 je ukázána geometrická reprezentace lineárního klasifikátoru na množině vstupních dat. Jedná se vlastně o nalezení optimální směrnice w separační přímky s hodnotou θ v bodě $x = 0$. Jelikož ve většině případů množina trénovacích vzorů lineárně separovatelná není, vybere se v každém kroku vždy klasifikátor s největší váženou mírou separability.



Obr. 4.23. Grafické znázornění separace dat pomocí lineárního perceptronu.

Abychom potlačili vliv osvětlení na proces detekce, je rozptyl intenzity každého vzoru z trénovací množiny normalizován [2]. Rozptyl σ^2 integrálního obrazu $I_{\text{int}}(x,y)$ se vypočítá podle (4.21), kde E představuje střední hodnotu:

$$\sigma^2 = E(I_{\text{int}}^2(x,y)) - E(I_{\text{int}}(x,y))^2. \tag{4.21}$$

Hodnoty příznaků jsou pak normalizovány pomocí rozptylu σ^2 následovně:

$$f_j'(s) = \frac{f_j(s)}{\sigma^2(s)}, \tag{4.22}$$

kde $f_j'(s)$ je hodnota j-tého příznaku vzoru s a f_j je pak hodnota normalizovaného příznaku.

Výsledný silný klasifikátor H je pak definován vztahem (4.23), kde $h_t(x_t)$ je slabý klasifikátor odpovídající t-tému příznaku, jehož hodnota je x_t, α_t je váha daného slabého klasifikátoru a P je práh silného klasifikátoru.

$$H = sign(\sum_{t=1}^{T}(\alpha_t \cdot h_t(x_t)) - P). \tag{4.23}$$

Celkový počet slabých klasifikátorů (příznaků) T může být buď stanoven přímo před samotným procesem trénování a nebo je odvozen od požadované úspěšnosti schopnosti klasifikace trénovací množiny vzorů (počet slabých klasifikátoru se neustále zvyšuje, dokud

není dosaženo požadované úspěšnosti klasifikace). V případě klasického monolitického detektoru je potom hodnota prahu P rovna:

$$P = \frac{1}{2} \cdot \sum_{t=1}^{T} \alpha_t . \tag{4.24}$$

V případě kaskádního detektoru obsahuje každá kaskáda nezávislý samostatný monolitický klasifikátor, jehož vstupem jsou všechny vzory klasifikované předchozím stupněm kaskády jako obličeje. Počet použitých slabých klasifikátorů T_k a práh silného klasifikátoru P_k, $k = 1 \ldots K$, je stanoven dle požadované přesnosti kaskády v závislosti na klasifikaci vzorů z trénovaní množiny viz. obr. 4.24. Celkový počet stupňů kaskády K může být opět určen přímo před vlastním procesem trénování a nebo je odvozen od požadované úspěšnosti schopnosti klasifikace trénovací množiny vzorů.

TP … požadovaná úspěšnost klasifikátoru při klasifikaci obličejů (*true positive*) ≈ 100%.
TN … požadovaná úspěšnost klasifikátoru při klasifikaci pozadí (*true negative*) = 20-50%.

1. Inicializace

T=0, *tp*=0, *tn*=0.

2. Opakuj dokud (*tp*<*TP*) nebo (*tn*<*TN*)

a) $T = T+1$.

b) Nalezení optimálního slabého klasifikátoru $h_T(x)$ a váhy α_T (viz. algoritmus *AdaBoost*)

c) $P = \sum_{t=1}^{T} \alpha_t$, *tp*=0 a *tn* = 100.

d) Opakuj dokud (*tp*<*TP*) a zároveň (*tn*>*TN*)

i) $P = P - \frac{P}{100}$.

ii) Stanovení úspěšností *tp* a *tn* na množině vstupních trénovacích vzorů.

3. Odstranění všech negativně klasifikovaných vzorů ze vstupní trénovaní množiny pro další stupně kaskády.

Obr. 4.24. Postup při trénování jednoho stupně kaskádního klasifikátoru.

Při trénování monolitického a kaskádního klasifikátoru byla stanovena požadovaná celková úspěšnost na 98%, a pro kaskádní detektor pak i *TP* = 99,9% a *TN*=30%. Monolitický klasifikátor obsahoval 180 slabých klasifikátorů a kaskádní klasifikátor pak 19 kaskádních stupňů s celkovým počtem 146 slabých klasifikátorů. Následně byl realizován proces klasifikace testovací sady obrazů z MIT CVCL Face Database v programovém prostředí Matlab. Tato testovací sada se skládala z 472 obrazů tváří a 5 tisíc obrázků pozadí.

Tab. 4.3. Výsledky klasifikace pomocí monolitického a kaskádního klasifikátoru.

Typ klasifikátoru	Úspěšnost klasifikace	Doba klasifikace
Monolitický	93 %	127.5 s
Kaskádní	92.3 %	23.5 s

Jak je možno vidět z tab. 4.3, je úspěšnost obou klasifikátorů téměř shodná a liší se tedy pouze dobou trvání klasifikačního procesu. Z tohoto důvodu budeme dále uvažovat pouze používání kaskádního klasifikátoru (detektoru).

V předchozí části je uveden princip detekce obličejů různých velikostí pomocí zvětšování posuvného pod-okna o faktor *s*. V našem případě byl zvolen jiný postup, místo zvětšování posuvného pod-okna o faktor *s*, dochází ke zmenšování (podvzorkování) celého vstupního obrazu o faktor 1/*s*. Jako nejvhodnější metoda použitá pro změnu měřítka vstupního obrazu byla zvolena metoda nejbližšího souseda (*nearest neighbor*) [60] viz. (4.25):

$$I'(x, y) = I(x \cdot s, y \cdot s), \tag{4.25}$$

kde I je původní šedo-tónový obraz a I' je zmenšený obraz o faktor s. Výhodou této metody je zejména její rychlost oproti jiným interpolačním metodám při zachování dostatečné kvality pro klasifikaci obrazu (zejména díky předpokládané homogenitě okolního prostředí). Nevýhodou pak je extrémně nízká odolnost proti případnému šumu v obraze.

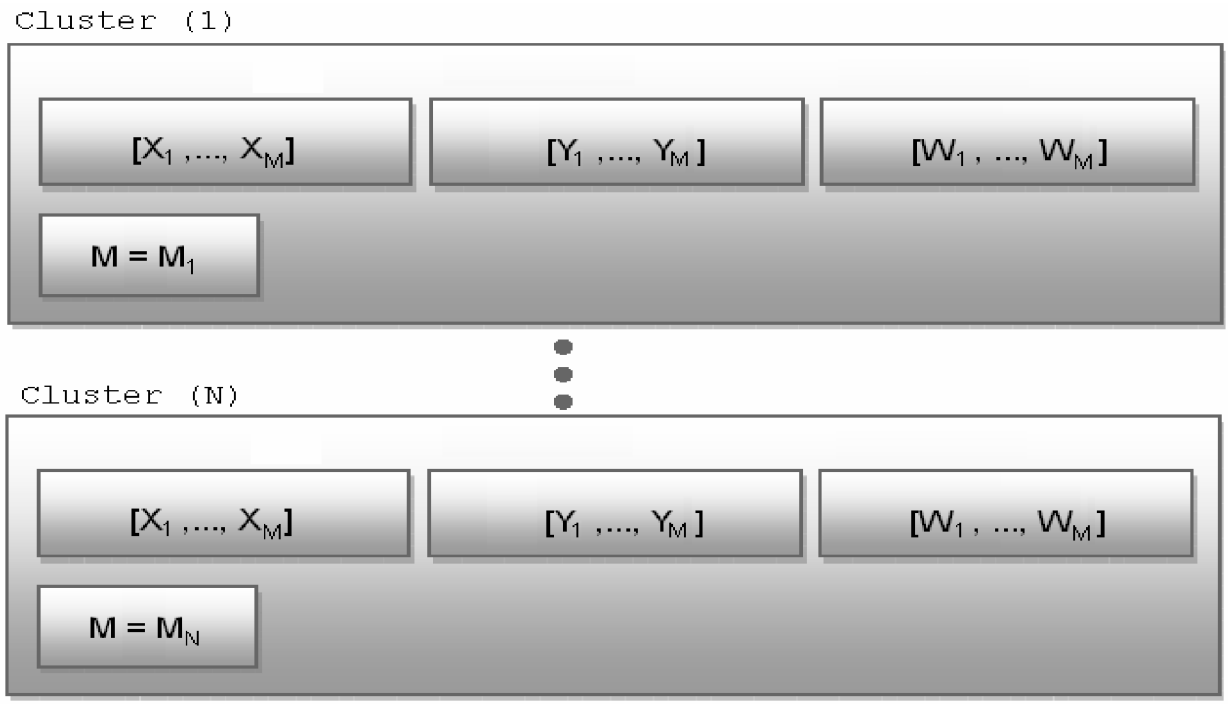
V prohledávaném obraze mohou vznikat mnohonásobné pozitivní detekce, což znamená, že tvář je zahrnuta v několika překrývajících se oblastech, které jsou od sebe jen málo posunuty a mají podobnou velikost. Stejně tak se mohou, ovšem v menší míře, vyskytnout chybně detekované oblasti. Toho lze využít k efektivnímu zjištění, zda se tvář na daném místě opravdu vyskytuje, pomocí *shlukové analýzy* (*cluster analysis*). Což je metoda dělení jednotek do kategorií (shluků) tak, aby si jednotky náležící do jedné kategorie byly podobnější než objekty z ostatních kategorií [35]. Toto dělení se provádí na základě určitých ukazatelů. Příkladem těchto ukazatelů mohou být korelační koeficienty, metriky a jiné. V našem případě byla jako metrika použita euklidovská vzdálenost:

$$D_{\mathrm{E}}((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \tag{4.26}$$

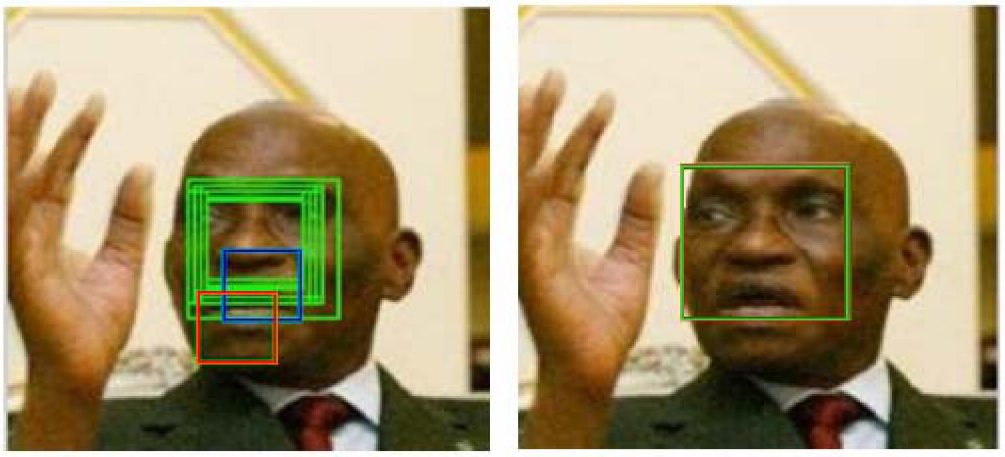
kde (x_1,y_1), (x_2,y_2) představují souřadnice dvou obrazových bodů.

Vzhledem k tomu, že každá pozitivní detekce *Viola-Jones* detektorem představuje vytvoření vektoru souřadnic $[X,Y,W]$, kde X a Y představují obrazové souřadnice středu detekovaného pod-okna a W pak jeho šířku. První vektor souřadnic vytvoří první kategorii shluku Cluster(1), viz obr. 4.25. Pokud nastane další pozitivní detekce, je testována hodnota metriky D_{E} pro všechny kategorie shluků Cluster(i), pro i=1…N, kde N je současný počet kategorií shluků (jedná se vlastně o maximální přípustný rozdíl pozic středu detekovaného pod-okna od pozic středů všech pod-oken, které jsou uloženy v dané kategorii Cluster(i)). Kromě metriky pozic středů je pro každou detekci také testován poměr velikosti pod-okna vzhledem k velikosti pod-oken uložených v každé kategorii Cluster(i). Jestliže jsou obě podmínky splněny, pak je detekované pod-okno zahrnuto do aktuální kategorie i. Pokud však obě podmínky nejsou splněny pro žádnou z N kategorií, pak je vytvořena nová kategorie shluků Cluster(N+1), a počet kategorií N je zvýšen. Po ukončení procesu detekce jsou všechny kategorie shluků s počtem pod-oken M_i nižším než je definovaná hodnota zamítnuty (pravděpodobný výskyt falešné detekce), u ostatních kategorií shluků jsou jednotlivé hodnoty aritmeticky zprůměrovány, takže každá nezamítnutá kategorie shluků obsahuje jedno

pod-okno odpovídající jednomu detekovanému obličeji. Tímto krokem tedy došlo ke snížení dimenze dat, viz. obr. 4.26.



Obr. 4.25. Znázornění struktury kategorií shluků.



Obr. 4.26. Využití shlukové analýzy k redukci datové dimenze. Jednotlivé barvy odpovídají jednotlivým kategoriím shluků (kategorie odpovídající modré a červené barvě jsou z důvodu nedostatečného počtu pod-oken zamítnuty).

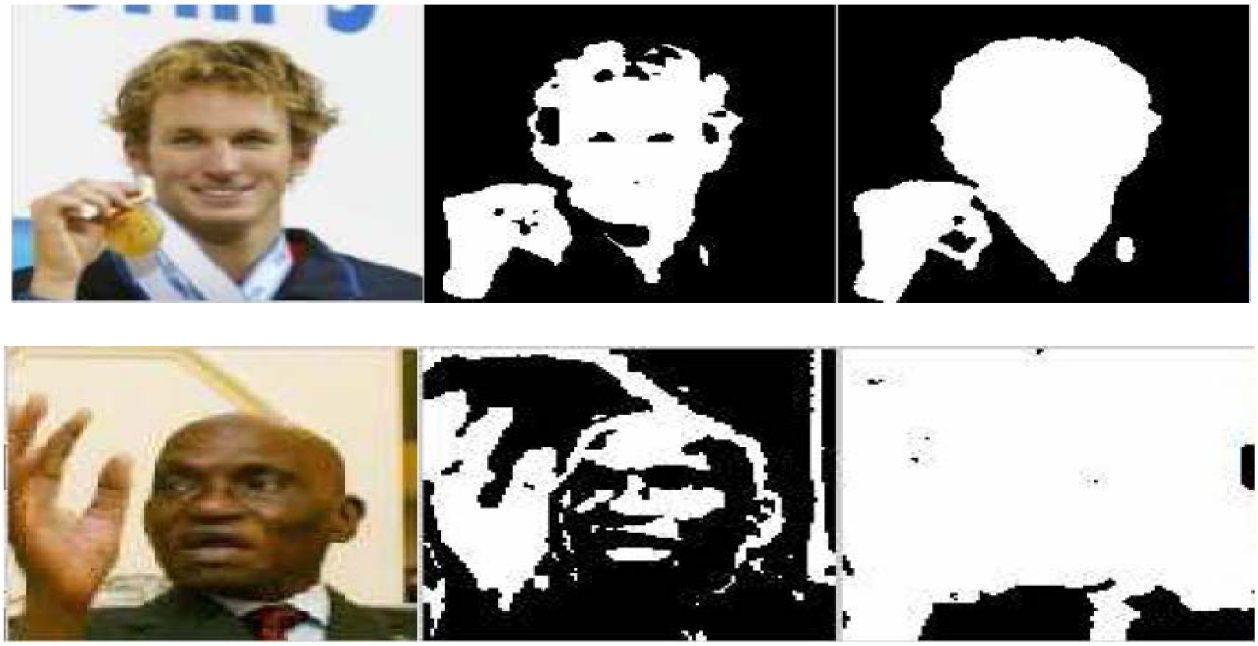
Abychom potlačili falešné pozitivní detekce, které se sice klasifikátoru jeví jako obličej, ale obličejem ve skutečnosti nejsou, využijeme informace nesoucí barevné složky obrazu. Vyjdeme přitom z předpokladů stanovených v podkapitole 4.1. V tomto případě ovšem nevyužíváme barvu kůže pro samotnou detekci obličeje, ale pouze pro ověření již detekovaných oblastí. Jelikož vstupní obraz byl podroben detekci v šedo-tónové oblasti,

odkud jsme získali jen velmi malé množství obličejových kandidátů, bude pravděpodobnost výskytu falešného obličejového kandidáta s barvou odpovídající barvě kůže poměrně nízká. Z tohoto důvodu není potřeba využívat složitých a robustních modelů uvedených v podkapitole 4.1, ale můžeme využít pouze jednoduchého, výpočetně nenáročného modelu. V této práci byly použity dva jednoduché parametrické barevné modely kůže vyjádřené explicitně dle (4.27) [26] a (4.28) [12]:

$$b_{\mathrm{RGB}}(x,y) = \left[R(x,y) > 95\right] \wedge \left[G(x,y) > 40\right] \wedge \left[B(x,y) > 20\right] \wedge \left\{\left[R(x,y) > G(x,y)\right] - 15\right\} \wedge \left[R(x,y) > B(x,y)\right], \tag{4.27}$$

$$b_{\mathrm{YCbCr}}(x,y) = \left[77 < Cb(x,y) > 127\right] \wedge \left[133 < G(x,y) > 173\right], \tag{4.28}$$

kde $b_{\mathrm{RGB}}(x,y)$ udává hodnotu binární masky barevného prostoru *RGB* na pozici *x* a *y*, podobně pak $b_{\mathrm{YCbCr}}(x,y)$ udává hodnotu binární masky na pozici *x* a *y* barevného prostoru *YCbCr*. Výhodou použití těchto dvou barevných modelů kůže je zejména rychlost výpočtu jednotlivých binárních masek a také velmi malá míra barev odpovídacích barvě kůže, jež těmito modely není zastoupena viz. obr. 4.27.



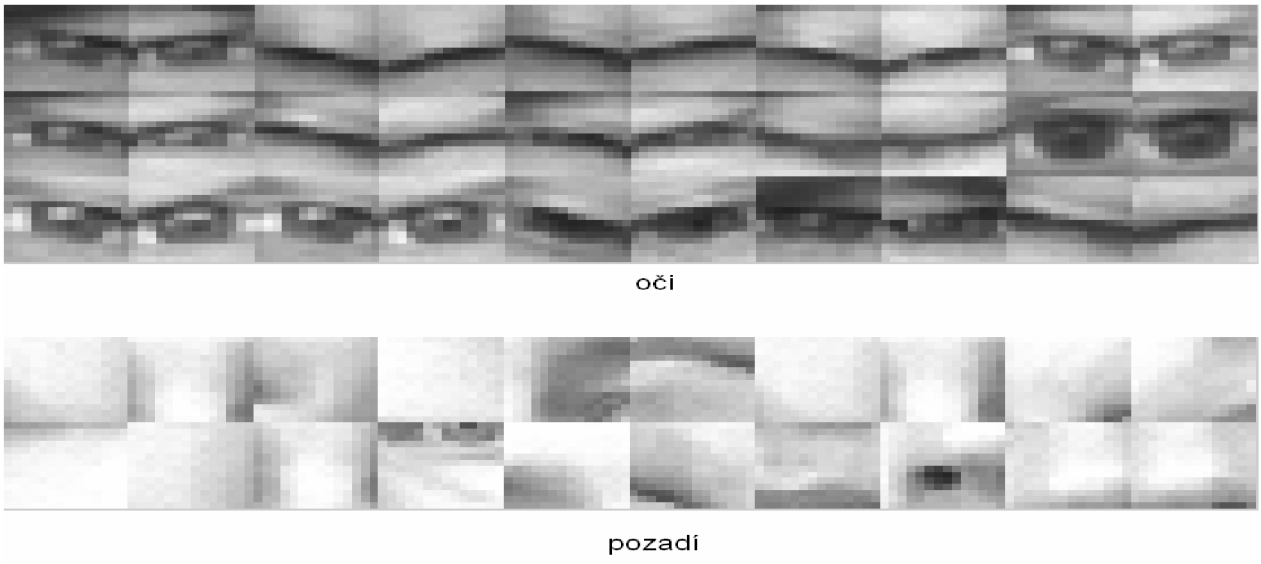
Obr. 4.27. Ukázky detekce barvy kůže (originální obraz, model *RGB*, model *YCbCr*).

Jak je možno vidět na předchozím obrázku, ani jeden z barevných modelů barvy kůže není ideální. V našem případě se však nesnažíme o nalezení obličeje v vstupním obraze, ale pouze o ověření, zda výsledné pod-okno dané kategorie shluků obličej obsahuje či nikoliv. K tomuto účelu je proveden součet všech bodů jednotlivých binárních masek odpovídajících danému pod-oknu váhovaných koeficientem o hodnotě 0,5. Pokud je hodnota tohoto součtu vyšší než $W_i^2.p$, kde W_i je šířka čtvercového pod-okna *i*–té kategorie shluků a $p \in \langle 0,1 \rangle$ je práh míry shody, je část vstupního obrazu odpovídající tomuto pod-oknu prohlášena za obličej.

Detekce očí

V podkapitole 4.1 byl pospán přístup k detekci jednotlivých obličejových částí (očí a úst) na základě barevných map. Pozice těchto obličejových částí byly následně využity pro ověření předpokládaných obličejových kandidátů. K tomuto účelu bylo využití barevným map při procesu lokalizace dostatečné, avšak zatímco u lokalizace rtů nalezená pozice poměrně přesně odpovídala skutečnosti, při lokalizaci očí se nalezené pozice od skutečnosti často velmi lišily (způsobeno vlivem obočí, částí vlasů a jiných tmavých a světlých částí obličeje). V této části tedy bude experimentální možnost využití detektoru *Viola-Jones* pro detekci očí v jichž nalezených obličejích [70].

Jak bylo zmíněno, využijeme opět výše popsaného detektoru objektů *Viola-Jones*, kde jako pozitivní vzory trénovací množiny použijeme různé obrázky očí. Jelikož však nyní budeme uvažovat detekci v rámci pod-okna obsahujícího pouze obličej, budou negativní vzory trénovací množiny různé obrázky jednotlivých částí obličejů (části neobsahující oči). Bohužel v současnosti není dostupná jakákoliv databáze splňující tyto požadavky, a proto jsme museli tuto databázi nejprve vytvořit. Databáze se skládá z 200 pozitivních vzorů (100 × levé oko a 100 × pravé oko) a z 2000 vzorů negativních. Jednotlivé vzory mají rozměr 14×14 pixelů a byly extrahovány z obličejové databáze *BioID* [29]. Na obr. 4.28 jsou některé vzory zobrazeny.



Obr. 4.28. Ukázka obrázků z trénování množiny pro detekci očí.

Tato trénovaní množina byla použita jako vstup trénovací fáze kaskádového detektoru objektů *Viola-Jones*, který byl natrénován se 100% přesností na této množině vzorů s celkovým počtem 81 slabých klasifikátorů (příznaků) rozdělených do 13 stupňů kaskádního zapojení.

Při testování spolehlivosti tohoto detektoru byla pro redukci mnohanásobné detekce opět použita shluková analýza způsobem popsaným výše. K potlačení falešné pozitivní detekce

byly opět využity informace nesoucí barevné složky obrazu a byla vytvořena tzv. *barevná oční mapa* popsaná v podkapitole 4.1 (4.12) a převedená na binární obraz (4.13) obr. 4.29.

Obr. 4.29. Ukázka binární mapy pro detekci očí.

Při vlastní detekci očí jsou tedy prohledávány pouze oblasti, u kterých hodnota binární masky odpovídající pozici středu posuvného pod-okna je rovna 1. Toto má za následek nejen snížení možnosti výskytu falešné pozitivní detekce, ale také zvýšení výpočetní rychlosti, protože se prohledává vždy jen přibližně 10% vlastního obrazu obličeje. V případě nalezení více oblastí odpovídajícím očí v jednom obličeji jsou vybrány pouze dvě ty oblasti, které nejlépe odpovídají možné pozici očí (nachází se v horní části obličeje, leží přibližně ve stejné vodorovné linii a jejich vzájemná vzdálenost je větší než ¼ šířky samotného obličeje).

Testování a výsledky

Výše popsaný postup pro detekci tváře a očí byl implementován v programovacím jazyce C/C++ a vytvořena aplikace umožňující jednoduché využití navrženého detektoru včetně přehledného nastavení jednotlivých parametrů zmiňovaných v samotném textu (faktor zvětšení posuvného pod-okna, minimální počet pod-oken v jedné kategorii shluků a jiné). Vlastní zpracování obrazu s rozlišením 640 × 480 pixelů trvá na počítači P4 2,8 GHz necelých 100 ms.

Popsaný algoritmus detektoru byl testován opět na obrazové databázi *Georgia Tech Face Database* [44]. Vlastní testování implementovaného detektoru bylo z hlediska vyhodnocení rozděleno na dvě části: na detekci tváře a na detekci očí.

Při detekci tváří byla vlastní detekce jednotlivého snímku považována za úspěšnou, jestliže byl v daném snímku nalezen obličej obsahující obě oči, nos a minimálně horní část úst. Jak je vidět z tab. 4.4, tak zatímco u detektoru obličejů založeném na detekci barvy kůže (podkapitola 4.1) byla míra specificity vyšší než míra senzitivity, u detektoru obličejů založeném na objektovém detektoru *Viola-Jones* je tomu naopak (tj. chybně detekovaných obličejů bylo méně než chybně detekovaných oblastí pozadí). Projevil se zde výrazný vliv barvy pozadí databáze *GTFD*, která z větší části barevně odpovídala barvě kůže, takže potlačení falešných pozitivních detekcí pomocí barevné informace se zde příliš neuplatnilo. Hlavní podíl chybných detekcí obličejů byl zapříčiněn určitými úhly natočení hlavy na které nebyl objektový detektor *Viola-Jones* trénován. Ačkoliv celková přesnost detekce byla ve srovnání s celkovou přesností detektoru obličejů založeném na barvě kůže vyšší, je třeba si uvědomit, že zvolená testovací databáze byla zaměřena primárně na odolnost proti barvě pozadí odpovídajícího barvě kůže. V reálných situacích, kdy se sice toto barevné pozadí na snímcích vyskytuje jen občas, by oba detektory dosahovaly přibližně stejné přesnosti. Avšak zatímco detekce obličeje trvala detektoru založeném na barvě kůže zmiňovaných 1,2 s, tak průměrná doba detekce pomocí detektoru *Vila-Jones* na stejném testovacím počítači byla necelých 100 ms.

Tab. 4.4. Výsledky testování detektoru obličejů založeném na detektoru *Viola-Jones* na obrazové databázi GTFD.

TP = 723	*TPR* = 96,4 %
TN = 711	*TNR* = 94,8 %
FP = 39	*ACC* = 95,6 %
FN = 27	*PPV* = 94,9 %

Při detekci pozice očí v oblastech vstupního obrazu odpovídajícím nalezeným tvářím byla detekce považována za úspěšnou, pokud se v detekované oblasti nacházela zornička (v případě zavřených očích by zde byl výskyt zorničky předpokládán). V tomto případě byl počet obrazů očí dvojnásobný oproti počtu obrazů pozadí $P = 1446$ a $N = 723$. Výsledky jsou uvedeny v tab. 4.5. Míra senzitivity je zde výrazně vyšší než míra specificity zejména z toho důvodu, že obraz zavřeného oka, kdy se na detekci podílí zejména vliv očních řas, je velmi

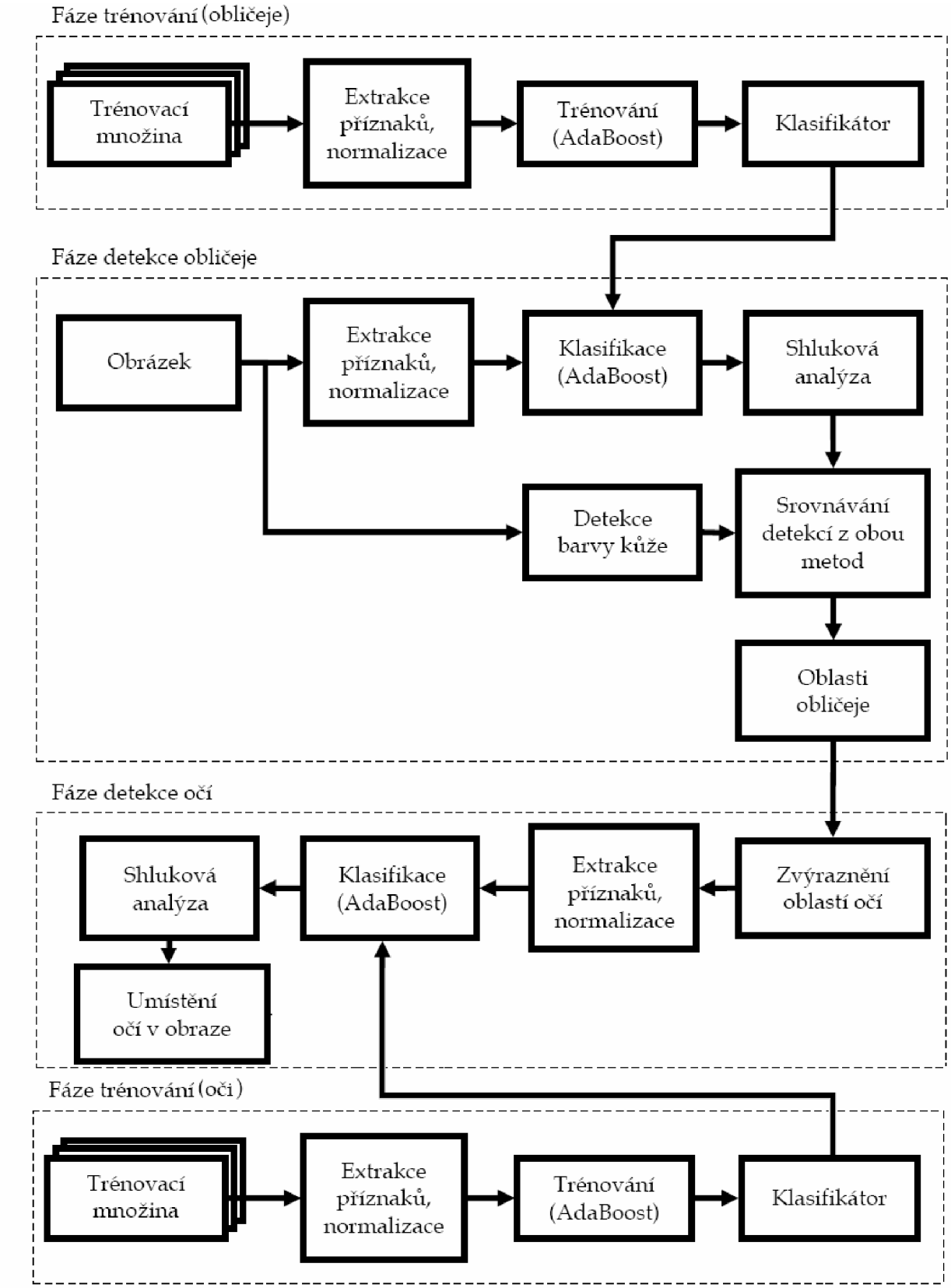
podobný obrazu obočí, v tomto případě docházelo poměrně často k záměně. Taktéž se občas projevil vliv nasazených brýlí, zejména vliv odrazu světla od skel a vliv částí obrouček, které při nižším obrazovém rozlišení mohou rovněž připomínat zavřené oko. Celková úspěšnost detekce očí je nižší než celková úspěšnost detekce obličejů, a to především z důvodu velké variability obrazu očí, který nemá na rozdíl od obličeje žádnou pevně danou strukturu (deformace celého oka při pohybu víček a pohyb zornice).

Tab. 4.5. Výsledky testování detektoru očí založeném na detektoru *Viola-Jones* na obrazové databázi GTFD.

TP = 1383	*TPR* = 95,6 %
TN = 526	*TNR* = 72,7 %
FP = 197	*ACC* = 88 %
FN = 63	*PPV* = 87,5 %

Níže jsou uvedeny obrázky jednotlivých detekcí na testovací databázi i na reálných obrázcích (obr. 4.30), na konec pak následuje stručné blokové schéma celého procesu lokalizace obličeje a obou očí (obr. 4.31).

Obr. 4.30. Ukázky výsledků navrženého detektoru (nahoře – obrázky z testovací databáze, dole – reálné obrázky).



Obr. 4.31. Blokové schéma procesu lokalizace obličejů a očí.

5 Extrakce příznaků

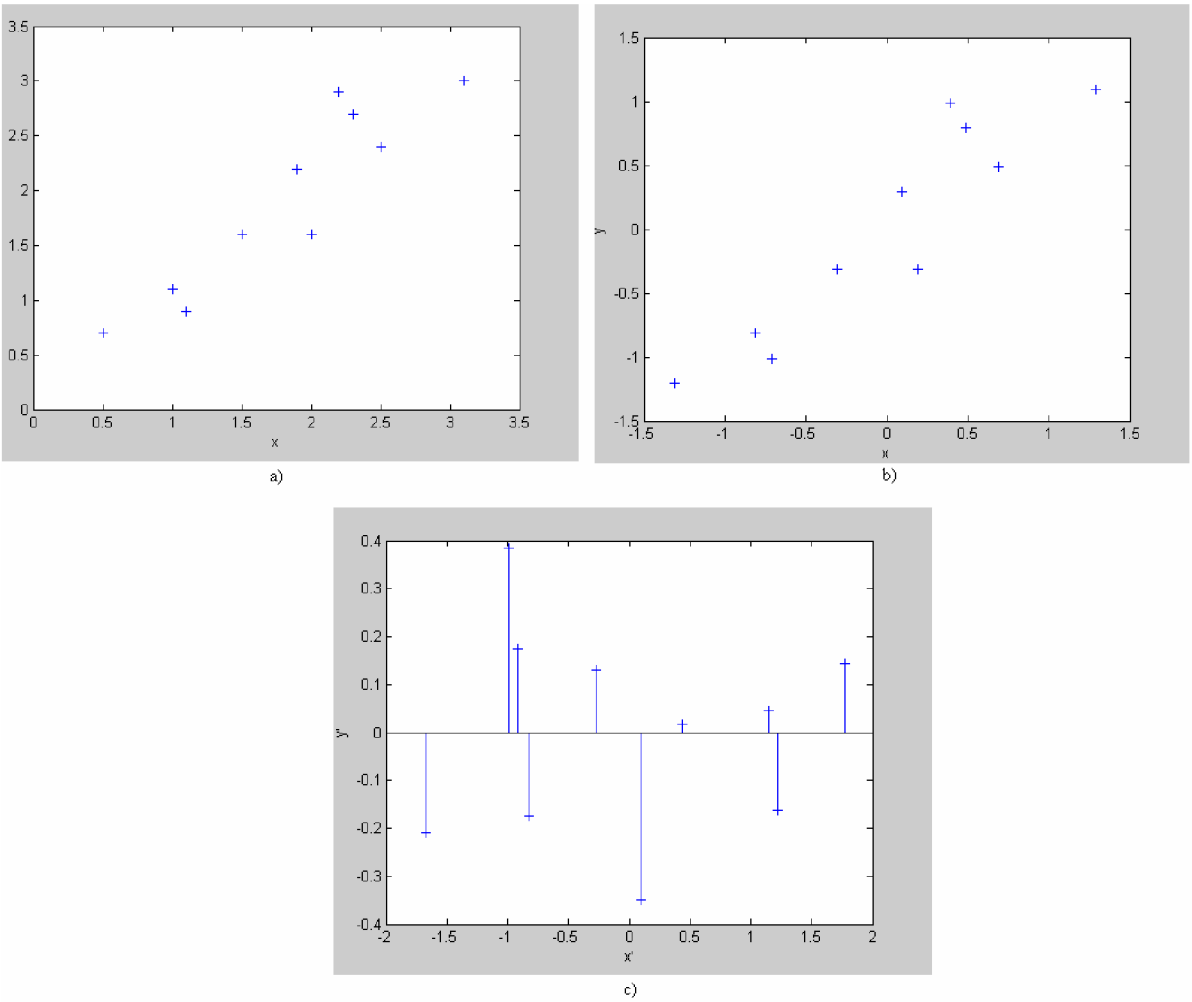
V předcházející kapitole byla rozebrána realizace automatického systému pro lokalizaci obličeje a jeho jednotlivých částí. Můžeme tedy předpokládat, že pro další zpracování známe pozici nejen samotného obličeje ale i pozici úst a očí. Tato kapitola bude zaměřena na extrakci optimálních univerzálních příznaků jedinečně popisujících určitý výraz v obličeji. Jelikož o výrazu v obličeji nese barevná složka pouze minimální informaci, bude z důvodu nižší výpočetní náročnosti dále pracováno pouze s jasovou složkou Y z barevného prostoru *YCbCr*. Dále pak bude jako vstupní obraz uvažován normalizovaný obraz obličeje s rozlišením 125x160 pixelů, který je získán změnou měřítka původního obrazu obličeje (tj. nadvzorkováním či podvzorkováním). Normalizace z hlediska osvětlení snímané scény a úhlu natočení obličeje nebude v této části práce uvažována. Pro vlastní extrakci příznaků budou použity dvě statistické metody a jedna metoda založená na prostorově orientované filtraci obrazu.

5.1 Analýza hlavních komponent

Analýza hlavních komponent *PCA* je klasická metoda lineární projekce pro zobrazení dat z prostoru s vyšší dimenzionalitou do prostoru s dimenzionalitou nižší [30]. Podstata analýzy hlavních komponent spočívá tedy ve snížení dimenzionality vstupní sady dat, která obsahuje velké množství navzájem souvisejících proměnných, takovým způsobem, aby bylo zachováno co největší množství informace sloužící k separaci jednotlivých kategorií navzájem nesouvisejících dat. Této redukce je dosaženo transformováním dané sady dat do nové sady proměnných tzv. *hlavních komponent*, které by měly být navzájem nekorelované a jsou řazeny tak, že prvních několik komponent obsahuje největší množství informace sloužící k separaci vstupních dat. Tato analýza je založena na statistických momentech druhého řádu, tj. na rozptylech vstupní sady dat. Tuto analýzu můžeme také nazvat jako dekorelaci vstupních dat.

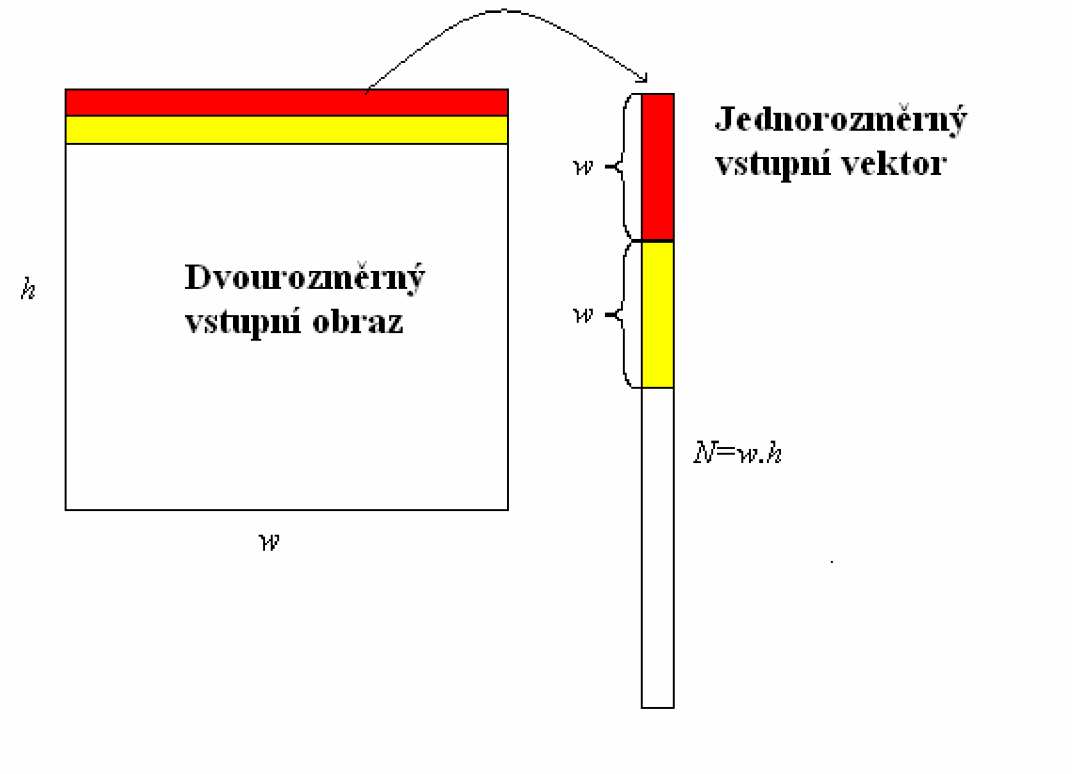
Tato analýza bývá často nazývána jako Karhunen-Loèveho transformace nebo Hotellingova transformace, která je matematicky definována jako ortogonální lineární transformace, jenž transformuje data do nového souřadnicového systému takovým způsobem, že největší variance vstupních dat bude promítnuta na první souřadnici tohoto nového souřadnicového systému (na první hlavní komponentu), druhá největší variance vstupních dat pak bude ležet na druhé souřadnici (druhé hlavní komponentě) a tak dále [30]. Analýza hlavních komponent je teoreticky optimální transformace vstupních dat při použití metody nejmenších čtverců, tj. je optimální z hlediska střední kvadratické chyby. Obecně lze tedy říci, že pro množinu náhodný vektor $X = \{X_1, X_2,\ldots X_n\}$ hledáme takový souřadnicový systém, ve kterém náhodný vektor X můžeme vyjádřit jako $Z = \{Z_1, Z_2,\ldots Z_n\}$, kde Z_1 je vyjádřena lineární kombinaci všech původních X_i , a to tak, aby rozptyl Z_1 byl ze všech existujících možností maximální. Dále pak Z_2 vznikne opět lineární kombinaci všech původních X_i , a opět tak, aby rozptyl Z_2 byl maximální, avšak zároveň nesmí být Z_1 a Z_2 navzájem korelovány.

Tímto způsobem se pokračuje až po Z_n. Grafické znázornění procesu analýzy hlavních komponent na dvourozměrné množině vstupních dat je zobrazeno na obr. 5.1.



Obr. 5.1. Grafické znázornění analýzy hlavních komponent na dvourozměrné množině vstupních dat: a)vstupní data, b) centralizovaná vstupní data podle aritmetického průměru, c) data po provedení analýzy hlavních komponent v novém souřadnicovém systému.

Metody založené na použití analýzy hlavních komponent jsou velmi často využívány při rozpoznávání objektů [36]. Poprvé byla tato analýza využita pro rozpoznávání jednotlivých obličejů v roce 1991 [62] a s různými modifikacemi se používá dodnes. Postup je následující: mějme sadu M obrázků o rozměrech $w \times h$, kde w horizontální a h udává vertikální rozměr obrázku v pixelech. Transformujme tuto sadu obrázků do M jednorozměrných normalizovaných vektorů $\Gamma_1, \Gamma_2, \ldots \Gamma_M$ ($\| \Gamma_i \| = 1$) délky N ($N = w \times h$) viz obr. 5.2, přičemž normalizace nám zajišťuje invariabilitu transformace vůči světelným podmínkám.



Obr. 5.2. Transformace dvourozměrného vstupního obrazu na jednorozměrný vstupní vektor.

Nyní můžeme vypočítat průměrný vektor Ψ dle rovnice (5.1).

$$\Psi = \frac{1}{M}\sum_{i}^{M} \Gamma_i . \tag{5.1}$$

Tento vektor můžeme považovat za průměrný vektor tváře, jehož odečtením od všech normalizovaných vektorů Γ_i dle rovnice (5.2) zajistíme, že při výpočtu hlavních komponent bude určen nový prostor, ve kterém dosáhne variance jednotlivých vektorů maximální hodnoty ve smyslu vzájemné korelace.

$$\Phi_i = \Gamma_i - \Psi, \quad i = 1..M. \tag{5.2}$$

Takto získaná množina vektorů Φ_i je následně podrobena analýze hlavních komponent, která má nalézt N ortogonálních vektorů u_n, kde n=1…N. Přičemž vektory u_n jsou vlastními vektory kovarianční matice C a jsou zároveň také požadovanými hlavními komponenty. Kovariační matice C je určena dle vztahu (5.3):

$$C = \frac{1}{M}\sum_{i}^{m} \Phi_i \cdot \Phi_i^{\mathrm{T}} = \frac{1}{M} \cdot A \cdot A^{\mathrm{T}}, \tag{5.3}$$

kde A je matice sloupcových vektorů Φ_i o rozměrech $N \times M$. Odtud vyplývá, že celková velikost kovariační matice je $N \times N$. Výpočet vlastních vektorů kovariační matice C je definován vztahem:

$$U^{-1} \cdot C \cdot U = D, \tag{5.4}$$

kde U *je* N-rozměrná matice sloupcových vlastních vektorů kovarinční matice C a D je N-rozměrná diagonální matice vlastních čísel kovariační matice C:

$$\begin{aligned} D(i,j) &= \lambda_n \quad \text{pro } i = j = n, n = 1..N \\ D(i,j) &= 0 \quad \text{pro } i \neq j. \end{aligned} \tag{5.5}$$

Vlastní čísla λ_n a vlastní vektory u_n jsou na sebe navzájem vázány, takže n-té vlastní číslo odpovídá n-tému vlastnímu vektoru.

Nový vektor příznaků ω_i je získán na základě řešení rovnice (5.6) za použití k vlastních vektorů u_n odpovídajících k nejvyšším vlastním číslům λ_n. Dojde tedy ke zobrazení původního N-rozměrného obrazového prostoru (N=$w \times h$) na k-rozměrný prostor příznaků, kde $k << N$:

$$\omega_{i,j} = u_i^{\mathrm{T}} (\Gamma_j - \Psi), \quad i = 1...k, j = 1..M. \tag{5.6}$$

Lze tedy říci, že dojde k redukci původního N-rozměrného vektoru Γ na k-rozměrný prostor příznaků, kde k-rozměrný vektor tvaru vah Ω_j=[$\omega_{1,j}, \omega_{2,j}, \ldots \omega_{k,j}$] vyjadřuje příspěvek každého vlastního vektoru pro reprezentaci j-tého obrázku. Následně pak lze srovnáním jednotlivých vektorů tvaru vah pomocí zvoleného klasifikátoru určit nejpravděpodobnější vzor.

Jak bylo dle literatury [30] prokázáno, tak určení vlastních vektorů a vlastních čísel kovariační matice C o rozměrech $N \times N$ je pro vyšší hodnoty N (tj. pro obrazové vzory s vyšším rozlišením) výpočetně extrémně náročný proces. Budeme-li však uvažovat, že počet vzorů bude výrazně nižní než jejich velikost (M<<N), pak lze využít techniku výpočtu vlastních čísel a vlastních vektorů kovariační matice o rozměrech $N \times N$ pomocí lineární kombinace vektorů získaných řešením výpočtu vlastních čísel matice o rozměrech $M \times M$. Postup je následující: vypočteme matici C' dle:

$$C' = \frac{1}{N} \sum_{i}^{m} \Phi_i^{\mathrm{T}} \cdot \Phi_i = \frac{1}{N} \cdot A^{\mathrm{T}} \cdot A. \tag{5.7}$$

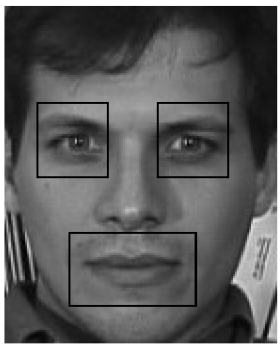
Nyní tedy máme $M \times M$ rozměrnou matici C', kterou podrobíme dekompozici na vlastní čísla a vlastní vektory jak bylo popsáno výše. Dostáváme tedy matici U' vlastních vektorů u_m' a diagonální matici D' vlastních čísel λ_m', kde m=1…M. Pro získání vlastních vektorů a vlastních čísel matice kovariační C využijeme vztahu (5.8):

$$\begin{aligned} U &= A \cdot U', \\ D &= A \cdot D'. \end{aligned} \tag{5.8}$$

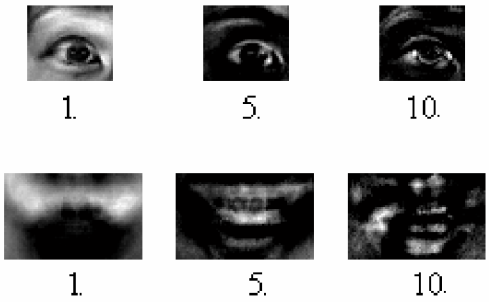
Odtud tedy získáme M vlastních vektorů u_m, které odpovídají M vlastním číslům λ_m kovariační matice C.

Využití analýzy hlavních komponent pro rozpoznání výrazu v obličeji vychází z práce publikované v literatuře [46]. Pro vlastní extrakci příznaků reprezentujících jedinečnost

daného výrazu v obličeji nebude analýzou hlavních komponent podroben celý obličej, ale pouze jeho tři vybrané zájmové oblasti, které se na utváření jednotlivých výrazů podílí největší měrou. První dvě oblasti se nacházejí kolem středu obou očí s rozlišením 35×40 pixelů, oblast třetí je potom umístěna kolem středu úst s rozlišením 40×65 pixelů. Umístění oblastí zájmu je ukázáno na obr. 5.3. a na obr. 5.4 je zobrazena ukázka několika vlastních vektorů oblastí pravého oka a úst.



Obr. 5.3. Umístění oblastí zájmu v analyzovaném obličeji.



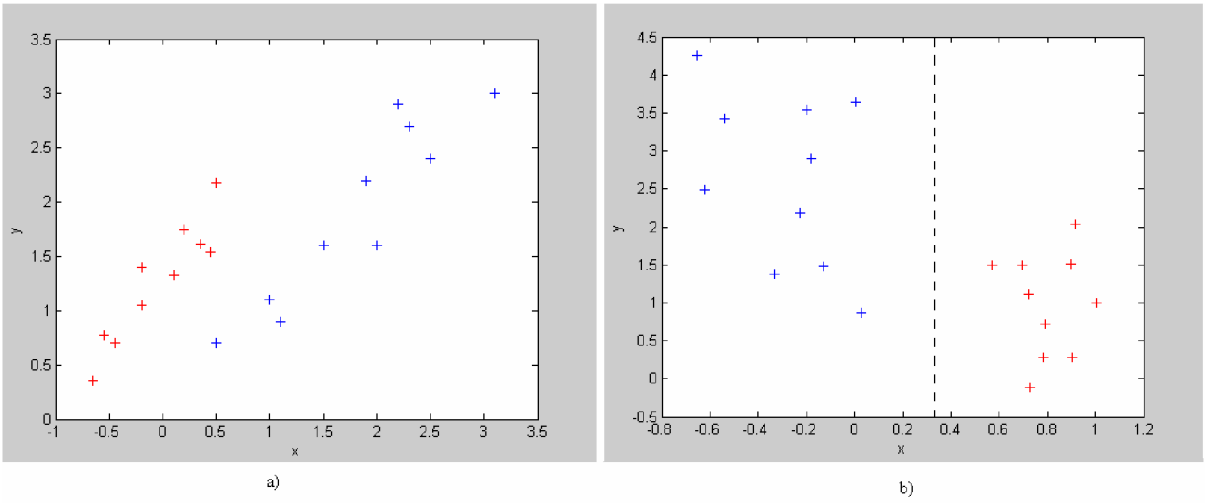
Obr. 5.4. Vlastní vektory odpovídající prvnímu, pátému a desátému nejvyššímu vlastnímu číslu oblasti pravého oka a úst.

5.2 Lineární diskriminační analýza

V předchozí podkapitole bylo uvedeno použití statistické metody analýzy hlavních komponent pro extrakci příznaků sloužících ke klasifikaci daného výrazu ve tváři. Nevýhodou této metody je ovšem fakt, že dochází k projekci vstupních obrazů do podprostorů, které se vyznačují maximální vzájemnou variancí napříč všemi vstupními obrazy, dokonce i těmi které mohou reprezentovat stejný výraz v obličeji. Tomuto omezení lze předejít využitím další statistické metody: lineární diskriminační analýzy *LDA*. Tato metoda se snaží nalézt novou sadu příznakových vektorů, které lze získat lineární kombinací příznaků s maximálním variancí napříč různými třídami a minimální variancí uvnitř jednotlivých tříd [56]. Tato metoda bývá také velmi často používána při úloze rozpoznávání obličejů a bývá také nazývána jako Fischerova lineární diskriminační analýza *FLDA*. Obecně lze tedy říci, že hledáme nový souřadnicový prostor definovaný vektory w v němž jsou jednotlivé třídy nejlépe separovatelné [56]:

$$Y_i = w^{\mathrm{T}} \cdot X_i, \tag{5.9}$$

kde X_i je vektor hodnot i-tého vzoru v původním souřadnicovém systému a Y_i je vektor hodnot i-tého vzoru v novém souřadnicovém systému. Grafické znázornění procesu lineární diskriminační analýzy na dvourozměrné množině vstupních dat se dvěmi třídami je zobrazeno na obr. 5.5.



Obr. 5.5. Grafické znázornění lineární diskriminační analýzy na dvourozměrné množině vstupních dat: a) původní vstupní data, b) transformovaná data.

Předpokládejme, že máme M vstupních obrazů o rozměrech $w \times h$, kde w horizontální a h udává vertikální rozměr obrázku v pixelech, a které můžeme rozdělit do L různých tříd podle vlastností jednotlivých obrazů (např. podle výrazu v obličeji). Stejně jako v předchozí podkapitole transformujme tuto sadu obrázků do N ($N=w \times h$) jednorozměrných normalizovaných vektorů $\Gamma_1, \Gamma_2, \ldots \Gamma_N$ ($\| \Gamma_i \| = 1$), přičemž normalizace nám opět zajišťuje

invariabilitu transformace vůči světelným podmínkám. Nyní můžeme vypočítat průměrné vektory jednotlivých tříd Ψ_l a celkový průměrný vektor Ψ dle rovnice (5.10) a (5.11):

$$\Psi_l = \frac{1}{J_l} \cdot \sum_{j}^{J_l} \Gamma_{l,j}, \quad l = 1 \ldots L, \tag{5.10}$$

$$\Psi = \sum_{l}^{L} p_l \cdot \Psi_l, \tag{5.11}$$

kde J_l udává počet obrazů v l-té třídě, $\Gamma_{l,j}$ udává j-tý vektor patřící do třídy l a p_l udává pravděpodobnost výskytu třídy l (v našem případě, kdy předpokládáme rovnoměrné rozložení obrázků do všech tříd platí pro všechna $p_l = 1/L$ a zároveň $J_l=J$ pro všechna l=1..L). Nyní můžeme definovat matici rozptylu uvnitř jednotlivých tříd S_w a matici rozptylu napříč všemi třídami S_b dle rovnic (5.12) a (5.13):

$$S_b = \frac{1}{L} \cdot \sum_{l}^{L} (\Psi_l - \Psi) \cdot (\Psi_l - \Psi)^{\mathrm{T}}, \tag{5.12}$$

$$S_w = \frac{1}{L \cdot J} \cdot \sum_{l}^{L} \sum_{j}^{J} (\Gamma_{l,j} - \Psi_l) \cdot (\Gamma_{l,j} - \Psi_l)^{\mathrm{T}}. \tag{5.13}$$

Vlastní analýza spočívá v maximalizaci variance rozptylové matice napříč všemi třídami a minimalizaci variance matice rozptylu uvnitř jednotlivých tříd:

$$J(w) = \frac{w^{\mathrm{T}} \cdot S_b \cdot w}{w^{\mathrm{T}} \cdot S_w \cdot w}. \tag{5.14}$$

Toho lze dosáhnout maximalizací poměru $det[S_b]/det[S_w]$. Jak bylo prokázáno Fischerem v roce 1938, tak tento poměr může být maximalizován za podmínky, že matice S_w není singulární. Matice S_w pak nebude singulární pokud budou sloupcové vektory projekční matice odpovídat vlastním vektorům součinu $S_w^{-1} S_b$ [20]:

$$S_b \cdot u = \lambda \cdot S_w \cdot u \Rightarrow S_w^{-1} \cdot S_b \cdot u = \lambda \cdot u, \tag{5.15}$$

kde u jsou vlastní vektory odpovídající vlastním číslům λ součinu matic $S_w^{-1} S_b$. Potom tedy hledané vektory transformace nalezeným odpovídají vlastním vektorům (w=u).

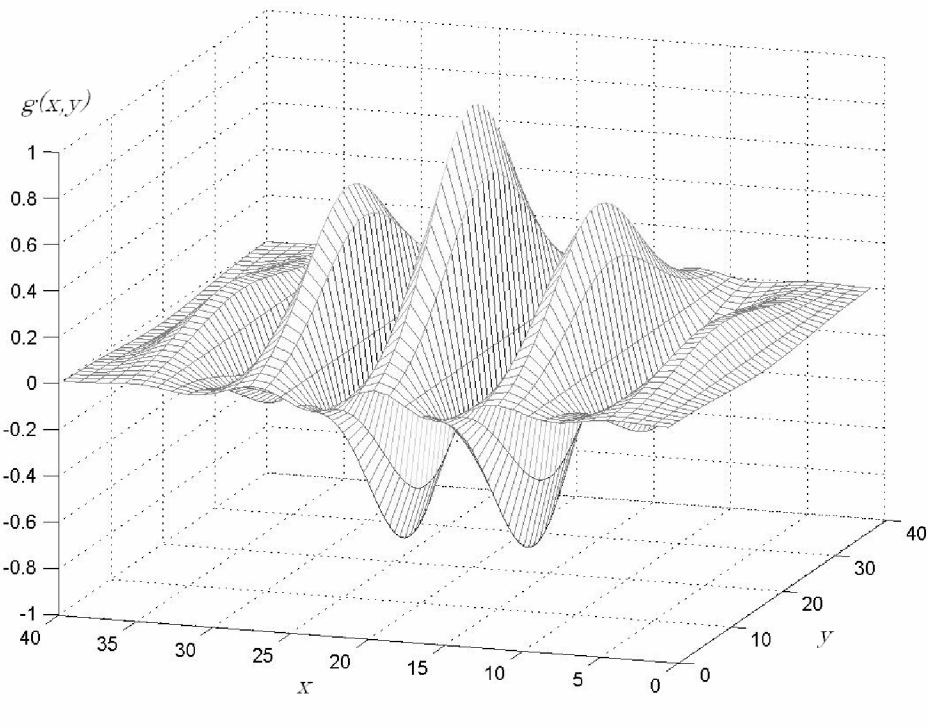
V našem případě byly pro rozpoznávání výrazů ve tváři použity stejné oblasti zájmu jako u analýzy hlavních komponent. Avšak aby matice S_w nebyla singulární, byla nejprve redukována dimenze vstupních dat z $N \times N \rightarrow M \times M$ ($N >> M$) pomocí aplikování analýzy hlavních komponent na vstupní vektory Γ_i. Pomocí lineární diskriminační analýzy poté jsme schopni vybrat k lineárních diskriminantů ($k \leq M$) popisujících jednotlivé příznaky.

5.3 Sada Gaborových filtrů

Obě doposud zmíněné matematické statistické metody *PCA* a *LDA* statisticky vyhodnocují intenzitu jasové složky Y jednotlivých pixelů vstupního obrazu, avšak již neuvažují jednotlivé vztahy mezi sousedícími pixely. To má za následek možnost chybných výstupů zejména v případě, pokud rozložení intenzity jasu v obraze není rovnoměrné (např. obraz je nasvícen z jedné strany, kdy jedna polovina obrazu se jeví světlejší a druhá polovina tmavší). Z tohoto důvodů se zdá vhodnější nepracovat přímo s intenzitou jasu jednotlivých pixelů, ale raději s její změnou v oblasti okolních pixelů. K tomuto je možno využít dvourozměrnou filtraci obrazu pomocí vhodně zvolených filtrů. V našem případě jsme vyšli metody navržené v literatuře [5], která pracuje se sadou Gaborových filtrů. Gaborův filtr je lineární filtr, jehož impulsní charakteristika je definována jako součin harmonické funkce a gaussova okna (5.16) viz obr. 5.6.

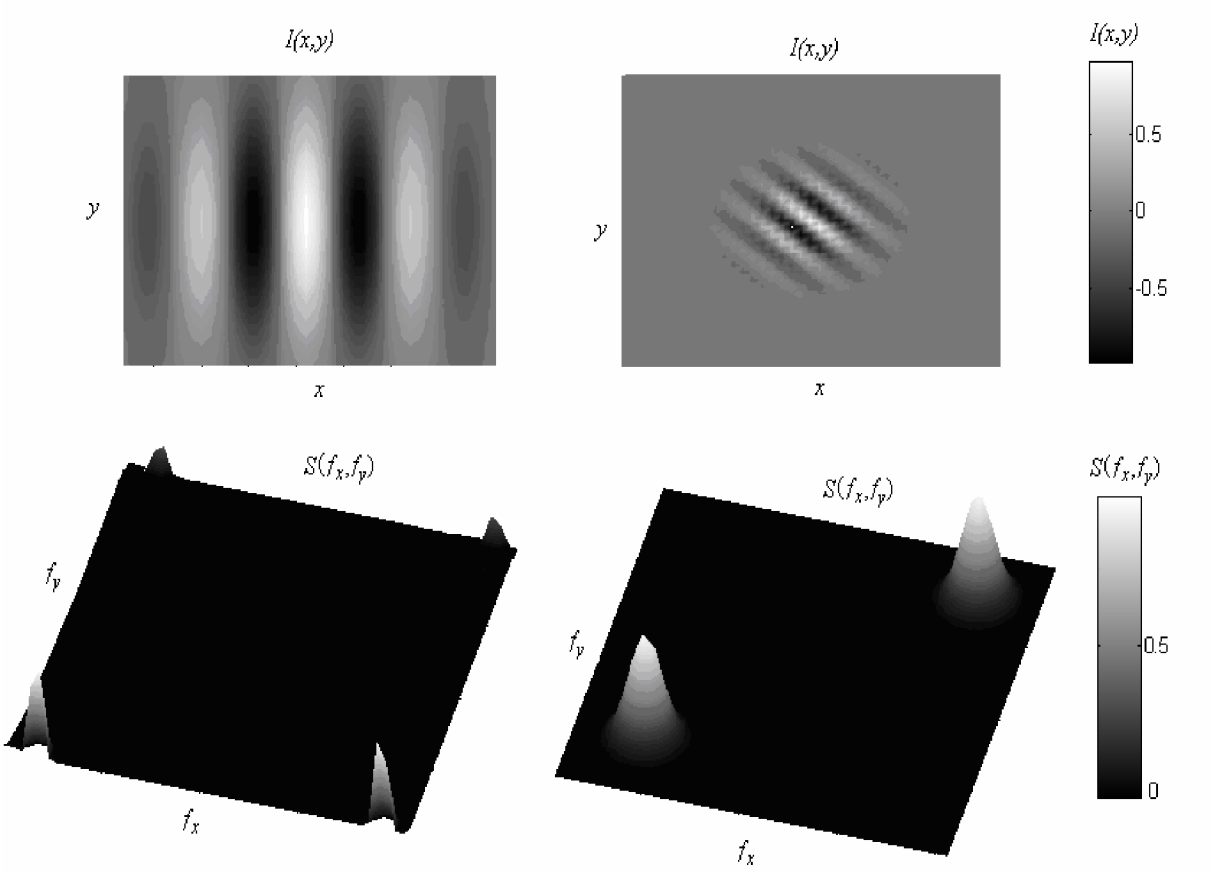
$$\begin{aligned} g(x, y, \lambda, \theta, \psi, \sigma) &= \mathrm{e}^{-\frac{x'^2+y'^2}{2\sigma^2}} \cdot \cos(2\pi \frac{x'}{\lambda} + \psi), \\ x' &= x\cos(\theta) + y\sin(\theta), \\ y' &= -x\sin(\theta) + y\cos(\theta), \end{aligned} \tag{5.16}$$

kde λ udává délku filtru (řád filtru), θ jeho prostorovou orientaci, ψ fázový posun a σ šířku gaussova okna.



Obr. 5.6. Impulsní charakteristika Gaborova filtru.

Obecně lze přirovnat použití gaborových filtrů k waveletové transformaci s mateřským waveletem odpovídajícím Gaborově funkci. Změnou délky Gaborova filtru a jeho prostorové orientace jsme schopni zachovat určitá prostorově-frekvenční pásma zatímco ostatní části jsou potlačeny, podobně jako při waveletové transformaci viz obr. 5.7. Změnou délky filtru λ dochází ke změně velikosti propustného pásma a změnou prostorové orientace θ dochází k frekvenčně-prostorovému posunu propustného pásma. Z tohoto důvodu bývají občas Gaborovy filtry označovány jako Gaborovy wavelety.

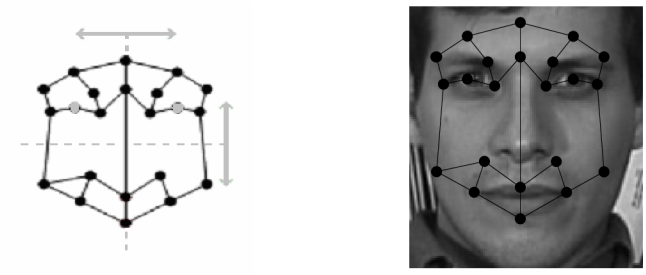


Obr. 5.7. Impulsní charakteristika (nahoře) a frekvenční charakteristika (dole) Gaborova filtru pro různé hodnoty λ a θ.

Vstupní obraz je tedy filtrován sadou 40 Gaborvých filtrů, realizovaných kombinací 8 různých prostorových orientací θ a 5 různých délek filtru λ a s nimi odpovídajících velikostí gaussova okna σ, dle (5.17):

$$\begin{aligned} \theta &= \left\{0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8}, \frac{7\pi}{8}\right\}, \\ \lambda &= \{4, 4\sqrt{2}, 8, 8\sqrt{2}, 16\}, \\ \sigma &= \lambda, \\ \psi &= 0. \end{aligned} \tag{5.17}$$

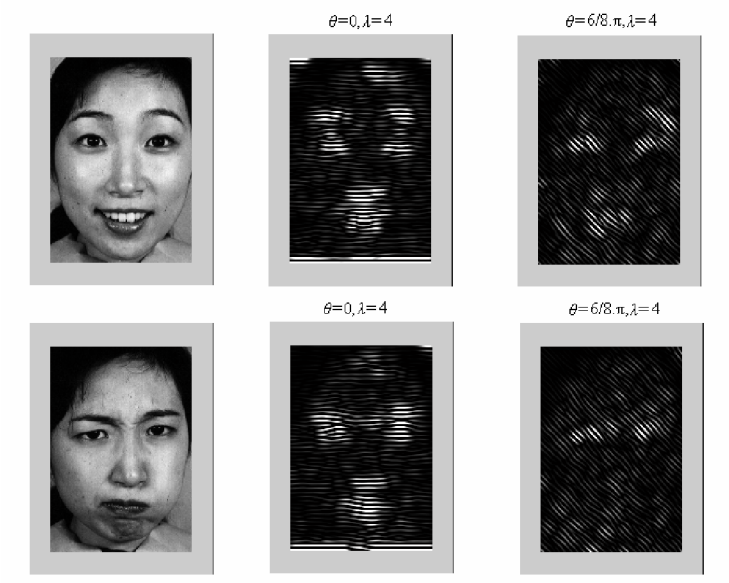
Celkový počet koeficientů je tedy roven $w \times h \times 40$, kde w a h jsou vertikální a horizontální rozměry vstupního obrazu. Uvážíme-li vysokou redundatnost a velkou vzájemnou korelaci sousedících koeficientů, můžeme provést redukci celkového počtu koeficientů výběrem vždy jednoho reprezentativního koeficientu ze skupiny sousedících koeficientů. Dále pak uvážíme-li, že ne všechny oblasti vstupního obrazu se podílejí na utváření jednotlivých výrazů v obličeji stejnou měrou, můžeme se zaměřit pouze na oblasti, které se při jednotlivých výrazech ve tváři mění nejvíce (tj. zejména oblasti kolem očí a úst). Na základě těchto úvah byla navržena 22-bodová obličejová maska viz obr. 5.8, která postihuje nejvýznamnější oblasti z hlediska změny při jednotlivých výrazech ve tváři. Maska je na obličej přikládána na základě znalosti pozice očí a úst, a aby vyhovovala různým antropologickým typům obličeje, je na základě poměru jednotlivých vzdáleností vertikálně a horizontálně přizpůsobována.



a) b)

Obr. 5.8. a) 22-ti bodová obličejová maska a b) příklad jejího použití.

Každý bod této obličejové masky reprezentuje sadu 40 koeficientů příslušející k danému pixelu. Celkem tedy nyní máme $22 \times 40 = 880$ koeficientů. Tento počet může být dále redukován např. použitím matematické statistické analýzy *PCA* či *LDA*.



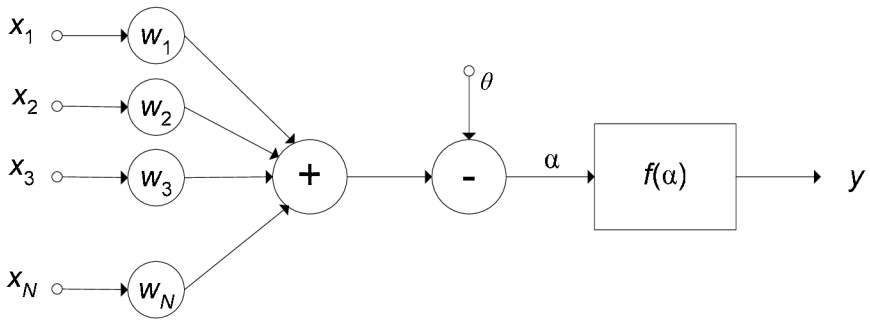
Obr. 5.9. Ukázka odezev několika Gaborových filtrů na vstupní obrazy představující vyjádření různých emocionálních stavů.

6 Klasifikace emocionálních výrazů

V předchozích kapitole byly uvedeny tři metody pro extrakci příznaků vhodných pro rozpoznávání emocionálních výrazů v obličeji. Pro vlastní rozpoznání daného emocionálního výrazu ze zvolených příznaků je vhodné použít matematický klasifikátor na bázi strojového učení s učitelem (viz kapitola 2). V našem případě byla jako matematický klasifikátor zvolena dopředná neuronová síť.

6.1 Dopředná neuronová síť

Jak již bylo zmíněno v kapitole 2, základními prvky dopředné neuronové sítě jsou umělé neurony, které jsou umístěny v několika vrstvách a které paralelně zpracovávají vstupní data dané vrstvy neuronové sítě. Matematický model umělého neuronu je vyjádřen rovnicí (6.1) viz obr. 6.1 [41].



Obr. 6.1. Matematický modelu umělého neuronu.

$$y = f\left(\sum_{i}^{N} w_i \cdot x_i - \theta\right), \tag{6.1}$$

kde x_i jsou vstupní hodnoty a N je počet vstupů daného umělého neuronu, w_i představují váhy jednotlivých vstupů, y odpovídá výstupní hodnotě daného umělého neuronu, θ je tzv. práh neuronu a $f(\alpha)$ je aktivační (přenosová) funkce neuronu.

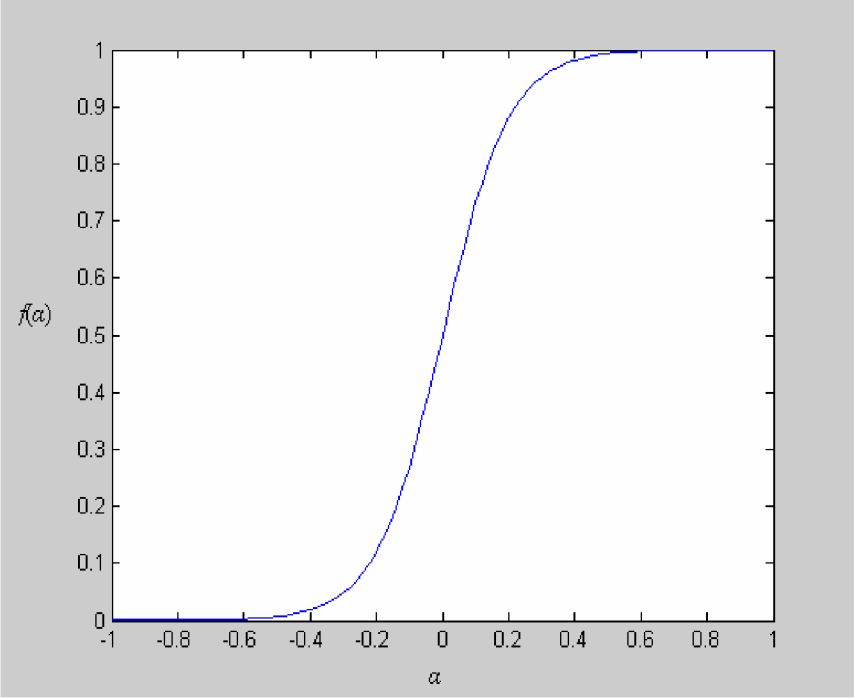
Váhy w_i ovlivňují jednotlivé vstupy do neuronů a tím i celou neuronovou síť. Každý vstup do neuronu je ohodnocen určitou hodnotou příslušné váhy spojení. Tato hodnota reprezentuje citlivost, s jakou příslušný vstup působí na výstup z neuronu. Váhy neuronu jsou vyjadřovány obvykle reálnými čísly, jejichž hodnoty vypovídají o průchodnosti, případně o důležitosti daného spojení. Váhy patří do skupiny parametrů, jejichž změnou je možné v procesu učení dosáhnout shody mezi výstupy zkoumaného procesu a výstupy neuronové sítě. Právě výpočty aktuálních hodnot vah a jejich postupné ladění představují podstatnou část učících algoritmů neuronových sítí [45].

Hodnota prahu θ určuje, kdy je neuron aktivní resp. neaktivní. Je-li hodnota vstupního signálu neuronu nižší než hodnota prahová, je na výstupu z neuronu signál odpovídající pasivnímu stavu neuronu. Jakmile dojde k překročení prahové hodnoty, stává se neuron aktivním. V modelech umělých neuronů je práh často používán k tomu, aby „posouval“ signál při vstupu do aktivační funkce $f(\alpha)$ [45].

Úkolem aktivační funkce $f(\alpha)$ je převést hodnotu vstupního proměnné α tzv. *aktivace* na výstupní hodnotu z neuronu. Konkrétní tvary přenosových funkcí bývají velmi různorodé. V principu se dají tyto funkce rozdělit na lineární a nelineární, případně na spojité a diskrétní. Výběr vhodné přenosové funkce je závislý na konkrétním typu řešené úlohy, případně na konkrétní poloze neuronu v neuronové síti [45]. Ve vícevrstvých dopředných neuronových sítích bývá nejčastěji používána jako aktivační funkce nelineární sigmoidní funkce (6.2) viz obr. 6.2:

$$f(\alpha) = \frac{1}{1 + \mathrm{e}^{\frac{-\alpha}{T}}}, \tag{6.2}$$

kde parametr T umožňuje měnit strmost přechodu funkce v okolí nuly.



Obr. 6.2. Průběh sigmoidní funkce.

Dopředné neuronové sítě bývají velmi často využívány při zpracování a klasifikaci signálů nejen pro svoji vhodnou architekturu, danou obecností realizovaného zapojení, ale zejména díky odvozenému formalizovanému optimálnímu postupu učení [28]. Tento postup nazývaný jako algoritmus zpětného šíření chyb (*backpropagation*) vychází z předpokladu existence trénovací množiny dat (x_p,d_p), kde x_p je určitý vstupní vektor příznaků délky N a d_p je jemu odpovídající požadovaný výstupní vektor délky M. Celkový postup se skládá

z učebních kroků tzv. epoch, které se opakují dokud není dosaženo požadovaného výsledku. Jednotlivá epocha se skládá z následujících částí [28]:

- postoupení vektoru x_p dané neuronové síti a zjištění její odezvy y_p,
- výpočet chybového vektoru $e_p=d_p-y_p$,
- výpočet rozpočetných chyb jednotlivých neuronů zpětným šířením chyb,
- oprava vektoru vah každého jednotlivého neuronu lokálně podle δ-pravidla s využitím známé aktivace tohoto neuronu při postoupení vektoru x_p neuronové síti.

Pro klasifikaci emocionálních výrazů ve tváři byla tedy použita třívrstvá dopředná neuronovou síť s učícím se algoritmem zpětného šíření chyb. Počet neuronů ve vstupní vrstvě závisí na počtu použitých příznaků, počet neuronů ve výstupní vrstvě je roven počtu rozpoznávaných výrazů a počet neuronů ve skryté vrstvě je na základě znalosti počtu neuronů ve zbývajících vrstvách stanoven explicitně. Jako aktivační funkce neuronů všech vrstev byla použita výše zmíněná sigmoidní funkce.

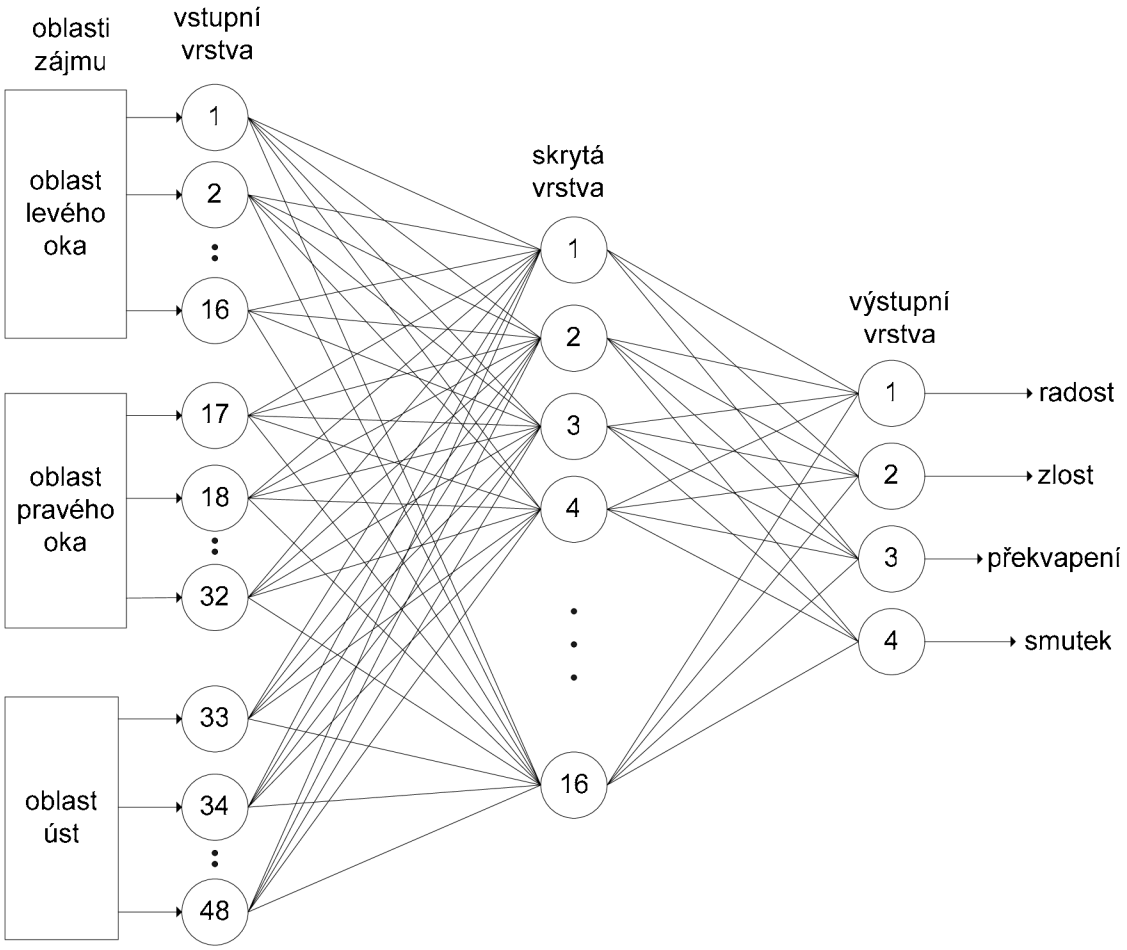
6.2 Testování rozpoznání emocionálních výrazů

V předchozích částech práce byly popsány metody extrakce jednotlivých příznaků a popsán matematický klasifikátor sloužící k rozpoznání jednotlivých emocionálních výrazů na základě těchto extrahovaných příznaků. Pro věrohodné ověření spolehlivosti jednotlivých metod je nezbytné použít kvalitní databázi zaměřenou na výrazy v obličeji. Bohužel v současnosti takových databází existuje velmi málo, a navíc ne všechny jsou volně dostupné.

V našem případě jsme zvolili databázi *JAFFE* (*Japanese Female Facial Expresion*) [31], která obsahuje 210 obrázků znázorňující 10 japonských modelek představujících 6 základních emocionálních výrazů a neutrální výraz viz obr. 6.3 (každý výraz je každým subjektem vyjádřen třikrát). Z důvodu malého počtu subjektů v databázi, byly pro učely testování vybrány pouze tyto čtyř emocionální výrazy: radost, zlost, překvapení a smutek. Jako trénovací množina pro nastavení neuronové sítě byly použity dva obrázky sedmi subjektů vyjadřující každý ze čtyř emocionálních výrazů. Zbývající obrázky těchto sedmi subjektů pro každý emocionální výraz byl společně s obrázky zbývajících tří subjektů použity pro účely testování.

Obr. 6.3. Příklad vyjádření jednotlivých základních emocionálních výrazů z databáze *JAFFE*, nahoře zleva: radost, zlost, překvapení, dole zleva: smutek, odpor, strach.

Stejně jako v [46] bylo pro vstup do neuronové sítě bylo zvoleno 16 nejvýznamnějších hlavních komponent ze tří oblastí zájmu (tj. celkem 48 příznaků). Aby mohla být porovnána vypovídací schopnost příznaků jednotlivých metod, byl zvolen stejný počet příznaků i pro zbývající metody. V případě lineární diskriminační analýzy tedy opět 16 nejvýznamnějších lineárních diskriminantů ze tří oblastí zájmu a v případě použití sady Gaborových filtrů pak 48 příznaků z celé obličejové oblasti redukovaných z původního počtu 880 příznaků pomocí analýzy hlavních komponent nebo lineární diskriminační analýzy. Dále pak byly experimentálně použity kombinace příznaků jednotlivých metod. Struktura klasifikační neuronové sítě je pak ukázána na obr. 6.4.



Obr. 6.4. Struktura neuronové sítě použité při rozpoznávání výrazů v obličeji.

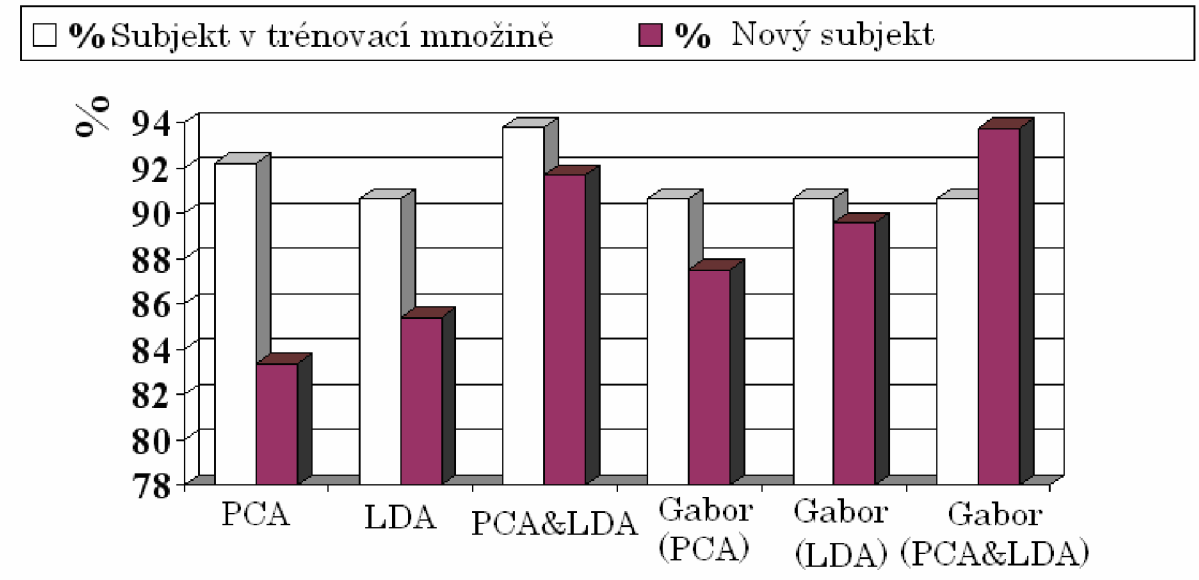
V případě testování jsme řešili dva úkoly: rozpoznání emocionálního výrazu představovaného subjektem, který byl zahrnut v trénovaní množině dat, a rozpoznání emocionálního výrazu nového subjektu [72]. Výsledky jednotlivých úloh jsou uvedeny v tab. 6.1 a tab.6.2. Celkové výsledky jsou potom shrnuty v grafu viz obr. 6.5.

Tab. 6.1. Úspěšnost klasifikace pro úlohu, kdy subjekt byl zahrnut v trénovací množině.

Použitá extrakční metoda	Zlost	Radost	Překvapení	Smutek	Celková přesnost
PCA	93 %	93 %	100 %	81 %	**92 %**
LDA	100 %	87 %	100 %	75 %	**90 %**
Kombinace PCA a LDA	93 %	93 %	100 %	87 %	**94 %**
Sada Gaborový filtrů, redukce pomocí PCA	100 %	87%	100 %	75 %	**90 %**
Sada Gaborový filtrů, redukce pomocí LDA	100 %	87 %	100 %	75 %	**90 %**
Sada Gaborový filtrů, redukce pomocí kombinace PCA a LDA	100 %	87 %	100 %	75 %	**90 %**

Tab. 6.2. Úspěšnost klasifikace pro úlohu, kdy subjekt nebyl zahrnut v trénovací množině.

Použitá extrakční metoda	Zlost	Radost	Překvapení	Smutek	Celková přesnost
PCA	83 %	92 %	92 %	66 %	**84 %**
LDA	100 %	100 %	92 %	50 %	**85 %**
Kombinace PCA a LDA	92 %	100 %	92 %	83 %	**92 %**
Sada Gaborový filtrů, redukce pomocí PCA	75 %	92 %	100 %	83 %	**87 %**
Sada Gaborový filtrů, redukce pomocí LDA	92 %	83 %	100 %	83 %	**89 %**
Sada Gaborový filtrů, redukce pomocí kombinace PCA a LDA	92 %	83 %	100 %	100 %	**94 %**



Obr. 6.5. Celkové výsledky procesu rozpoznávání emocionálních výrazů.

Na základě průběhu testování jsme provedli vyhodnocení vlastností jednotlivých metod použitých pro extrakci příznaků. Statistické metody *PCA* a *LDA* pracují přímo s intenzitou jasu jednotlivých pixelů. Jak bylo již dříve uvedeno, případné rozdílné intenzity jasu jednotlivých obrázků jsou potlačeny normalizací vstupních vektorů. Avšak pokud je intenzita jasu v daném obrázku rozložena nerovnoměrně, tyto metody selhávají. Tato problematika je potlačena použitím sady Gaborových filtrů, které pracují s rozdílem intenzit jasu sousedních pixelů. Avšak tyto filtry jsou velmi citlivé na změnu orientace, to znamená, že tato metoda selhává již při mírném natočení hlavy. Dále pak se tato metoda vyznačuje poměrně velkou výpočetní náročností (až 25-ti násobek výpočetního času u *PCA*

nebo *LDA*). Jednotlivé metody lze stručně charakterizovat na základě jejich vlastností následovně:

- Analýza hlavních komponent – rychlá a snadno implementovatelná metoda, ale podléhá vlivu nerovnoměrného rozdělení intenzity jasu.
- Lineární diskriminační analýza – rychlá a snadno implementovatelná metoda, která umožňuje rozdělení do tříd, ale podléhá vlivu nerovnoměrného rozdělení intenzity jasu a je vyžadováno předzpracování např. pomocí *PCA*.
- Sada Gaborových filtrů – metoda s nejlepšími výsledky, která je nezávislá na světelných podmínkách, ale vyžaduje vyšší výpočetní výkon a je závislá na úhlu natočení hlavy.

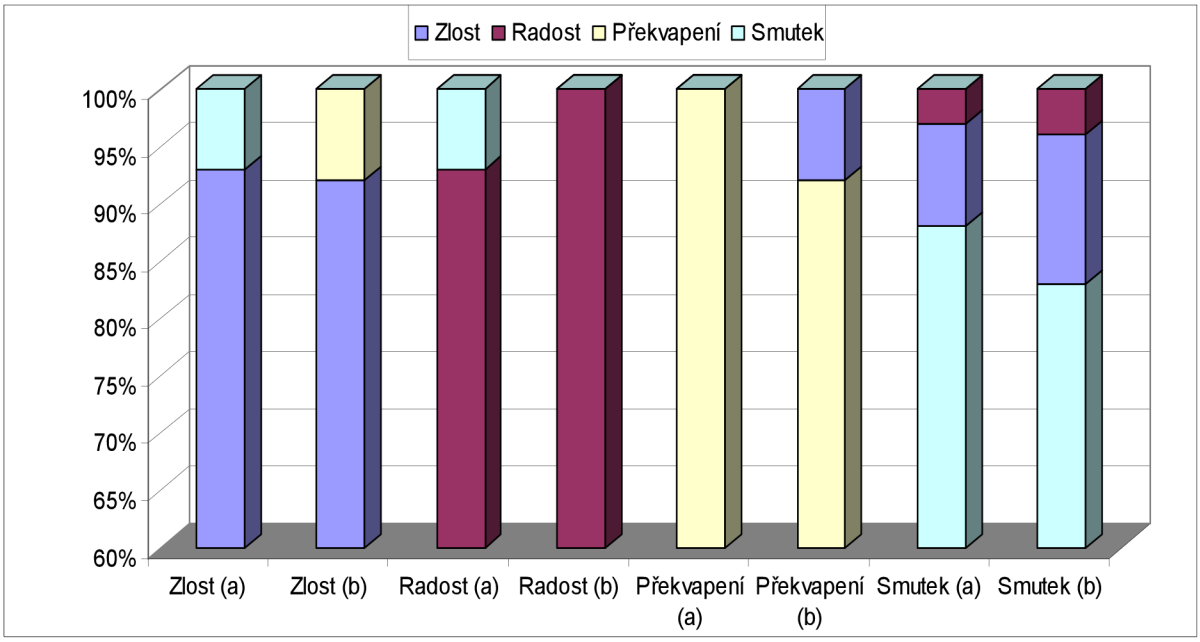
Z jednotlivých výsledků lze dále formulovat následující pozorování:

- Obecně platí, že úspěšnost rozpoznávání je vyšší pokud byl subjekt zahrnut v trénovaní množině dat. Jedinou výjimkou je situace, kdy bylo použito příznaků získaných filtrací sadou Gaborových filtrů a redukovaných pomocí kombinace *PCA* a *LDA*, kde se projevil vliv natočení hlavy jednoho ze subjektů v trénovací množině.
- Úspěšnost rozpoznávání závisí také na rozpoznávaném výrazu, kdy pro výrazy, které jsou většinou lidí vyjadřovány velmi podobně s jasně definovanými prvky (tj. např. překvapení – široce rozevřené oči, nadzvednuté obočí) byla úspěšnost výrazně vyšší, než pro výrazy, které jsou lidmi vyjadřovány odlišně a nemají žádný přesně definovaný charakter (tj. např. smutek).
- Kombinací příznaků extrahovaných pomocí různých metod jsme v mnoha případech schopni dosáhnout lepších výsledků než pomocí příznaků extrahovaných pouze pomocí jedné metody. Míra zlepšení závisí na celkové úspěšnosti jednotlivých metod, pokud jedna z metod dosahuje výrazně nižší úspěšnosti rozpoznání, tak se celková úspěšnost jejich kombinace blíží úspěšnosti metody druhé. Pokud obě metody dosahují nízké úspěšnosti při rozpoznání, může jejich kombinací dojít ke zvýšení celkové úspěšnosti nad úspěšnost jednotlivých metod.

Dále pak byly pro dvě nejúspěšnější metody stanoveny tabulky záměn jednotlivých klasifikovaných emocionálních výrazů (*confusion matrices*) viz tab. 6.3 a tab. 6.4, které vyjadřují procentuální zastoupení jednotlivých emocionálních výrazů při rozpoznávání daného výrazu. Graficky jsou tyto tabulky vyneseny v obr. 6.6 a obr. 6.7.

Tab. 6.3. Tabulka záměn získaná na základě kombinace *PCA* a *LDA* příznaků, a) subjekt byl zahrnut trénovací množině, b) nový subjekt.

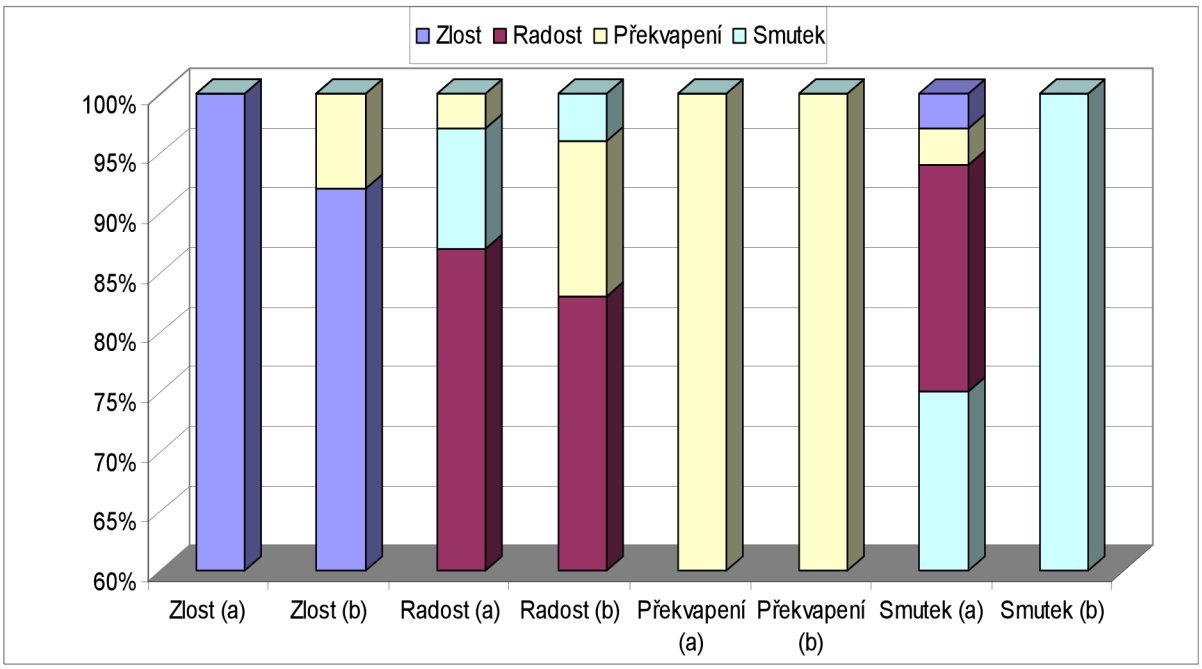
	Zlost		Radost		Překvapení		Smutek	
	a)	b)	a)	b)	a)	b)	a)	b)
Zlost	93 %	92 %	0 %	0 %	0 %	8 %	7 %	0 %
Radost	0 %	0 %	93 %	100 %	0 %	0 %	7 %	0 %
Překvapení	0 %	8 %	0 %	0 %	100 %	92 %	0 %	0 %
Smutek	3 %	4 %	9 %	13 %	0 %	0 %	87 %	83 %



Obr. 6.6. Grafická reprezentace tabulky záměn pro kombinaci *PCA* a *LDA* příznaků**,** **(**a) subjekt byl zahrnut trénovací množině, (b) nový subjekt.

Tab. 6.4. Tabulka záměn získaná na základě příznaků extrahovaných sadou Gaborových filtrů a redukovaných pomocí kombinace *PCA* a *LDA*, a) subjekt byl zahrnut trénovací množině, b) nový subjekt.

	Zlost		Radost		Překvapení		Smutek	
	a)	b)	a)	b)	a)	b)	a)	b)
Zlost	100 %	92 %	0 %	0 %	0 %	8 %	0 %	0 %
Radost	0 %	0 %	87 %	83 %	3 %	13 %	10 %	4 %
Překvapení	0 %	0 %	0 %	0 %	100 %	100 %	0 %	0 %
Smutek	3 %	0 %	19 %	0 %	3 %	0 %	75 %	100 %



Obr. 6.7. Grafická reprezentace tabulky záměn pro příznaky extrahovány sadou Gaborových filtrů a redukovaných pomocí kombinace *PCA* a *LDA*, (a) subjekt byl zahrnut trénovací množině, (b) nový subjekt.

Z jednotlivých výsledků v tabulkách záměn lze formulovat tato následující pozorování:

- Existuje zde nezanedbatelné procento nesprávných rozpoznání, které je nezávislé na použité metodě extrakce příznaků, tj. některé emocionální výrazy ve tváři byly nesprávně zařazeny do stejné třídy nezávisle na typu použitých příznaků.
- Existuje zde závislost mezi rozpoznávaným emocionálním výrazem a použitou metodou extrakce příznaků, tedy pro metodu založenou na sadě Gaborových filtrů byla úspěšnost rozpoznávání emocionálních výrazů překvapení a smutku vyšší než u metody založené na *PCA* a *LDA* a naopak u těchto metod byla vyšší úspěšnost rozpoznání emocionálních výrazů zlosti a radosti než u sady Gaborových filtrů.
- Existuje vazba (v obou úlohách) mezi jednotlivými emocionálním výrazy, tj. nejčastějším nesprávně klasifikovaným emocionálním výrazem pro radost byl smutek a obráceně. Toto lze pozorovat také pro emocionální výraz zlosti a překvapení.

Použitá databáze emocionálních výrazů *JAFFE* se bohužel nevyznačuje přílišnou variabilitou subjektů (subjekty jsou stejného pohlaví, stejné národnosti a přibližně stejného věku) a zároveň je celkový počet subjektů poměrně nízký. Abychom mohli ověřit robustnost použitých metod extrakce příznaků byly pro další účely testování vybrány obrázky z databáze *ORL Database of Faces* [54]. Tato databáze sice není orientovaná na rozpoznávání emocionálních výrazů, ale obsahuje dostatečný počet subjektů s velkou mírou variability (pohlaví, národnost, věk, vzhled), které se buď usmívají či neusmívají. Úkolem této části testování bylo tedy rozhodnout zda se daný subjekt usmívá či nikoliv. Z tohoto důvodu byla tedy pro extrakci příznaků uvažována pouze oblast okolo úst (35 × 40 pixelů okolo středu

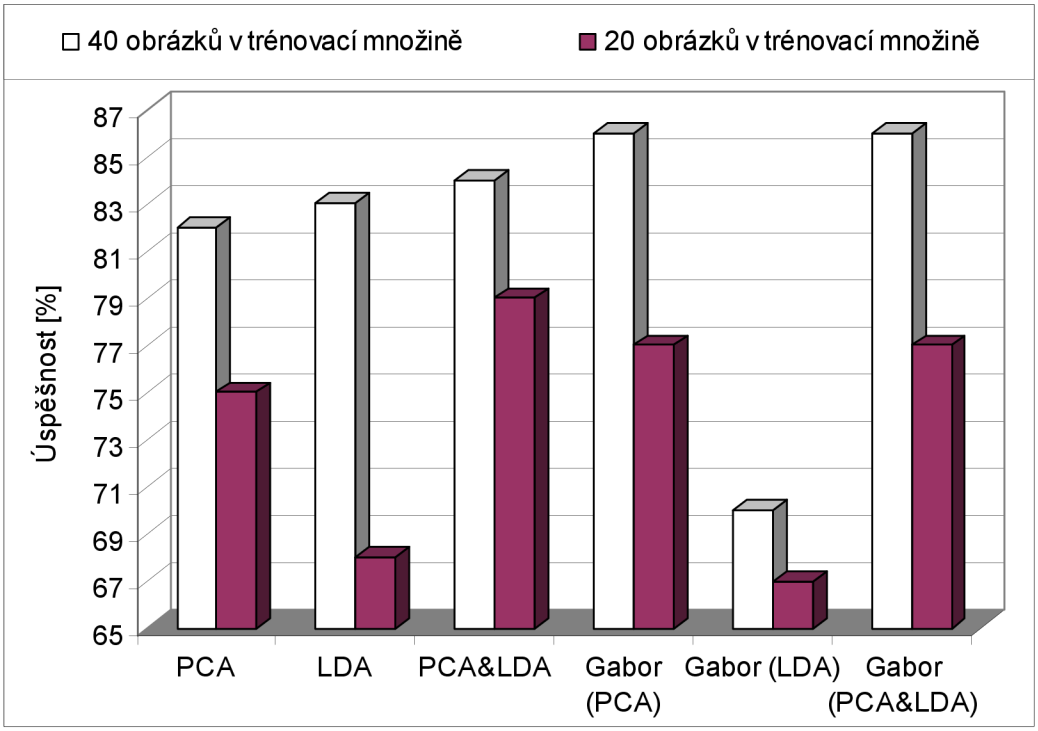
úst). Opět byla tuto úlohu rozdělena na dvě části: v první části bylo do trénovací množiny zahrnuto 40 různých obrázků pro každý z obou emocionálních stavů a v druhé části jich bylo do trénovací množiny zahrnuto pouze 20 pro každý z obou emocionálních stavů. Pro obě části pak byl počet testovaných obrázků roven 120 (60 usmívajících se tváří a 60 neusmívajících se tváří). Výsledky jsou uvedeny v tab. 6.5 a v tab. 6.6 Na obr. 6.8 je pak zobrazena jejich grafická reprezentace.

Tab. 6.5. Výsledky rozpoznávání úsměvu ve tváři - část 1 (40 obrázků v trénovací množině).

Použitá extrakční metoda	Úsměv?		Celková přesnost
	ANO	NE	
PCA	90 %	73 %	**82 %**
LDA	93 %	72 %	**83 %**
Kombinace PCA a LDA	93 %	73 %	**84 %**
Sada Gaborový filtrů, redukce pomocí PCA	80 %	92 %	**86 %**
Sada Gaborový filtrů, redukce pomocí LDA	67 %	73 %	**70 %**
Sada Gaborový filtrů, redukce pomocí kombinace PCA a LDA	80 %	92 %	**86 %**

Tab. 6.6. Výsledky rozpoznávání úsměvu ve tváři - část 2 (20 obrázků v trénovací množině).

Použitá extrakční metoda	Úsměv?		Celková přesnost
	ANO	ANO	
PCA	85 %	65 %	**75 %**
LDA	72 %	63 %	**68 %**
Kombinace PCA a LDA	87 %	72 %	**79 %**
Sada Gaborový filtrů, redukce pomocí PCA	72 %	82 %	**77 %**
Sada Gaborový filtrů, redukce pomocí LDA	65 %	68 %	**67 %**
Sada Gaborový filtrů, redukce pomocí kombinace PCA a LDA	72 %	82 %	**77 %**



Obr. 6.8. Celkové výsledky procesu rozpoznávání úsměvu ve tváři.

Z těchto výsledků můžeme opět odvodit určitá pozorování:

- Úspěšnost rozpoznání se zvyšuje s rostoucím počtem obrázků v trénovací množině, míra zvýšení závisí na míře variability subjektů v trénovací množině.
- V první části se neprojevil žádný významný rozdíl mezi samotnými příznaky a jejich vzájemnými kombinacemi, zatímco ve druhé části, kdy bylo vlivem menšího počtu obrázků v trénovací množině dosaženo nižší rozpoznávací schopnosti, se pozitivní vliv kombinace příznaků *PCA* a *LDA* projevil.
- Úspěšnost rozpoznání pomocí příznaků extrahovaných lineární diskriminační analýzou byla celkově nižší než bylo předpokládáno v návaznosti na testy s databází *JAFFE*. Pravděpodobně velká variabilita subjektů neumožnila přesné přiřazení jednotlivých příznaků do odpovídajících tříd, zejména při redukci příznaků extrahovaných sadou Gaborových filtrů pomocí LDA bylo dosaženo úspěšnosti jen o málo vyšší než by byla úspěšnost náhodného odhadu (tj. přibližně 50%). Z tohoto důvodu se nezdá využití lineární diskriminační analýzy pro rozpoznání emocionálních výrazů ve tváři při vyšší variabilitě subjektů vhodné.

Při procesu extrakce příznaků byly příslušné oblasti vybírány na základě pozic očí a úst lokalizovaných během procesu detekce tváře. Avšak tyto pozice jsou pouze přibližné, čímž jsme se při vlastní extrakci dopustili jistého zkreslení požadovaných příznaků vlivem nepřesného umístění daných oblastí, což mohlo negativně ovlivnit celý proces rozpoznávání emocionálních výrazů. Z tohoto důvodu bude dále pro přesnou lokalizaci očí a rtů použita metoda aktivního modelu tvaru.

6.3 Aktivní model tvaru

Princip metody aktivního modelu tvaru byl stručně popsán v kapitole 2. Jedná se tedy o vytvoření normalizovaného modelu tvaru sledovaného objektu na základě statistického zpracování manuálně vytvořených modelů. Tato metoda bývá velmi často používána pro nalezení popisu tvaru sledovaného objektu v biometrických a biomedicínckých aplikacích (popis tvaru obličeje, postavy, kostí a podobně). Metoda aktivního modelu tvaru se skládá ze dvou hlavních částí: statistického modelu tvaru a modelu profilů intenzit jasu. Zatímco statistický model tvaru vyjadřuje vzájemné vazby mezi pozicemi jednotlivých bodů modelu (reprezentuje geometrickou informaci), model profilů intenzit jasu popisuje obrazové okolí daného bodu (reprezentuje pozici modelu v obraze).

Statistický model tvaru

Tvar daného objektu může být popsán pomocí dvourozměrného vektoru obsahujícího souřadnice (x_i, y_i) N vhodně zvolených bodů. Tyto body by měly být voleny tak, aby co nejlépe popisovaly tvar zvoleného objektu a zároveň byly navzájem v určitém vztahu invariantní [9]. To znamená, že pokud bychom daný objekt posunuli, podrobili rotaci či změně měřítka, budou tyto body stále reprezentovat tentýž tvar objektu.

Nejprve je tedy nezbytné vytvořit trénovací množinu modelu tvarů pomocí manuálního umístění jednotlivých bodů pro různé variace tvaru daného modelu. Pro zjednodušení budeme dále uvažovat pouze jednorozměrný vektor tvaru $x' = (x_1, x_2, \ldots, x_N, y_1, y_2, \ldots, y_N)$, pro M vytvořených modelů tedy dostaneme M jednorozměrných vektorů x_i', pro $i=1\ldots M$, o délce $2\times N$. Následně z těchto modelů tvaru vytvoříme normalizované modely tvaru $x_i'^{(n)}$ a průměrný model tvaru $\bar{x}'$ tak, aby byl minimalizován součet rozdílů mezi jednotlivými normalizovanými modely tvaru a průměrným modelem tvaru měřený podle euklidovské vzdálenosti:

$$D_E(\bar{x}', x_i'^{(n)}) = \sqrt{(\bar{x}' - x_i'^{(n)})^2}, \tag{6.3}$$

Vlastní normalizace probíhá ve smyslu vzájemné změny posunu, změny velikosti a rotaci jednotlivých modelů. Popis vytvoření normalizovaných modelů tvaru a průměrného modelu tvaru sledovaného objektu je ukázán na obr. 6.9. Využívá se zda tzv. transformace podobnosti modelů (*similarity transformation*), která provádí rotaci modelu o úhel θ, posun modelu o souřadnice x_p a y_p a změnu velikosti o parametr s [9]:

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_p \\ y_p \end{pmatrix} + \begin{pmatrix} s\cdot\cos\theta & s\cdot\sin\theta \\ -s\cdot\sin\theta & s\cdot\cos\theta \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}. \tag{6.4}$$

1. Posun souřadnic každého modelu tak, aby jeho těžiště $t = (t_x, t_y)$ leželo v počátku souřadnicového systému, tj. $t_x = \sum_i^{2\cdot N} x_i = 0$ a $t_y = \sum_i^{2\cdot N} y_i = 0$.
2. Náhodný výběr jednoho modelu tvaru jako počátečního průměrného modelu tvaru $\bar{x}_0' = x_i'$ a normalizace jeho velikosti tak, aby platilo $|\bar{x}_0'| = \sqrt{\sum_i^{2\cdot N} \bar{x}_{0,i}' \cdot \bar{x}_{0,i}'} = 1$.
3. Zarovnání všech modelů tvaru k průměrnému modelu tvaru $\bar{x}_0'$ pomocí transformace podobnosti. Zarovnáním vznikají nové normalizované modely tvaru $x_i'^{(n)}$, které nahrazují původní modely tvaru.
4. Výpočet průměrného modelu tvaru odpovídajícímu střední hodnotě všech normalizovaných modelů tvaru $\bar{x}' = \frac{1}{M} \cdot \sum_i^M x_i'^{(n)}$.
5. Zarovnání průměrného modelu tvaru $\bar{x}'$ ku počátečnímu průměrnému modelu tvaru $\bar{x}_0'$ pomocí transformace podobnosti a normalizace jeho velikosti tak, aby platilo $|\bar{x}'| = \sqrt{\sum_i^{2\cdot N} \bar{x}_i' \cdot \bar{x}_i'} = 1$.
6. Opakování celého od bodu 4, dokud nedojde k požadované konvergenci (tj. hodnoty průměrného modelu tvaru se již významně nemění).

Obr. 6.9. Popis procesu normalizace modelů tvaru a vytvoření průměrného modelu tvaru [9].

V literatuře [9] je dále popsán způsob řešení transformace podobnosti pomocí metody nejmenších čtverců. Zavedeme-li substituci $a = s \cdot \cos\theta$ a $b = -s \cdot \cos\theta$, potom dostaneme:

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_\mathrm{p} \\ y_\mathrm{p} \end{pmatrix} + \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}. \tag{6.5}$$

Nyní hledáme takové parametry a, b, x_p a y_p, které minimalizují rozdílovou funkci E mezi transformovaným $T(x')$ a referenčním modelem x_r':

$$E(a, b, x_\mathrm{p}, y_\mathrm{p}) = \left(T(x') - x_\mathrm{r}'\right)^2 = \sum_i^N \left[\left(a \cdot x_i - b \cdot y_i - x_{\mathrm{r},i}\right)^2 + \left(b \cdot x_i + a \cdot y_i - y_{\mathrm{r},i}\right)^2\right]. \tag{6.6}$$

Odtud pak za předpokladu, že těžiště referenčního modelu tvaru se nachází v počátku souřadnicového systému, dostáváme:

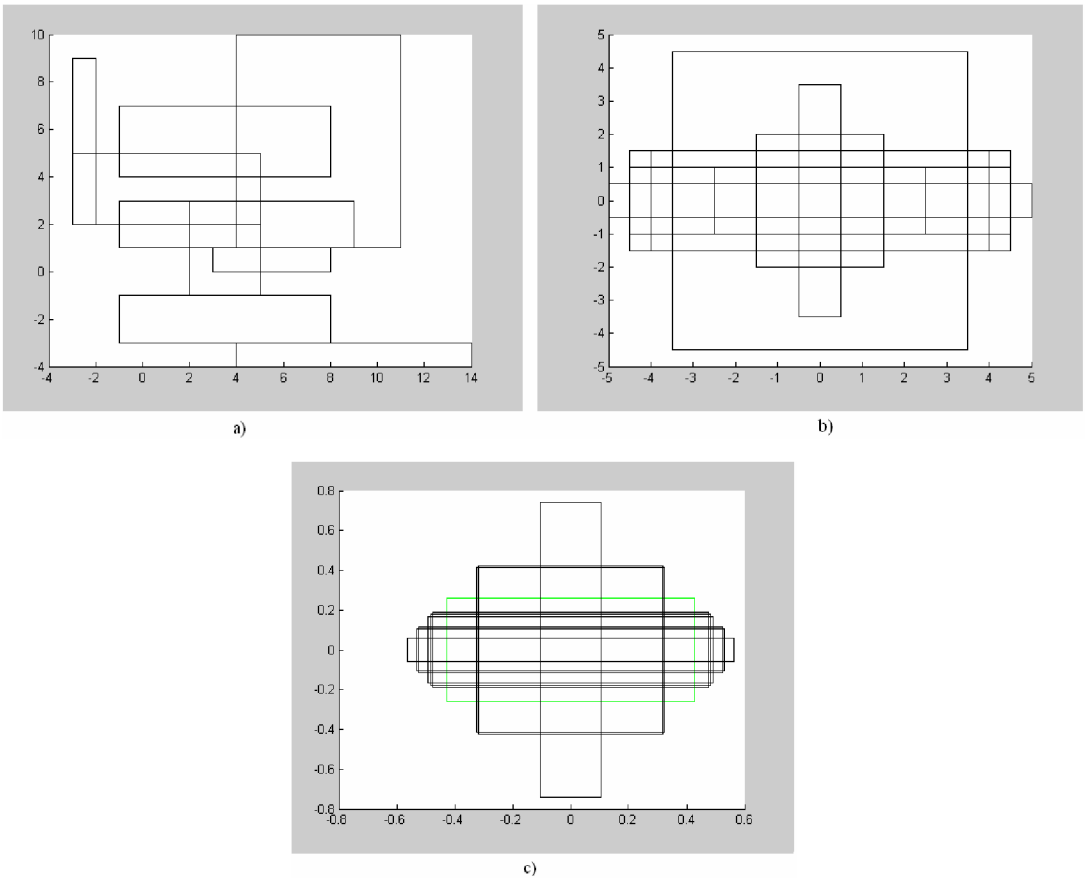
$$x_{\mathrm{p}} = \frac{1}{N} \cdot \sum_{i}^{N} x_i ,$$

$$y_{\mathrm{p}} = \frac{1}{N} \cdot \sum_{i}^{N} y_i ,$$

$$a = \frac{x_{\mathrm{r}}' \cdot x'}{\left|x'\right|^2}, \tag{6.7}$$

$$b = \left(\frac{\frac{1}{N} \sum_{i}^{N} x_i \cdot y_{\mathrm{r},i} - \frac{1}{N} \sum_{i}^{N} y_i \cdot x_{\mathrm{r},i}}{\left|x'\right|^2} \right).$$

Na obr. 6.10 je ukázán příklad použití normalizace modelů tvarů pro pravidelný čtyřúhelník s různým poměrem stran. Model tvaru zde zastupují čtyři dvojice souřadnic odpovídajícím rohovým bodům daného čtyřúhelníku. Jednotlivé čtyřúhelníky se kromě různých poměrů stran liší v celkové velikosti v posunutí oproti počátku souřadnicového systému, úhel rotace θ je konstantní.



Obr. 6.10. Normalizace modelů tvaru pravidelných čtyřúhelníků (spojnice mezi jednotlivými rohovými body nejsou součástí modelu tvaru a byly přidány pouze pro větší přehlednost): a) vstupní modely tvaru, b) vstupní modely tvaru s těžištěm posunutým do počátku souřadnicového sytému, c) normalizované modely tvaru a průměrný model tvaru (zvýrazněn zeleně).

Nyní tedy máme k dispozici normalizované modely tvaru a průměrný model tvaru, jak je ovšem vidět z obr. 6.10 jednotlivé normalizované modely tvaru se od průměrného modelu tvaru i od sebe navzájem značně liší. Z tohoto důvodu je potřeba vytvořit statistický model tvaru, který bude schopen reprezentovat všechny normalizované modely tvaru z trénovací množiny. Postup pro vytvoření takové statistického modelu je uveden v literatuře [9]. Pomocí analýzy hlavních komponent popsané v kapitole 5 vypočítáme matici vlastních vektorů U kovariační matice C množiny normalizovaných modelů tvaru:

$$C = \frac{1}{M}\sum_{i}^{M}(x_i'^{(n)} - \bar{x}') \cdot (x_i'^{(n)} - \bar{x}')^{T}. \tag{6.8}$$

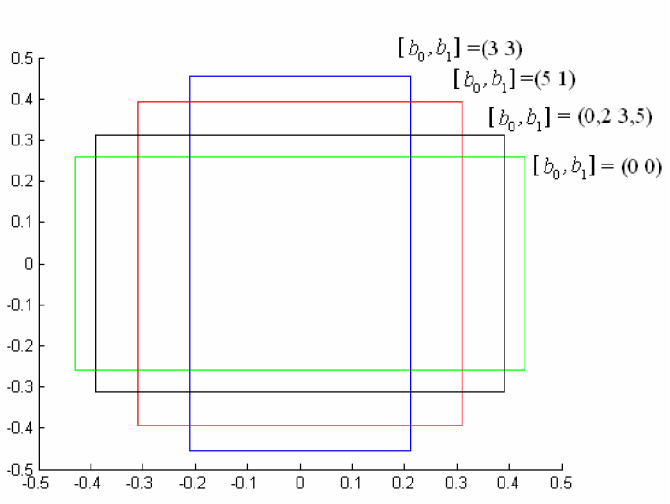
Nyní pak můžeme vyjádřit libovolný model tvaru z trénovací množiny jako:

$$x' \approx \bar{x}' + \Phi \cdot b, \tag{6.9}$$

kde Φ je matice obsahující t vlastních vektorů z matice U odpovídajících t největším vlastním číslům a b je t-rozměrný vektor definovaný vztahem:

$$b = \Phi^{T} \cdot (x' - \bar{x}'). \tag{6.10}$$

Vektor b zde definuje sadu parametrů deformačního modelu. Změnou jednotlivých prvků vektoru b jsme nyní schopni měnit tvar zvoleného modelu. Variance i-tého prvku vektoru b je dána hodnotou i-tého nejvyššího vlastního čísla λ_i. Pokud zajistíme, aby hodnota i-tého prvku vektoru b nepřekročila mez $\pm 3 \cdot \sqrt{\lambda_i}$, pak modely tvaru generované pomocí tohoto vektoru budou přibližně odpovídat modelům z trénovací množiny. Počet vlastních vektorů t je volen tak, aby byla zachována dostatečná variance modelu a zároveň došlo k potlačení určitých nepřesností (obvykle bývá volen počet nejvyšších vlastních čísel jejichž součet odpovídá 98% součtu všech vlastních čísel). Na obr. 6.11 je ukázána deformace průměrného modelu tvarů z obr 6.10 pomocí změny parametrů b_0 a b_1.

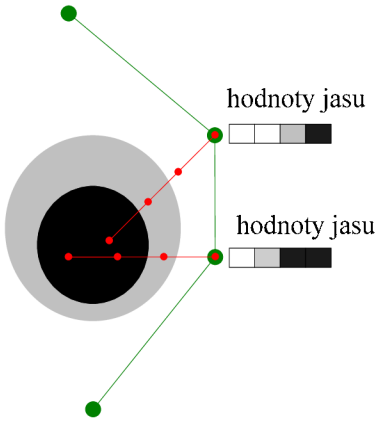


Obr. 6.11. Deformace průměrného modelu tvarů pomocí parametrů b_0 a b_1.

Model profilů intenzity jasu

Úkolem této části aktivního modelu tvaru je vytvoření modelu, který bude nejlépe reprezentovat daný statistický model tvaru v konkrétním obraze. Tento model bude obsahovat stejný počet bodů se stejným rozmístěním jako model statistického tvaru, ovšem nyní nás nebude zajímat pozice a vzájemný vztah těchto bodů, ale vztah těchto bodů vzhledem k samotnému obrazu [9]. Jedná se tedy o zprostředkování vazby statistického modelu tvaru na daný obraz. Měli bychom tedy pro každý bod statistického modelu tvaru vytvořit ideální soubor příznaků, který bude tento bod v daném obraze jednoznačně reprezentovat, tj. na základě znalosti těchto příznaků bychom měli být schopni daný bod kdekoliv v obraze nalézt. Tato podmínka však není reálně splnitelná pro všechny typy obrazů a všechny body statického modelu tvaru. Uvážíme-li ovšem existenci samotného statistického modelu tvaru, pak není zapotřebí vyhledávat daný bod v celém obraze, ale pouze v nejbližším okolí, kde se tento bod v současnosti nachází, tj. kam byl umístěn pomocí statistického modelu tvaru. Dojde tedy pouze k upřesnění jeho pozice na základě podobnosti s obrazem. Jelikož bude prohledávána pro každý bod pouze malá část obrazu, nemusí být zvolené příznaky příliš robustní, ale spíše jednoduché, aby proces vyhledávání byl co nejrychlejší. Často jsou proto k tomuto využívány profily intenzity jasu [3].

Profilem intenzity jasu je v tomto případě míněn vektor délky N obsahující hodnoty odpovídající rozdílu hodnot jasů jednotlivých pixelů v okolí dané bodu, které se nacházejí ve směru kolmém na spojnici dvou sousedících bodů viz obr. 6.12. Pro každý bod každého statistického modelu tvaru určíme tento profil intenzity jasu a tím získáme model profilů intenzity jasu g_i pro každý statistický model tvaru z trénovací množiny. Pomocí nich určíme průměrný model profilů intenzit jasu $\bar{g}$ a kovarianční matici C_g [9].



Obr. 6.12. Příklad určení profilů intenzit jasu (zeleně – body statistického modelu tvaru a jejich spojnice, červeně – kolmice na spojnici sousedících bodů a jim odpovídající pixely). Vykreslena pouze jedna polovina každého profilu jasu.

Při vyhledávání odpovídajícího umístění daného bodu v obraze pak postupujeme podél spojnice sousedících bodů a určíme K profilů intenzity jasu kolmých na tuto spojnici a od sebe navzájem vzdálených o P pixelů ve směru této spojnice. Vyhledávání probíhá v obou směrech od současné pozice daného bodu, dostáváme tedy $2K + 1$ profilů intenzity jasu. Nové

pozici bodu pak odpovídá pozice profilu intenzity jasu s minimální Mahalanobisovou vzdáleností od průměrného profilu intenzity jasu pro daný bod:

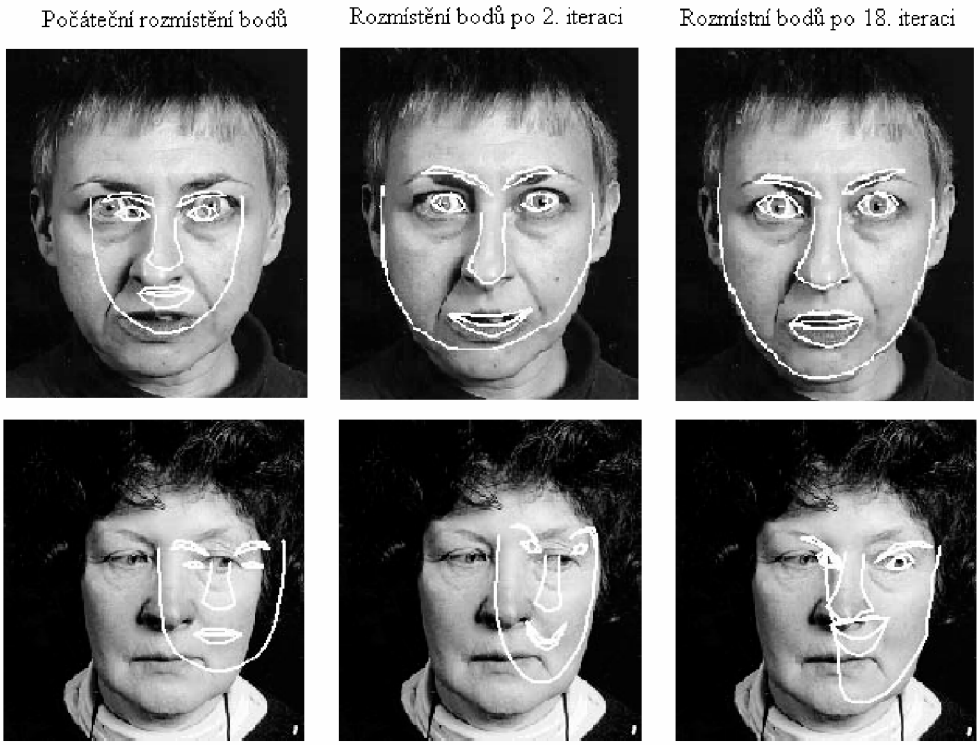
$$D_{\mathrm{M}}(g_{n,k}) = (g_{n,k} - \bar{g}_n)^{\mathrm{T}} \cdot C_{\mathrm{g}} \cdot (g_{n,k} - \bar{g}_n), \text{ pro } k = 1..2 \cdot K + 1 \text{ a } n = 1..N, \tag{6.11}$$

kde $g_{n,k}$ odpovídá *k*-tému profilu intenzity jasu pro *n*-tý bod.

Celkový postup při aplikaci aktivního modelu tvaru pomocí statistického modelu tvaru a modelu profilů intenzity jasu v testovaném obraze je popsán na obr. 6.13. Přičemž významnou úlohu zde hraje odpovídající počáteční rozmístění bodů modelu tvaru v tomto obraze viz obr. 6.14.

1. Počáteční umístění jednotlivých bodů aktivního modelu tvaru v obraze.
2. Nalezení předpokládaných pozic těchto bodů v jejich okolí pomocí modelu profilů intenzity jasu.
3. Aplikování statistického modelu tvaru na nově nalezené pozice bodů, tj. změna pozic bodů dle jejich vzájemných pozičních vztahů.
4. Opakování celého procesu od bodu 2, dokud nedojde k požadované konvergenci (tj. hodnoty nalezeného modelu tvaru se již významně nemění).

Obr. 6.13 Postup při aplikaci aktivního modelu tvaru na nový obraz [9].

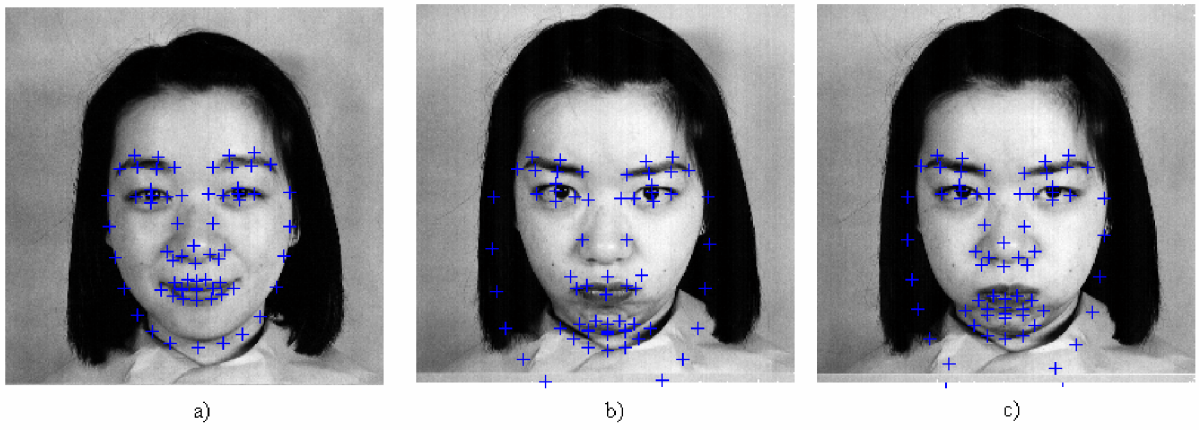


Obr. 6.14 Ukázky aplikace aktivního modelu tvarů při popisu tvaru obličeje v závislosti na počátečním rozmístění jednotlivých bodů modelu tvaru [9] (model tvaru je zde reprezentován pouze spojnicemi vybraných bodů a nikoli body samotnými).

6.4 Testování rozpoznání emocionálních výrazů s použitím aktivního modelu tvaru

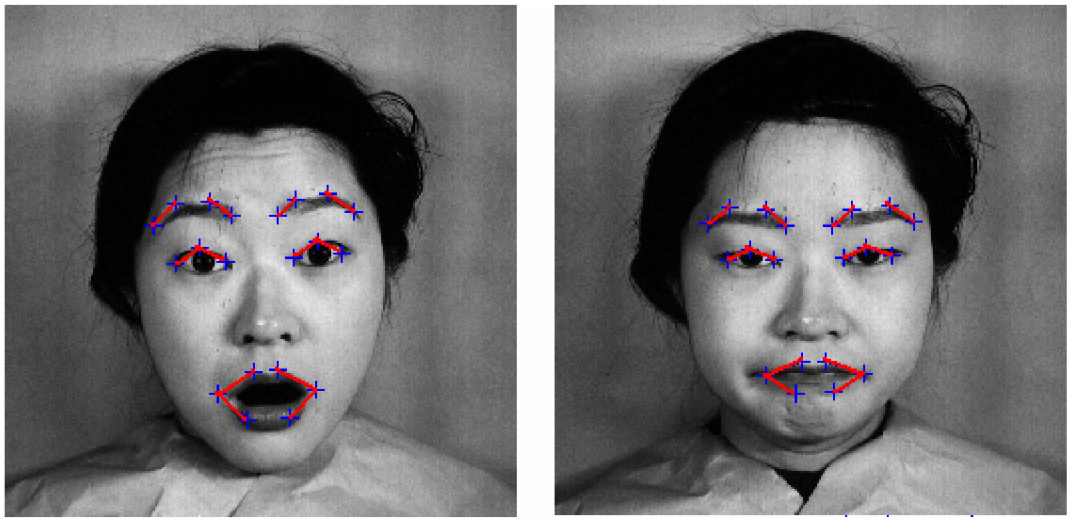
V předchozí části byl popsán přístup k využití aktivního modelu tvaru k popisu tvaru lidského obličeje. Aby byl výsledný model tvaru v praxi použitelný, vyžaduje velké množství modelů tvaru v trénovací množině. Tyto modely jak již bylo zmíněno je nutno vytvářet manuálně, což je časově velmi náročné. Z tohoto důvodu byl v rámci této práce využit již existující aktivní model tvaru lidského obličeje, který je realizován pomocí knihovny *Stasm* [42] popsané v literatuře [43]. Tento aktivní model tvaru se skládá z 68 bodů viz obr. 6.15, kde pro každý bod je definován profil intenzity jasu o délce 17 pixelů.

Tento aktivní model tvaru byl aplikován na databázi *JAFFE*, přičemž pro co nejpřesnější počáteční rozmístění bodů bylo využito nejen znalosti pozice samotného obličeje, ale i přibližné pozice očí, viz kapitola 4. Tímto jsme získali pro každý obrázek z databáze *JAFFE* odpovídající model tvaru daného obličeje, pouze u dvou obrázků tento model neodpovídal skutečnosti viz obr. 6.15.



Obr. 6.15. Aplikace aktivního modelu tvaru na obrázky z databáze *JAFFE*, a) správně umístěný model tvaru,, b) a c) nesprávně umístěný model tvaru.

Pro přesnou lokalizaci dané oblasti pro extrakci příznaků nám v tomto případě postačí pouze znalost pozice tří bodů (střed obou očí o střed rtů). Avšak může využít také znalosti pozic vybraných bodů jako samotných příznaků, které budou použity v procesu rozpoznání emocionálních výrazů ve tváři. Nezdá se však příliš vhodné použít přímo dané pozice bodů jako vlastní příznaky (vzhledem k variabilitě antropologického uspořádání lidského obličeje), ale je výhodnější jako příznaky použít vzájemné vazby mezi jednotlivými vybranými body. Jako vhodná vazba mezi jednotlivými body byla vybrána směrnice spojnice těchto bodů. Celkem bylo použito 12 vybraných směrnic pro doplnění vstupního souboru příznaků. Spojnice bodů odpovídající vybraným směrnicím jsou zobrazeny na obr. 6.16.



Obr. 6.16. Vliv emocionálního výrazu na směrnice spojnic vybraných bodů .

Nyní pro účely testování zvolíme stejný počet subjektů jako v předcházející části, avšak pokusíme se o rozpoznání všech šesti základních emocí i neutrálního výrazu. Vlastní rozpoznávání bude provedeno pomocí dvou nejúspěšnějších metod z předchozí části testování (tj. kombinace příznaků *PCA* a *LDA* (metoda 1) a sada Gaborových filtrů redukovaná pomocí kombinace *PCA* a *LDA* (metoda 2)) a nové metody založené na kombinaci *PCA* příznaků, na sadě Gaborových filtrů redukovaných pomocí *PCA* a na směrnicích získaných z modelu tvaru (metoda 3). Přičemž pro poslední jmenovanou metodu bude počet *PCA* příznaků z každé oblasti snížen na 10, dále bude použito pouze 18 Gaborových filtrů (6 různých prostorových orientací a 3 různé délky viz (6.12)), jejichž výstup bude redukován na 30 hodnot, a bude nově použito 12 směrnic spojnic vybraných bodů modelu tvaru. U všech metod bude využita přesná pozice oblasti potřebná k extrakci daných příznaků lokalizovaná pomocí aktivního modelu tvaru. Výsledky jsou uveden v tab. 6.7 a v tab. 6.8 a graficky znázorněny na obr. 6.17.

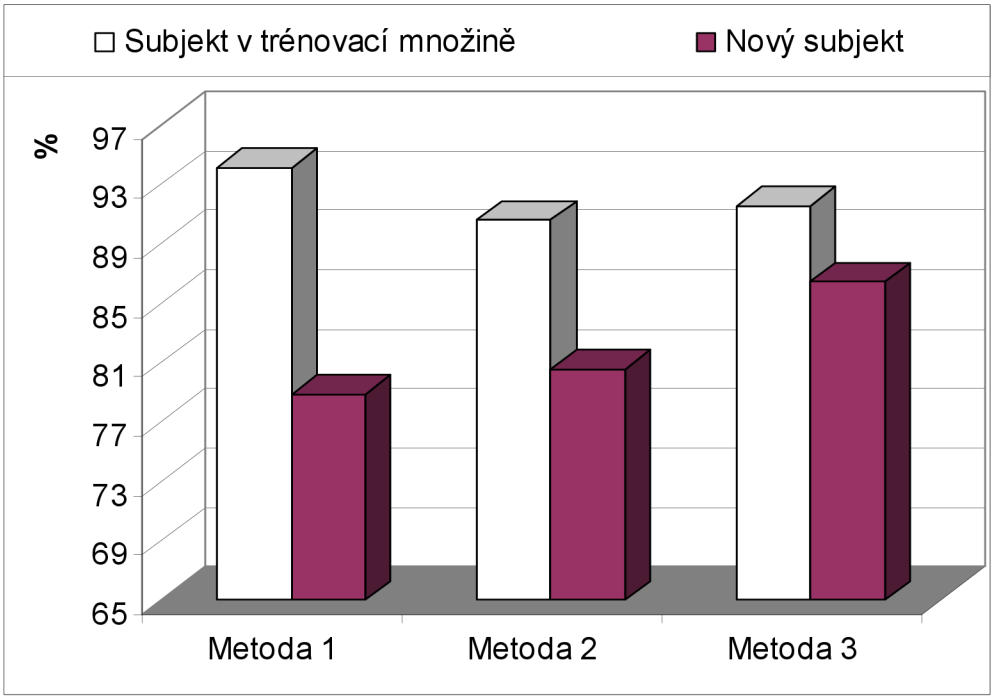
$$\begin{aligned} \theta &= \left\{0, \frac{\pi}{6}, \frac{2\pi}{6}, \frac{3\pi}{6}, \frac{4\pi}{6}, \frac{5\pi}{6}\right\}, \\ \lambda &= \left\{4, 4\sqrt{2}, 8\sqrt{2}\right\}, \\ \sigma &= \lambda, \\ \psi &= 0. \end{aligned} \tag{6.12}$$

Tab. 6.7. Úspěšnost klasifikace pro úlohu, kdy subjekt byl zahrnut v trénovací množině.

Použitá extrakční metoda	Zlost	Radost	Překvap.	Smutek	Strach	Odpor	Neutr.	Celková přesnost
Metoda 1	93 %	100 %	100 %	88 %	87 %	91 %	100 %	**94 %**
Metoda 2	88 %	93 %	100 %	83 %	87 %	83 %	100 %	**91 %**
Metoda 3	88 %	100 %	100 %	88 %	78 %	87 %	100 %	**92 %**

Tab. 6.8. Úspěšnost klasifikace pro úlohu, kdy subjekt byl zahrnut v trénovací množině.

Použitá extrakční metoda	Zlost	Radost	Překvap.	Smutek	Strach	Odpor	Neutr.	Celková přesnost
Metoda 1	83 %	100 %	79 %	83 %	46 %	68 %	93 %	**79 %**
Metoda 2	85 %	89 %	83 %	100 %	55 %	59 %	93 %	**81 %**
Metoda 3	89 %	93 %	86 %	93 %	87 %	64 %	93 %	**86 %**



Obr. 6.17. Celkové výsledky procesu rozpoznávání základních emocionálních výrazů.

Z dosažených výsledků můžeme odvodit následující závěry:

- Pomocí přesné lokalizace jednotlivých částí potřebných při extrakci příznaků bylo dosaženo lepších výsledků než při prvním testování. To je vidět zejména při úloze, kdy byl sledovaný subjekt zahrnut v trénovací množině (radost a smutek). A to i přes to, že celková přesnost je nižší než při prvním testování, což je ovšem způsobeno vyšším počtem rozpoznávaných emocionálních výrazů.
- Použitím kombinace různých příznaků (metoda 3) je možné dosáhnout lepších výsledků i při nižším počtu samotných příznaků (metody 1 a 2 využívaly 96 příznaků, zatímco metoda 3 pouze 72). To se projevilo zejména v druhé části testu, kdy subjekt nebyl zahrnut v trénovací množině.
- U metody 3 byl snížením počtu Gaborových filtrů zároveň snížen výpočetní čas (více než o 50% oproti metodě 2), a to při dosažení vyšší spolehlivosti rozpoznání daného emocionálního výrazu v obou částech testu.

7 Závěr

V rámci této práce byl popsán návrh automatického systému pro rozpoznávání emocionálních výrazů ve tváři ze statických obrazů. Tento systém se skládá ze tří částí, kde každé části byla v práci věnována samostatná kapitola popisující způsob řešení dané problematiky. V první části byl navržen vlastní detektor obličeje založený na detekci barvy lidské kůže. Při návrhu tohoto detektoru byla jedním z hlavních kritérií jeho robustnost, což se ovšem v důsledku ukázalo jako jeho největší slabina, jelikož při testování na obrazové databázi *Georgia Tech Face Database* dosáhl sice úspěšnosti přibližně 93%, avšak celková doba detekce jednoho obrázku přesahovala jednu sekundu. Z tohoto důvodu byl pro detekci obličeje zvolen objektový detektor *Viola-Jones,* který pro tento účel bývá v praxi často používán. V návaznosti na předchozí poznatky získané při návrhu detektoru založeném na barvě kůže, byl detektor *Viola-Jones* pro zvýšení robustnosti doplněn o zpracování barevné informace obrazu a o shlukovou analýzu. Spolehlivost takto modifikovaného detektoru *Viola-Jones* byla opět testována na obrazové databázi *Georgia Tech Face Database,* kde bylo dosaženo přibližně stejné spolehlivosti jako pro detektor založený na barvě kůže, ovšem potřebná doba detekce byla téměř dvanáctkrát nižší. Dále pak byl detektor *Viola-Jones* experimentálně využit pro detekci očí v již nalezených tvářích, kde bylo dosaženo úspěšnosti 88%. V rámci druhé části navrženého systému byla pro extrakci příznaků použita převzatá metoda analýzy hlavních komponent na jejímž základě bylo navrženo i použití lineární diskriminační analýzy. Jako třetí metoda byla pro extrakci příznaků použita sada 40 Gaborových filtrů doplněná o 22-bodovou obličejovou masku. Všechny tyto tři metody byly ve třetí části použity pro rozpoznání čtyř základních emocionálních výrazů z obrazové databáze *Japanese Female Facial Expresion Database,* kde pro úlohu klasifikace byla zvolena třívrstvá dopředná neuronová síť. Experimentálně zde bylo také využito kombinace příznaků extrahovaných pomocí různých metod. Pomocí těchto kombinací bylo dosaženo lepších výsledků než při použití příznaků metod samotných. Z důvodu malé variability subjektů v použité databázi bylo použito několika obrázků z databáze *ORL Database of Faces* pro rozpoznání úsměvu. Zde se projevila zejména nevhodnost použití lineární diskriminační analýzy pro subjekty s velkou variabilitou vzhledu. Na závěr bylo použito pro přesnou lokalizaci jednotlivých oblastí extrakce příznaků metody aktivního modelu tvaru. Pomocí obrazové databáze *Japanese Female Facial Expresion Database* bylo provedeno testování úspěšnosti rozpoznání všech šesti základních emocionálních výrazů a neutrálního výrazu pro dvě nejúspěšnější metody z první skupiny testů, nyní však navíc využívající aktivního modelu tvaru pro přesnou lokalizaci jednotlivých oblastí extrakce příznaků. Dále pak byla navržena nová metoda kombinující příznaky získané pomocí analýzy hlavních komponent, redukované sady Gaborových filtrů a směrnic spojnic vybraných bodů aktivního modelu tvaru. Tato metoda dosáhla ze všech tří výše zmíněných metod nejvyšší míry úspěšnosti (přibližně 86%) a to i přes to, že zde byl použit menší počet příznaků než u metod ostatních.

Literatura

Seznam použité literatury

[1] AUGUSTEIJN, M. F., SKJUCA, T.L.: Identification of Human Faces through Texture-Based Feature Recognition and Neural Network Technology, *Proceedings IEEE Conference Neural Networks*, 1993.

[2] BALLARD, C., BROWN, C.: *Computer Vision*. Prentice-Hall, 1982.

[3] BARTLETT, M. S.: *Face Image Analysis by Unsupervised Learning and Redundancy Reduction*. PhD thesis, University of California, san Diego. 1998.

[4] BELUR, V. D.: *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*, 1991. ISBN 0-8186-8930-7.

[5] BLACK, M. J., FLEET, D.J., YACOOB, Y. A.: A Framework for Modeling Appearance Change in Image Sequences. *Sixth International Conference on Computer Vision*, IEEE, Computer Society Press, 1998.

[6] BOOMGAARD, B.: Image Transforms Using Bitmapped Binary Images, *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing*, May, 1992.

[7] BURGES, J. C.: A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery 2*, s. 121 - 167, 1998.

[8] COLANTONI, P: *ColorSpace*. Dostupný z *http://www.couleur.org.*

[9] COOTES, T., TAYLOR, C., COOPER, D., GRAHAM, J.: Active shape models-their training and application, *Computer Vision and Image Understanding* 61(1), s. 38-59, 1995.

[10] DARWIN, C.: The Expression of the Emotions in Man and Animals. J.Murray, London, 1872. ISBN: 10-0195112717.

[11] DUDA, R. O., HART, P. E., STORK, D. G.: Unsupervised Learning and Clustering, Ch. 10 in *Pattern classification*. Wiley. s. 571, 2001. ISBN 0-471-05669-3, 2001.

[12] DUDA, R. O., HART, P. E., STORK, D. G.: *Pattern Classification.* Wiley-Interscience, 2000. 0471056693.

[13] EDWARDS, G. J., COOTES, T. F., TAYLOR, C. J.: Face Recognition using Active Appearance Models. *Proceedings of the 5th European Conference on Computer Vision*, University of Freiburg, Germany, 1998.

[14] EKMEN, P., FRIESEN, W.: Constants Across Cultures in the Face and Emotion. *Journal of Personality and Social Psychology*. 1971. ISSN: 0022-3514.

[15] EKMAN, P. FRIESEN, W. V.: *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, Calif.: Consulting Psychologists Press (1978).

[16] ESSA, I., PENTLAND, A.: Coding, Analysis, Interpretation and Recognition of Facial Expressions. *IEEE Transactions on pattern analysis and machine intelligence*. 1997.

[17] ETHEM, A.*: Introduction to Machine Learning*. s. 445, 2004 ISBN 978-0-262-01211-9.

[18] FASEL, B., LUETTIN, J.: Automatic Facial Expression Analysis: A Survey. *Pattern Recognition*. 2003.

[19] FAWCETT, T.: An introduction to ROC analysis. *Pattern Recognition Letters. Vol.* 27, 2006.

[20] FISHER, R.A.: The Statistical Utilization of Multiple Measurements. *Annali of Eugenics*, 8 (1938) 376-386.

[21] FELLENZ, W. A., TAYLOR, J. G.: Comparing Temple-based, Feature-based and Supervised Classification of Facial Expression from Static Images. *Proceedings of Circuits, Systems, Communications and Computers*, 1999.

[22] FRIEDMAN, J., HASTIE, T., TIBSHIRANI, R.: Additive logistic regression: a statistical view of boosting. *Ann. Statist.*, Vol. 28, pp. 337--407, 2000.

[23] GARCIA, C., ZIKOS, G., TZIRITAS, G.: Face Detection in Color Images using Wavelet Packet Analysis. *IEEE Int. Conf. on Multimedia Computing and Systems*, 1999.

[24] HAN, C.: Fast Face Detection via Morphology-Based Pre-Processing. *Proceedings of Ninth International Conference on Image Analysis and Processing,* 1998. s. 469-476.

[25] HSU, R-L., ABDEL-MOTTALEB, M., JAIN, A. K.: Face Detection in Color Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, 2002.

[26] CHAIN, D., NGAN, K. N.: Face segmentation using skin-color map in videophone applications. *Circuits and Systems for Video Technology.* Vol. 9, 1999.

[27] CHALOUPKA, J.: *Rozpoznávání akustického signálu řeči s podporou vizuální informace*, disertační práce Technická univerzita v Liberci, 2005.

[28] JAN, J.: *Číslicová filtrace, analýza a restaurace signálů*. Brno: VUTIUM Publishing, 2002. 427 pages. ISBN 80-214-1558-4.

[29] JESORSKY, O., KIRCHBERG, K., FRISCHHLOZ. R.: Robust Face Detection Using the Hausdorff Distance In J. Bigun and F. Smeraldi, editors, *Audio and Video based Person Authentication - AVBPA 2001*, s. 90-95. Springer, 2001.

[30] JOLLIFE I.T.: *Principal Component Analysis*, New York, Springer Verlag (1986).

[31] KAMACHI, M., LYONS, M., GYOBA, J.: *Japanese Female Facial Expresion Database,* Psychology Department in Kyushu University, *http://www.kasrl.org/jaffe.html.*

[32] KANADE, T., COHN, J.: *Cohn-Kanade AU-Coded Facial Expression Database*, Robotics Institute of Carnigie Mellon University, *http://vasc.ri.cmu.edu/idb/html/face/facial_expression.*

[33] KAWATO, S., OHYA, J.: Automatic Skin-color Distribution Extraction for Face Detection and Tracking, *ICSP2000: The 5th Int. Conf. on Signal Processing*, August 2000, Beijin, China.

[34] KOBAYASHI, H., HARA, F.: Facial Interaction between Animated 3D Face Robot and Human Beings. *Proceedings of the International Conference on Systems*, *Man and Cybernetics*. 1997.

[35] KOTSIANTIS, S., PINTELAS, P.: Recent Advances in Clustering: A Brief Survey, *WSEAS Transactions on Information Science and Applications*, Vol. 1, No 1 (73-81), 2004.

[36] KRÁTKÝ, M., SKOPOAL, T., SNÁŠEL, V.: *Efektivní vyhledávání v kolekcích obrázků tváří*, DATAKON 2003, BRNO, 2003.

[37] LANITIS, A., TAYLOR, C.J., COOTES, T.F.: An automatic Face Identification System Using Flexible Appearance Models. *Image and Vision Computing*, 1995. Volume 13. s 393-401.

[38] LIESETTI, C. L., RUMELHART, D. E.: Facial Expression Recognition using a Neural Network. *Proceedings of the 11th International Flairs Conference*. AAAI Press, 1998.

[39] LIENHART, R. and MAYDT, J.: An extended set of Haar-like features for rapid object detection, *ICIP02*, s. 900-903, 2002.

[40] LYONS, M. J., BUDYNEK J., AKAMTSU, S.: Automatic Classification of Single Facial Images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1999.

[41] McCULOCH, W. S., PITTS, W. H.: A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, s.115-133,1943.

[42] MILBORROW, S.: *Stasm library*, University of Cape Tlen, *http://www.milbo.users.sonic.net*.

[43] MILBORROW, S.: *Locating Facial Features with Active Shape Models*, Master's thesis. University of Cape Town (Department of Image Processing. 2007.

[44] NEFIAN, A. V.: *Georgia Tech face database*, *http://www.anefian.com/face_reco.htm.*

[45] NOVÁK, M. a kol.: *Umělé neuronové sítě teorie a aplikace*. Praha: C.H.BECK, s. 382, 1998. ISBN 80-7179-132-6.

[46] PADGETT, C., COTTRELL, G. W.: Representing Face Image for Emotion Classification. *Advances in Neural Information Processing Systems*, Cambridge, MA, 1997.

[47] PANTIC, M., ROTHKRANTZ, J.M.: Automatic Analysis of Facial Expression: The State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12) (2000), s. 1424-1445.

[48] PANTIC, M., ROTHKRANTZ M. J.: Expert System for Automatic Analysis of Facial Expression, *Image and Vision Computing Journal*, 2000.

[49] PETKOV, N., WIELING, M.B.: Gabor Filtering Augmented with Surround inhibition for improved contour detection by texture suppression. *Perception*, 33 (68c) 2004.

[50] PHILIPS, P.J., FLYNN, P.J., SCRUGGS, T., BOWYER, and K.W: Overview of the face recognition grand challenge. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, (2005).

[51] ROJAS, R.: *Neural Networks - A Systematic Introduction*. Berlin : Springer-Verlag, S. 502, 1996. ISBN 978-3540605058.

[52] ROSENBLATT, F.: *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.* 6, American Psychological Association, 1958.

[53] ROWLEY, H., BALUJA, S.,KANADE, T.: Neural Network-Based Face Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 1996. s. 203-208.

[54] SAMARIA, F., HARTER, A.: *The ORL Database of Faces*, AT&T Laboratories Cambridge University, *http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html*.

[55] SINGH, S. K., CHAUHAN, D. S., VATSA, M.: A Robust Skin Color Based Face Detection Algorithm, *Tamkang Journal of Science and Engineering*, 2003, Vol. 6.

[56] SCHNEIDERMAN, H., KANADE, T.: A statistical method for3D object detection applied to faces and cars. *International Conference on Computer and Pattern Recognition*, 1 (2000) 746-751.

[57] SOKOLOVA, M., JAPKOWICZ, N., SZPAKOWICZ, S.: Beyond Accuracy, F-score and ROC: a Family of Discriminant Measures for Performance Evaluation. *AI 2006: Advances in Artificial Inteligence.* Springer, s. 1015-1021, 2006. ISSN: 0302-9743.

[58] ŠOCHMAN, J., MATAS, J.: *AdaBoost,* Centre for Machine Perception Czech Technical University, Prague.

[59] ŠOCHMAN, J.: Adaboost. *Cvičení z RPZ* [online]. 2005.

[60] THÉVENAZ, P., BLU, T., UNSER, M., BANKMAN, I. N.: Image Interpolation and Resampling, *Handbook of Medical Imaging, Processing and Analysis*, s. 393-420, 2000.

[61] TOMKINS, S.S.: Affect theory. In Scherer, K.R., Ekman, P.(eds.), *Approaches to emotion*, Hillsdale, N.J.: Erlbaum, (1984) 163-196.

[62] TURK, M., PENTLAND, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience Volume 3, Number 1,* 1991.

[63] VIOLA, P., JONES, M.: *Robust Real-time Object Detection*, 2001. Vancouver, Canada.

[64] WEYRAUCH, B., HUANG, J., HEISEL, B.: Component-based Face Recognition with 3D Morphable Models, *First IEEE Workshop on Face Processing in Video*, Washington, D.C., 2004.

[65] YANG, M.: Detecting Faces in Images: A Survey. *IEEE Transactions on pattern analysis and machine intelligence*. January 2002, Vol. 24, 1.

[66] YANG, G., HUANG, T.S.: Human Face Detection in Complex Background. *Pattern Recognition,* 1994. Volume 27. s. 53-63.

[67] YOUNG, J. W.: *Head and Face Anthropometry of Adult U.S. Civilians*, U. S. Government Printing Office, 1993.

[68] YOW, K. C., CIPOLLA, R.: Feature-Based Human Face Detection, *Image and Vision Computing*, 1997, vol. 15.

Seznam vlastních prací

[69] MÍČA, I., PŘINOSIL, J., VLACH, J.: Framework for Support of Digital Signal Processing Education. In *Telecommunications and Signal Processing TSP - 2007.* Brno: VUT BRNO, 2007. s. 97-100. ISBN: 978-80-214-3445-5.

[70] PŘINOSIL, J., KROLIKOWSKI, M.: Využití detektoru Viola-Jones pro lokalizaci obličeje a očí v barevných obrazech. *Elektrorevue - Internetový časopis (http://www.elektrorevue.cz)*, 2007, roč. 2008, č. 04, s. 1-12. ISSN: 1213-1539.

[71] PŘINOSIL, J., MÍČA, I.: Supporting System for Digital Signal Processing. In *Research in Telecommunication Technology RTT - 2007.* Liptovský Ján, Slovenská republika: 2007. s. 341-344. ISBN: 978-80-8070-735-4.

[72] PŘINOSIL, J., SMÉKAL, Z., ESPOSITO, A.: Combining Features for Recognizing Emotional Facial Expressions in Static Images. *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction,* Springer, Berlin. 2008. s. 59-73. ISBN: 3540764410.

[73] PŘINOSIL, J., VLACH, J.: Human face localization in color images. In *Proceedings of the 12th IFIP International Conference, Prague.* 2007. s. 533-544. ISBN: 978-0-387-74158-1.

[74] PŘINOSIL, J., VLACH, J.: Face detection in image with complex background. *Mobile and Wireless Communication Networks,* Springer. 2007, č. 9, s. 533-544. ISSN: 1571-5736.

[75] PŘINOSIL, J., SCHIMMEL, J.: Supporting System for Education of Digital Audio Signals Processing on Digital Signal Processors. In *ICSES´06 International Conference on Signals and Electronic Systems.* Lodz, Poland: Technical University of Lodz, Institute of Circuit Theory, Metrology and Materials Science, 2006. s. 673-676. ISBN: 83-921172-5-5.

[76] SCHIMMEL, J., PŘINOSIL, J.: Surveillance System for Audio Broadcast. In *7th Nordic Signal Processing Symposium NORSIG 2006.* Reykjavik, Icelnad: 2006. s. 298-301. ISBN: 1-42244-0413-4.

[77] SCHIMMEL, J., PŘINOSIL, J.: Integration of DSP Plug-in System into Virtual Studio Technology. In *Proceedings of the International Conference on Research in Telecommunication Technology, RTT 2005.* Praha: 2005. s. 1-6. ISBN: 80-210-3699-0.

[78] SCHIMMEL, J., PŘINOSIL, J.: Digital Audio Signal Processing in DSP Using Plug-Ins. *GESTS International Transaction on Computer Science and Engineering,* 2005, roč. 2005, č. 12, s. 149-155. ISSN: 1738-6438.

[79] VLACH, J., PŘINOSIL, J.: Lokalizace obličeje v obraze s komplexním pozadím. *Elektrorevue - Internetový časopis (http://www.elektrorevue.cz)*, 2007, roč. 2007, č. 04, s. 1-12. ISSN: 1213-1539.

[80] VLACH, J., RAJMIC, P., PŘINOSIL, J., VYORAL, J., MÍČA, I.: Optimized discrete wavelet transform to real-time digital signal processing. *Mobile and Wireless Communication Networks,* Springer. 2007, č. 9, s. 514-520. ISSN: 1571-5736.

[81] VLACH, J., RAJMIC, P., PŘINOSIL, J.: New algorithm of discrete wavelet transform optimized to real-time digital signal processing. *In Proceedings of the 12th IFIP International Conference, Prague.* 2007. s. 514-520. ISBN: 978-0-387-74158-1.

Jiří Přinosil

Curriculum Vitae

Osobní data:

Adresa:	Fleischnerova 5, 63500, Brno
Telefon:	+420 723 946 654
E-mail:	prinosil@ feec.vutbr.cz
Datum narození:	12.8.1981
Národnost:	Česká
Stav:	svobodný

Vzdělání:

2000-2005 Magisterské studium – VUT v Brně, Fakulta elektrotechniky a komunikačních technologií, obor: Elektronika a sdělovací technika. Téma diplomové práce: Systém podpory zpracování audiosignálů.

2005-2008 Doktorské studium – VUT v Brně, Fakulta elektrotechniky a komunikačních technologií, obor Teleinformatika. Téma disertační práce: Analýza emocionálních stavů na základě obrazových předloh.

Další vzdělání:

2007 Tří měsíční zahraniční odborná stáž – *International Institute of Advanced Scientific Studies* (IIASS), Vietri sul mare, Italy. Náplň stáže: využití metod číslicového zpracování obrazů při analýze emocionálních stavů.

Ostatní:

Člen organizace IEEE (*Institute of Electrical and Electronics Engineers*) v roce 2008.

Vybrané publikace:

PŘINOSIL, J., SMÉKAL, Z., ESPOSITO, A.: Combining Features for Recognizing Emotional Facial Expressions in Static Images. *Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction,* Springer, Berlin. 2008. s. 59-73. ISBN: 3540764410.

PŘINOSIL, J., VLACH, J.: Face detection in image with complex background. *Mobile and Wireless Communication Networks,* Springer. 2007, č. 9, s. 533-544. ISSN: 1571-5736.

PŘINOSIL, J., SCHIMMEL, J. *Supporting System for Education of Digital Audio Signals Processing on Digital Signal Processors* In ICSES´06 International Conference on Signals and Electronic Systems. ICSES '06 International Conference on Signals and Electronic Systems. Lodz, Poland: Technical University of Lodz, Institute of Circuit Theory, Metrology and Materials Science, 2006, page 673 - 676, ISBN 83-921172-5-5.

Účast na projektech:

Projekt MŠMT OC-COST OC08057 2008: *Analýza a zvýraznění řečových a obrazových signálů ze šumu pro vzájemnou analýzu verbální a neverbální komunikace*. Ústav telekomunikací fakulta elektrotechniky a komunikačních technologií, VUT v Brně. SMÉKAL, Z.

Projekt FRVŠ G1 1730/2008: *Sofistikovaný systém pro analýzu a zpracování obrazových dat*. Ústav telekomunikací fakulta elektrotechniky a komunikačních technologií, VUT v Brně. MÍČA, I.

Projekt FRVŠ F1 1767/2008: *Inovace výuky předmětů zaměřených na zpracování zvukových signálů*. Ústav telekomunikací fakulta elektrotechniky a komunikačních technologií, VUT v Brně. SCHIMMEL, J.

Projekt FRVŠ G1 1023/2007: *Podpůrný systém pro zpracování číslicových signálů a jeho zavedení do výuky předmětu Číslicové zpracování signálů*. Ústav telekomunikací fakulta elektrotechniky a komunikačních technologií, VUT v Brně. PŘINOSIL, J.

Projekt FRVŠ F1 1469/2007: *Inovace laboratorních cvičení předmětu Číslicové filtry*. Ústav telekomunikací fakulta elektrotechniky a komunikačních technologií, VUT v Brně. SYSEL, P.

Projekt AV 1ET110540521 2005: *Výzkum nové generace infúzních pump s centrálním dispečinkem*. Ústav telekomunikací fakulta elektrotechniky a komunikačních technologií, VUT v Brně. ŠILHAVÝ, P.

Produkty:

SCHIMMEL, J.; PŘINOSIL, J.: Prototyp; *Audified DSP Solution for Motorola DSP56307EVM*. DISK Multimedia, s.r.o., Sokolská 13, 680 01 Boskovice.

V Brně, 22. srpna 2008.