VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

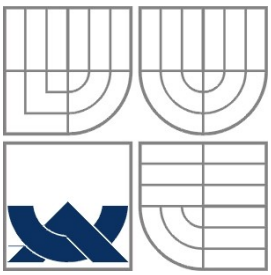
POLOAUTOMATICKÉ POŘÍZENÍ ROZSÁHLÉ DATABÁZE LIDSKÝCH OBLIČEJŮ

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

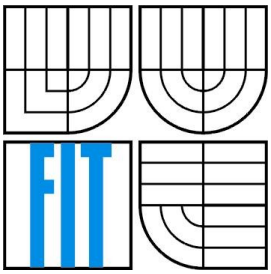
AUTOR PRÁCE
AUTHOR

Bc. MAREK MICHALÍK

BRNO 2011



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÝCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER SYSTEMS

POLOAUTOMATICKÉ POŘÍZENÍ ROZSÁHLÉ DATABÁZE LIDSKÝCH OBLIČEJŮ

SEMIAUTOMATIC COLLECTION OF LARGE DATABASE OF HUMAN FACES

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

Bc. MAREK MICHALÍK

VEDOUCÍ PRÁCE
SUPERVISOR

doc. Ing. ADAM HEROUT, Ph.D.

BRNO 2011

Abstrakt

Diplomová práce je zaměřena na metody získávání rozsáhlého počtu snímků lidských obličejů. Takto vzniklá databáze obličejů by měla sloužit jako datová sada pro detekci a rozpoznání lidských tváří pomocí strojového učení s učitelem. Práce se zabývá základními principy strojového učení s učitelem a dostupnými datovými sadami pro toto učení. Součástí práce je návrh technik a implementace algoritmů vhodných pro extrahování potencionálních tváří z videa a zpracování pomocí uživatelského rozhraní pro poloautomatickou akceptaci a anotaci nalezených snímků.

Abstract

The project is focused on methods of obtaining large number of images of human faces. Such database should then serve as a set of data for face detection and recognition by the means of supervised machine learning. The work deals with the basic principles of supervised machine learning and available data sets for this procedure. Project contains proposals of techniques and implementation of algorithms suitable for acquiring images from video and a concept of user interface for semi-automatic acceptation and annotation of located images.

Klíčová slova

zpracování obrazu, detekce obličeje, detekce očí, detekce úst, databáze obličejů, viola-jones, OpenCV

Keywords

Image processing, face detection, eye detection, mouth detection, face database, viola-jones, OpenCV

Citace

Marek Michalík: Poloautomatické pořízení rozsáhlé databáze lidských obličejů, diplomová práce, Brno, FIT VUT v Brně, 2011

Poloautomatické pořízení rozsáhlé databáze lidských obličejů

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením doc. Ing. Adama Herouta, Ph.D.
Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Marek Michalík
25. 5. 2011

Poděkování

Na tomto místě bych velmi rád poděkoval doc. Ing. Adamu Heroutovi, Ph.D. za odborné vedení, cenné rady, připomínky a konzultace, které mi poskytoval během řešení této práce.

Obsah

1 Úvod

Při detekci a identifikaci lidských obličejů se v dnešní době používají především metody umělé inteligence založené na strojovém učení. Pro správnou funkčnost těchto algoritmů je jim třeba nejprve předkládat trénovací data a následně zjistit kvalitu učení na testovacích datech. Součástí obou těchto skupin dat je i množina lidských obličejů, čím větší a rozmanitější bude databáze obličejů, tím je možné kvalitnější trénování. K samotnému roztřídění obličejů a ostatních dat je pro potřeby strojového učení vhodná také anotace důležitých rysů na obličeji (oči, ústa, nos, uši atd.). Získání takto rozsáhlé databáze lidských obličejů s anotací některých rysů bez pomoci automatizovaného postupu je ovšem časově (tím i finančně) velmi náročné.

Tato práce má za úkol přiblížit strojové učení s učitelem, popsat dostupné datové sady vhodné pro toto učení a především nastínit způsob pro vytvoření vlastní rozsáhlé databáze lidských obličejů. Pro vytvoření této databáze je nutné získat vhodné snímky, následně je maximálně automaticky zpracovat a předložit uživateli k posouzení (popř. úpravu anotovaných částí).

Druhá kapitola práce se zabývá strojovým učením a jeho využitím při detekci a rozpoznání lidské tváře, včetně dostupných trénovacích dat pro toto učení. Ve třetí kapitole jsou nastíněny některé způsoby získání snímků lidských tváří z obrazových dat. Zpracování těchto dat pro potřeby rozsáhlé databáze lidských obličejů rozebírá čtvrtá kapitola. Slouží jako informace o použitých metodách při tvorbě aplikace generující databázi. Pátá kapitola se zabývá především způsobem implementace tohoto programu, který by měl umožnit vytvořit rozsáhlou databázi lidských obličejů s co nejmenším podílem lidské práce. V šesté kapitole je čtenář seznámen s výsledky praktické části práce, jaké data se podařilo získat a jakou měrou při jejich pořízení musel přispívat uživatel.

2 Strojové učení

Pro detekci lidských obličejů se často používají metody strojového učení. Myšlenkou této podoblasti umělé inteligence je možnost postupného zdokonalování algoritmu na základě získaných zkušeností. V závislosti na čase systém založený na strojovém učení zvyšuje pravděpodobnost správného rozhodování [1]. V případě lidských obličejů je možné vytvářet různé systémy na samotné rozpoznání tváří, natočení hlavy, rozpoznání výrazu obličeje apod. Podle přístupu k získávání zkušeností (učení) se dělí metody strojového učení do dvou hlavních skupin – strojové učení bez učitele a strojové učení s učitelem. První z těchto metod se pro rozpoznávání obličejů příliš nepoužívá. Učení bez učitele [2] zachycuje pravidelnosti ve vstupních vektorech bez přijímání jakékoliv další informace (nemá informaci o požadovaných hodnotách na výstupu). Vstupní vektory jsou brány jako náhodné proměnné, ve kterých se pomocí pravděpodobnostních metod nebo pomocí extrakce charakteristických vlastností získávají statistické zákonitosti. Pokud je reálná možnost získat dostatečné množství vstupních hodnot, ke kterým lze jednoznačně přiřadit odpovídající výstupní hodnoty, je vhodnější použít učení s učitelem.

2.1 Učení s učitelem

Strojové učení s učitelem [3] využívá trénovacích dat tvořených vstupním vektorem rysů a požadovaným výstupem. Tímto způsobem si systém vytváří klasifikátor, který předpovídá výstupní hodnotu funkce pro každý platný vstupní vektor. Algoritmus se učením na trénovacích datech adaptuje na rozhodování určitého obecného problému. Může potom s určitou pravděpodobností rozřazovat data stejného typu, podle kterých byl trénován. Trénovacími daty může být skupina snímků, u kterých je označeno zda se jedná o obličej nebo ne. Oběma těmto skupinám je přiřazena určitá hodnota (např. 1 a -1), podle které si systém upraví svou vnitřní strukturu rozhodování.

Pro zjištění kvality natrénovaného systému se používá testovací sada dat, struktura těchto dat by měla být stejná s trénovacími daty. Často jsou testovací i trénovací data na počátku pohromadě a rozdělí se náhodně na dvě skupiny, jedna skupina je použita k trénování a druhá k následnému testování. Pokud jsou trénovací data málo rozmanitá nebo pokud je trénování příliš intenzivní, může u některých systémů (neuronové sítě) dojít k přetrénování (přeučení). V takovém případě systém ztrácí schopnost generalizace, jeho rozpoznávací schopnost je zaměřena na konkrétní data z trénovací množiny a selhává na testovací množině dat.

2.2 Datové sady

Pro trénování systémů založených na strojovém učení s učitelem je nutné získat sadu trénovacích i testovacích dat. V případě rozpoznávání lidských obličejů je tedy potřeba sada lidských tváří. Existuje celá řada zdrojů nabízející sady těchto dat (přehled viz Tabulka 1, Tabulka 2). Jejich pořízením se velmi často zabývají vědecké týmy na univerzitách po celém světě, a tak je řada z nich volně dostupná pro vědecké použití, u některých je nutná registrace či žádost o zpřístupnění, některé jsou plně komerční a jejich získání je možné pouze za poplatek. Problémem těchto dat může být jejich úzká zaměřenost, databáze většinou vznikaly v rámci určitého projektu, kterému se podřídilo

jejich pořízení. Řada těchto databází obsahuje snímky podobné dokladovým fotografiím, tedy čelní pohled hlavy na homogenním pozadí, popř. natočení z profilu. Některé datové sady (např. MIT-CBCL Face Recognition Database) již přímo obsahují připravená data pro učení neuronových sítí, vedle původních snímků tváří dávají k dispozici i vyextrahované oblasti obličeje převedené na velmi malé rozlišení (19 x 19 px) ve stupních šedi (viz Obr. 1). K dispozici jsou trénovací i testovací data rozdělená na obličeje a ostatní objekty. Data z těchto databází už dnes nemusí být tolik žádaná, protože na samotnou lokalizaci obličeje je již dostupná celá řada rozpoznávacích algoritmů. Novější databáze lidských obličejů se proto zaměřují na specifičtější snímky - různé výrazy stejných lidí, osvětlení obličeje z různých úhlů, změny obličeje (vousy, brýle apod.). V těchto případech si i rozsáhlá databáze většinou vystačí s poměrně malým množstvím zaznamenaných osob. Ty jsou zaznamenávány opakovaně v různých polohách a situacích. Jinou variantou je snímání stejných osob ve větším časovém intervalu (rok a více), u tohoto postupu je poměrně obtížné udržet jednotné prostředí na všech snímcích. Zvláštní skupinou jsou pak databáze zaměřené na rozpoznávání hlásek nebo databáze využívající 3D snímků obličejů. Nevýhodou všech těchto datových sad je jejich laboratorní pořízení. Při trénování systému na těchto datech hrozí v praktickém nasazení systému riziko horší pravděpodobnosti rozpoznávání. Další nevýhodou, která může ovlivnit trénování je přílišná homogennost snímaných národností a ras. Jednotlivé univerzity se zaměřují na snímání lidí ze svého prostředí, vznikají tak databáze např. výhradně indických, čínských aj. osob (bohužel ani u dat z evropský databází nejsou národnosti rozmanitější). Další ukázky datových sad jsou zobrazeny na Obr. 2 a Obr. 3.

Snímky pro databázi se většinou pořizují ručně, kdy jsou vybrané osoby jednotlivě fotografované. Určitá snaha o automatizaci procesu se provádí u snímků jedné osoby z více pohledů. Některé univerzity sestrojily konstrukce pro snímání hlavy z více míst několika kamerami, nebo otáčení snímané osoby před kamerou.

Některé databáze obličejů jsou doplněny o anotaci základních rysů tváře, nejčastěji se lokalizuje pozice očí a úst, dále nosu popř. uší. Méně často se snímky doplňují informacemi o věku snímané osoby a o jejím vzhledu, v tomto případě se uvádí zda má osoba vousy, brýle apod. Počet anotovaných bodů opět záleží na způsobu použití trénovaného systému. Zatímco u lokalizace obličeje postačí několik základních bodů (pozice středů očí, koutky úst atd.), pro složité biometrické systémy určené k autentizaci osob je zapotřebí mnohonásobně více bodů v obličeji (tvar lícních kostí, očí a nadočnicových oblouků, celkový obrys obličeje apod.). Specifikem některých databází (Iranian Face Database) mohou být informace o vadách kůže na obličeji nebo podrobnější záznamy o snímané osobě (např. vykonávané zaměstnání). Anotování tváře se v drtivé většině případů provádí ručně a je tak dostupné především u méně rozsáhlých datových sad.

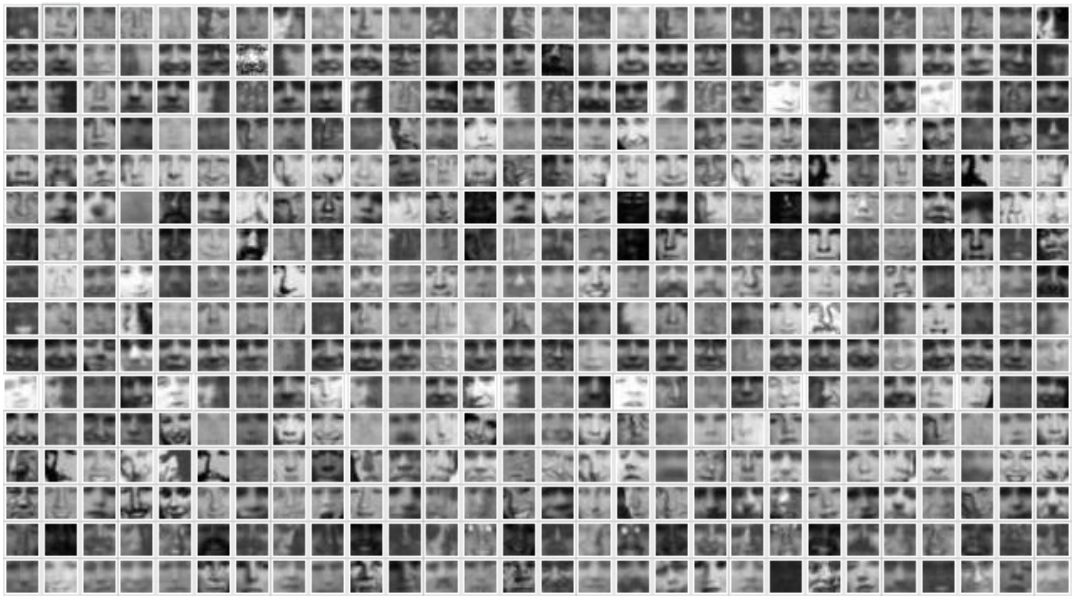
Pro trénování nových systémů je potřeba co nejrozmanitější databáze tváří. V databázi by se měly vyskytovat nejenom zcela jednoznačné snímky obličejů (čelní pohled), ale i snímky s různými polohami obličejů, výrazů tváře i různým osvětlením. Vhodnými prvky takové databáze by měly být i částečně zakryté a neúplné obličeje, popř. další netypické snímky tváří, na které nejsou trénované běžné rozpoznávací systémy. Aby měla vzniklá databáze nějaký přínos, není vhodné při pořizování snímků obličejů používat detektory natrénované na současných datových sadách, které dokáží detekovat především jasně zřetelné obličeje a můžou vynechávat netypické polohy dalších obličejů. Datové sady zaměřené na tuto problematiku jsou ojedinělé a většinou se spokojí pouze s dioptrickými brýlemi nebo několika grimasy.

Zdroj databáze	Poč.	Roz.	Anot.	Dostupnost
The Color FERET Database, USA http://www.face.nist.gov/colorferet/	14126 (1199)	?	žádná	• stáhnutelné po e-mailové žádosti • dříve rozesílaná na DVD
• vytvořena v letech 1993 - 1996 • stupně šedi i barevná verze			• unifikované prostředí (dokladové foto)	
SCface - Surveillance Cameras Face Database http://www.scface.org/	4160 (130)	až 2048 x 3072	oči ústa nos	• přístupná pro vědecké účely (na žádost zaměstnance, ne studenta)
• různorodé vnitřní prostředí • různě kvalitní kamery • barevné a IR snímky			• 115 mužů, 15 žen, běloši 20 – 75 let • informace o vzhledu (stáří, vousy, brýle) • součástí jsou snímky 130 předmětů	
Multi-PIE http://www.multipie.org/	750000 (337)	?	žádná	• placená
• rozsáhlá databáze menšího počtu lidí • různé výrazy			• barevné snímky • foceny osoby v křesle z 15 míst při 20 různých podmínkách osvětlení	
The Yale Face Database http://cvc.yale.edu/projects/yalefaces/yalefaces.html	165 (15)	320 x 243	žádná	• volně dostupná pro vědecké účely (nutné vyplnit formulář)
• 11 snímků každé osoby za různého osvětlení, výrazu, s brýlemi • stupně šedi (formát GIF)				
The Yale Face Database B http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html	5760 (10)	640 x 480	oči ústa	• volně dostupná
• stupně šedi • nehomogenní pozadí			• snímky s jednobodového zdroje světla (foceno pomocí plošiny, 9 póz se 64 blesky)	
PIE Database, CMU http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261	41368 (68)	?	?	• kontakt e-mailem
• u každé osoby 13 různých póz, 4 výrazy, 43 různých světelných podmínek • barevné snímky				
AT&T "The Database of Faces" http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html	400 (40)	92 x 118	žádná	• Volně dostupná
• stupně šedi • každá osoba snímána 10x při různém osvětlení a s různým výrazem				
Cohn-Kanade AU Coded Facial Expression Database http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html	486 sekvencí (97)	640 x 480	žádná	• Podepsání dohody
• stupně šedi • video záznam převedený na obrázky s uvedeným pořadím snímků • studenti (18-30 let), 65% ženy, 15% Afroameričani, 3% Asiati a Hispánci				
MIT-CBCL Face Recognition Database http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html	60 (10)	až 2048 x 1536	žádná	• volně dostupná
• dvě trénovací sady, kromě klasických fotografií (barevných) tváří 10 lidí z různých pohledů je k dispozici 3240 vytvořených 3D obličejů (ve stupních šedi) těchto lidí s různým osvětlením a natočením tváře, příprava na strojové učení • 2000 trénovacích snímků (osoby s různým natočením hlavy a očí)				
Poč.= Počet snímků v databázi (počet osob), Roz. = Rozlišení [px], Anot.= Anotované rysy tváře				

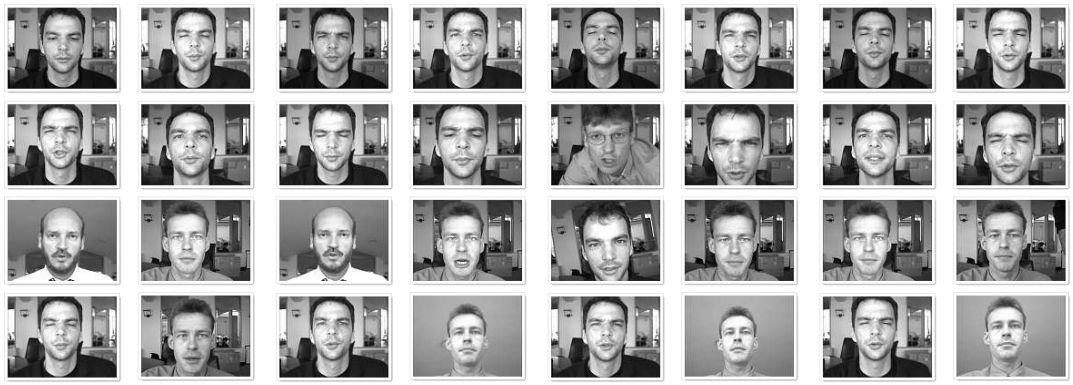
Tabulka 1: Přehled některých databází obličejů

Zdroj databáze	Poč.	Roz.	Anot.	Dostupnost
Iranian Face Database http://kiau.ac.ir/bastanfard/IFDB_index.htm	3600 (616)	640 x 480	jiné	• na požádání vědeckým osobám (publikovaný článek) • podepsání dohody
• rozsáhlá databáze obyvatel středního východu • osoby různých věkových kategorií (2-85 let) • unifikované prostředí (dokladové foto) osvětlené denním světlem			• u každé osoby čelní pohled a dva profily, popř. různé výrazy nebo snímky s brýlemi • informace o zaměstnání, kožních vadách apod. • 487 mužů, 129 žen	
The Hong Kong Polytechnic University (PolyU) NIR Face Database http://www4.comp.polyu.edu.hk/~biometrics/polyudb_face.htm	34000 (335)	?	pohlaví	• získání hesla po schválení žádosti • pro výzkumnou činnost
• IR snímky • výhradně asiaté			• unifikovaná pozice snímané osoby • stejné osoby snímané v časovém intervalu	
The Hong Kong Polytechnic University (PolyU) Hyperspectral Face Database (HSFD) http://www4.comp.polyu.edu.hk/~biometrics/hyper_face.htm	300 (25)	220 x 180	pohlaví změny vzhledu oči	• získání hesla po schválení žádosti • pro výzkumnou činnost
• unifikované prostředí (osvětlení, vzdálenost osoby) • snímaná osoba postupně otáčena o 360° • černobílé snímky (čelní a dva profilové)			• 3D data lidských hav • 17 mužů, 8 žen • snímání stejných osob v časovém intervalu (2007 - 2008)	
Indian Face Database http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase	560 (40)	640 x 480	pohlaví	• volně dostupná pro vědecké účely • vhodné informovat autory o užívání databáze
• černobílé snímky osob indické národnosti			• čelní pohledy (7 snímků pohledů, 7 snímků výrazů)	
Georgia Tech Face Database http://www.anefian.com/research/face_reco.htm	750 (50)	640 x 480	pozice obličeje	• volně dostupná
• různé pozice hlavy (čelní pohled, nakloněná hlava) • změny osvětlení a výrazy tváře			• výřezy obličejů (rozl. kolem 150x150 px) • ručně anotovaná pozice hlavy	
Caltech Faces http://www.vision.caltech.edu/html-files/archive.html	450 (27)	896 x 592	žádná	• volně dostupná
• pouze mladší osoby			• různé osvětlení, výraz a pozadí	
The UCD Colour Face Image Database for Face Detection http://ee.ucd.ie/~prag/	720	různé	žádná	• dostupná pro vědecké účely po ověření registrace
• barevné snímky z nejrůznějších zdrojů (skenované, stáhnuté z internetu, digitální fotografie)			• původní snímky a ručně segmentované obličeje	
MoBio http://www.idiap.ch/dataset/mobio	1824 (152)	640 x 480	žádná	• potvrzení EULA
• video nahrávky lidských tváří z mobilního telefonu • 100 mužů, 12 žen			• osoby snímané při odpovědi na otázky (anglicky) • 12 sezení s 32 otázkami	
Put Face Database http://biometrics.cie.put.poznan.pl/	9971 (100)	2 048 x 1 536	oči ústa nos	• emailový kontakt • volně dostupná pro výzkumné účely
• unifikované pozadí (částečně i osvětlení) • ruční anotace dat			• různé pozice hlavy s neutrálním výrazem (otáčení hlavy v přímém pohledu, se sklopenou a zvednutou hlavou, přikyvování) a s uvolněným výrazem či grimasou	
Poč.= Počet snímků v databázi (počet osob), Roz. = Rozlišení [px], Anot.= Anotované rysy tváře				

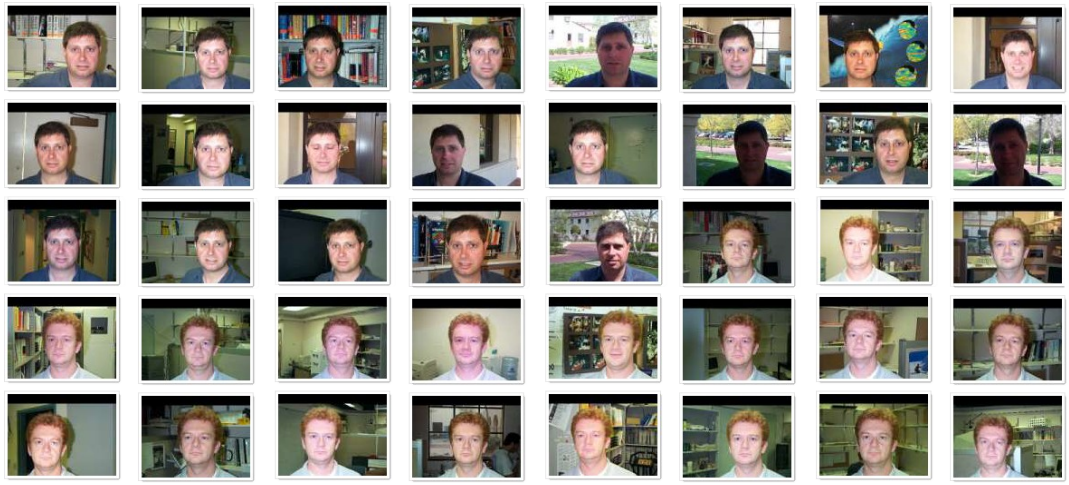
Tabulka 2: Přehled některých databází obličejů



Obr. 1: MIT-CBCL Face Recognition Database



Obr. 2: BioID Face - DB - HumanScan AG, Švýcarsko



Obr. 3: Caltech Faces

3 Detekce obličeje

Datové sady pro strojové učení s učitelem obsahují pouze snímky hlavy člověka, popř. jenom samotné tváře. K pořízení těchto dat je tedy nutné detekovat v obraze s různorodým okolím lidský obličej. Z důvodu větší rozmanitosti nasbíraných dat je vhodnější použít jednodušší detektory, které rozpoznají tvář v obraze s menší pravděpodobností. Vznikne tak větší množství snímků na kterých žádná tvář není, ale dokáží zachytit i méně zřetelné tváře, které by natrénovaný detektor nedokázal vyhodnotit jako lidský obličej. Uživatel pak musí projít snímky a vytřídit ty, na kterých není žádný obličej.

Pro pořízení takto specifických snímků se příliš nehodí znalostní metody [9], které jsou založené na popisu základních rysů tváře (oči, nos, ústa) a jejich vztahy. Problémem těchto metod může být detekování obličejů při různých výrazech tváře (pro každý výraz by se musela specifikovat vlastní pravidla).

Metody založené na invariantních rysech často využívají hranových detektorů ke zvýraznění obličejových rysů, podle kterých detekují obličej (často podle očí a obočí) [7]. Při špatném osvětlení může například stín vytvořit velké množství ostrých hran a znemožnit tak správnou detekci obličeje. Do skupiny těchto metod spadá i detekce obličeje podle barvy lidské kůže. Tato metoda není náchylná na výrazy tváře nebo natočení obličeje. Barva lidské kůže je natolik specifická, že není problém její jednoduché detekování v různém prostředí. Více tuto metodu popisuje kapitola 3.2.

Metody založené na porovnávání se šablonou [6] mají výhodu jednoduché implementace. Využívají standardní vzor tváře (ručně vytvořená šablona, popř. skupina šablon), který se porovnává s daným vstupním obrazem a nezávisle na sobě se počítají korelační hodnoty pro obrys tváře, očí, nosu a úst. Detekce tváře záleží na míře shody vstupního obrazu se vzorovou šablonou. Tento způsob ale není efektivní pro různé rozměry obličejů, jejich natočení i tvar. Částečně lze tento problém kompenzovat několika vzorovými šablonami, přesto tato metoda není vhodná k detekci částečně skrytých a netypicky umístěných obličejů.

Metody založené na vzhledu [5] využívají pro detekci tváře strojového učení (viz kap. 2), pomocí kterého nachází charakteristiky rozlišující lidskou tvář od okolí. Tyto metody dobře detekují obličeje podobného rázu, jaký byl v jejich trénovací množině. Jelikož je účelem této práce vytvořit podobnou množinu dat, není vhodné používat dobře natrénovaných systémů na konkrétní podobu tváří. Lze použít algoritmů založených na propojených jednoduchých klasifikátorech, u kterých nehrozí přetrénování a dokáží detekovat obličeje více obecněji (např. metody založené na AdaBoost - Viola-Jones).

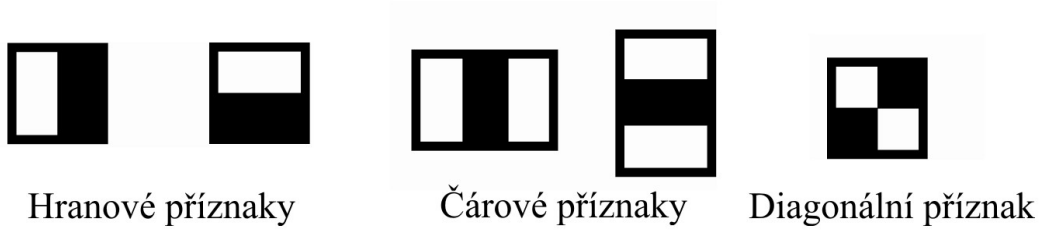
3.1 Viola – Jones

Detektor Viola – Jones je obdobou algoritmu AdaBoost vhodný pro detekci obličeje. Nejlépe funguje při detekci objektů nepodléhajících velkým změnám [4]. U obličeje tak může kvalitně rozpoznávat např. jenom čelní pohledy, na profilové a další pohledy je vhodné natrénovat nový detektor, popř. zahrnout do trénovací množiny snímky tváře v různých polohách.

Obraz není zpracováván jako celek, ale pomocí posuvného okna jsou procházeny jednotlivé oblasti (okno mění polohu i velikost). Pokud by se plně zpracovávalo každé okno byla by detekce

výpočetně náročná. Proto se používá kaskáda klasifikátorů, které postupně odmítají jednotlivé oblasti bez hledaného objektu. Kaskáda zamítne většinu oblastí (70 - 80%) už v prvním nebo druhém uzlu. Hledané objekty by ale měla propouštět stále dál. Teprve projdou-li celou kaskádou jsou přijaté. Jednotlivé uzly kaskády musí mít maximální pravděpodobnost správného přijetí (v praxi až 99,9%). U oblastí neobsahující hledaný objekt nemusí být klasifikátory zdaleka tak přesné a můžou propouštět až 50% oblastí, které hledaný objekt neobsahují. Při dostatečném množství klasifikátorů těchto vlastností se může míra správně přijatých oblastí udržet až na 98%, zatímco míra chybně přijatých klesne na zanedbatelné množství. Sada slabých klasifikátorů tak vytváří silný a rychlý algoritmus detekce.

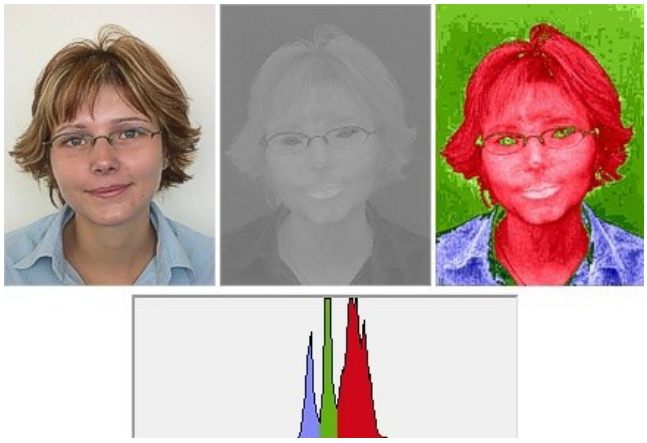
Klasifikátory používají příznaky podobné tzv. Haarovým vlnkám [4]. K urychlení výpočtu se využívá převedení vstupu na integrální obraz. Nemusí se pak počítat suma pixelů pro každý příznak. Výhodou této kombinace je rychlá odezva jednotlivých klasifikátorů, čas odezvy je navíc konstantní, i když se mění velikost použitého příznaku. Nevýhodou integrálního obrazu můžou být nároky na ukládání dat u obrazů ve velkém rozlišení.



Obr. 4: Příznaky podobné Haarově vlnce

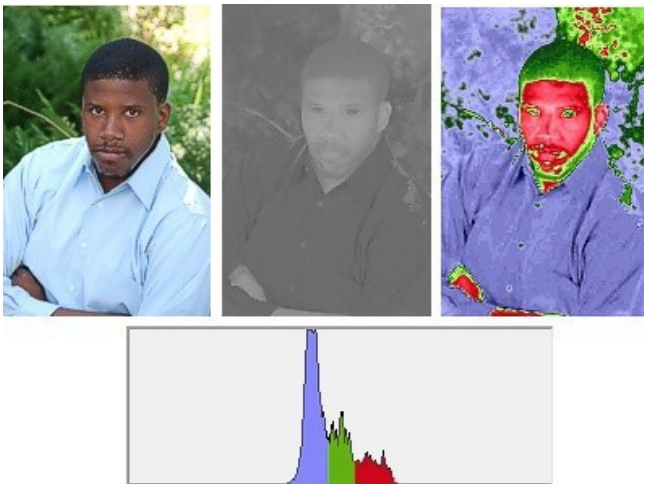
3.2 Hledání pomocí barvy kůže

Metody detekce obličeje založené na barvě lidské kůže využívají podobnosti této vlastnosti u velkého množství lidí. Metody využívají převodu obrazu do vhodného barevného modelu. U některých složek těchto modelů je pak určen interval ve kterém leží barva kůže. V praxi tento postup může úspěšně fungovat se stejným intervalem výběru u lidí rozličných barev pleti. Vhodným barevným modelem pro detekci lidské kůže je YC_bC_r (na Obr. 5 a Obr. 6 ukázka využití C_r kanálu při detekci lidské kůže u různých pletí), výhodou tohoto modelu je i snadný převod z používaného RGB, lze ale použít i dalších modelů např. HLS či CIECAM [7].



Obr. 5: Histogram C_r kanálu u osoby se světlou pletí

- *nahoře barevná fotografie, její C_r kanál a zvýraznění částí podle histogramu*
- *dole histogram C_r kanálu s barevným vyznačením úrovní*



Obr. 6: Histogram C_r kanálu u osoby s tmavou pletí a složitým pozadím

- *nahoře barevná fotografie, její C_r kanál a zvýraznění částí podle histogramu*
- *dole histogram C_r kanálu s barevným vyznačením úrovní*

3.2.1 Barevný model YC_bC_r

Tento barevný prostor je převážně využíván v oblasti digitálního videa a to z důvodu jeho kompresních vlastností (využívá se i pro formát JPEG). Hlavní výhodou je to, že informace o barvě je nesena pouze dvěma komponentami C_b a C_r. Komponenta C_b určuje rozdíl mezi modrou složkou a referenční hodnotou, komponenta C_r pak obdobně určuje rozdíl mezi červenou složkou a referenční hodnotou [8]. Komponenta Y obsahuje informaci o luminanci (luminance udává sílu jasu). Jasová

složka je tedy reprezentována pouze jedinou komponentou Y. Převod z nejběžnějšího modelu RGB je založen na jednoduchém přepočtu složek:

$$Y = 0{,}299 \cdot R + 0{,}587 \cdot G + 0{,}114 \cdot B \tag{3.1}$$

$$Cb = -0{,}1687 \cdot R - 0{,}3313 \cdot G + 0{,}5 \cdot B + 128 \tag{3.2}$$

$$Cr = 0{,}5 \cdot R - 0{,}4187 \cdot G - 0{,}0813 \cdot B + 128 \tag{3.3}$$

Obr. 7: Rozložení snímku obličeje na složky RGB (nahoře) a YC_bC_r

V barvě lidské kůže je pochopitelně malé zastoupení modré barvy, proto jsou při zobrazení C_b kanálu oblasti s obličeji hodně tmavé. Naopak červená barva je zastoupena výrazně a tak v zobrazení C_r kanálu obličeje tvoří jasnější oblasti. Intervalem hodnot z obou kanálů (C_b i C_r) lze automaticky vybrat oblasti ve kterých se pravděpodobně nachází barva lidské kůže. Interval musí být zvolen vhodně, aby eliminoval nejen místa s malým zastoupením červené barvy, ale naopak i místa s příliš velkým podílem červené (červené oblečení). Zároveň ale musí zohlednit co největší škálu barev pleti.

4 Zpracování snímků

Ze získaných snímků nelze přímo sestavit databázi obličejů; v databázi se nesmí vyskytovat falešná data, která by mohla vést k chybnému strojovému učení. Bohužel není možné se spolehnout na plně automatické třídění chybných snímků z velkého množství zaznamenaných dat. Konečné rozhodnutí o tom, zda se jedná o lidský obličej, stejně jako anotace důležitých bodů obličeje, bude na uživateli. Lidskou práci ale může usnadnit automatické vyřazení snímků, které sice našly jedny detektory, ale neodpovídají kriteriím ostatních detektorů. Např. mají tvar i výraz obličeje, ale neobsahují barvu lidské kůže (může se jednat o kreslené postavy, malované obrazy nebo tvary připomínající obličej). K filtrování těchto snímků lze opět použít metody popsané v kapitole 3.2. Dalším urychlením práce obsluhy je automatické nalezení důležitých rysů obličeje (umístění očí, rtů, nosu apod.), aplikací dalších detektorů na rozeznávání objektů v lidské tváři. Mnohým detektorům je vhodné upravit vstupní data pro kvalitnější rozhodování o umístění jednotlivých částí.

4.1 Předzpracování obrazu

V mnoha případech získaných snímků je vhodné před detekcí částí obličeje upravit vlastnosti obrazu a tím zvýšit pravděpodobnost úspěšné lokalizace hledané oblasti v obraze. V situaci, kdy jsou vstupní data pro získání snímků volena co nejkvalitnější, není předpokládána nutnost zbavovat obraz chyb a šumu. V úvahu přichází spíše úprava kontrastu a jasu, popř. doostření a zvýraznění hran. Tyto úpravy ovšem nepřinášejí do získaného snímku žádnou novou informaci (naopak některé ztrácejí), proto je nutné neprovádět úpravy snímků v databázi, ale jen u aktuálně zpracovávaného snímku před detekcí rysů.

4.1.1 Úprava histogramu

Barevný histogram je reprezentací rozložení barev v obrazu; vyjadřuje poměrné zastoupení počtu pixelů každého z daných barevných rozsahů. Nejčastěji je prezentován jako graf s osou x rozdělenou na hodnoty od 0 do 255, na ose y je pak četnost jednotlivých odstínů v obraze (počet pixelů stejné barvy v jednom sloupci). U správně exponovaného snímku by měly být zastoupeny všechny odstíny a rozdíl v zastoupení mezi okolními odstíny by neměl být velký. Normalizací histogramu lze roztáhnout hodnoty odstínů po délce celého intervalu [10].

$$q = \tau(p) = \frac{(p - p_0)\cdot(q_k - q_0)}{(p_k - p_0)} + q_0 \qquad (4.1)$$

kde:

$\langle p_0, p_k \rangle$ - původní jasová stupnice histogramu, kde $p = \langle p_0, p_k \rangle$,
$\langle q_0, q_k \rangle$ - nová jasová stupnice histogramu.

Ekvalizace histogramu umožní rozložení intenzit v obraze v co nejširším rozmezí s přibližně stejnou četností. U obrazů s vysokým kontrastem umožňuje ekvalizace zvýraznit těžko rozpoznatelné nízko kontrastní detaily [10].

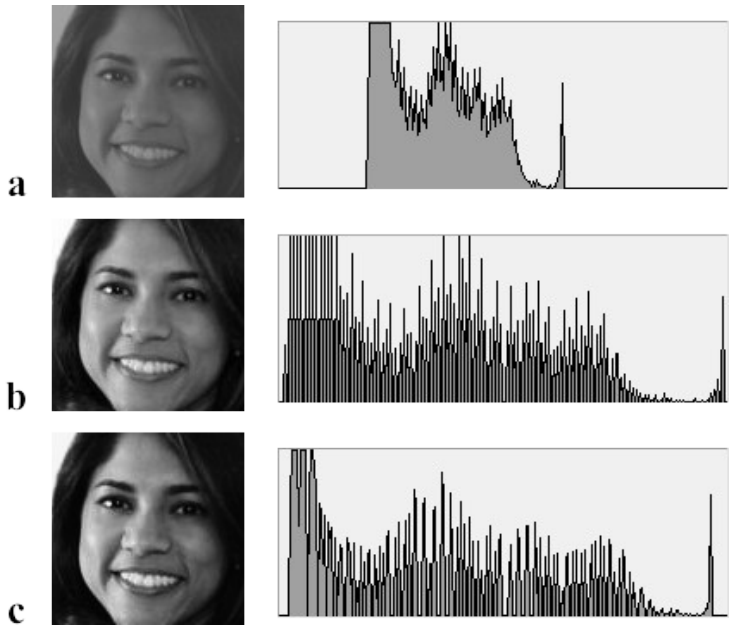
$$q = \tau(p) = \frac{q_k - q_0}{X \cdot Y} \cdot \sum_{i=p_0}^{p} H(i) + q_0 \qquad (4.2)$$

kde:

$\langle p_0, p_k \rangle$ - původní jasová stupnice histogramu, kde $p = \langle p_0, p_k \rangle$,

X, Y - rozměry obrázku

$\langle q_0, q_k \rangle$ - nová jasová stupnice histogramu.



Obr. 8: Zpracování histogramu

a - původní snímek

b - ekvalizace histogramu

c - normalizace histogramu

4.2 Detekce očí

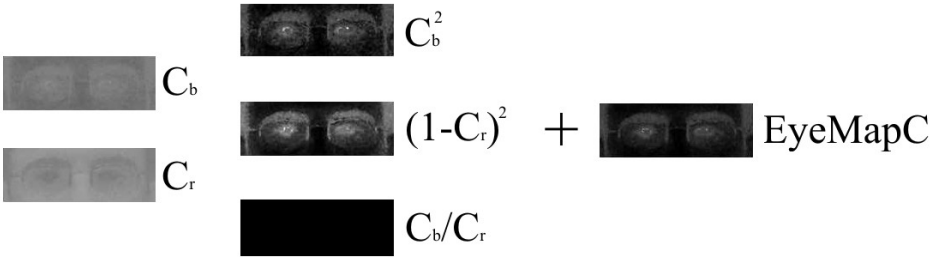
Automatická lokalizace očí opět využívá barevného modelu YC_bC_r, pro zpracování se využívají všechny tři složky modelu. Z jasového kanálu Y se pomocí šedotónové eroze a dilatace vytvoří mapa EyeMapL, kanály nesoucí informaci o barvě se zpracují do mapy EyeMapC [12].

$$EyeMapC = \frac{1}{3}((C_b)^2 + (255 - C_r)^2 + (\frac{C_b}{C_r})) \tag{4.3}$$

kde:

C_b - barevná složka modelu YC_bC_r nesoucí informaci o modré barvě

C_r - barevná složka modelu YC_bC_r nesoucí informaci o červené barvě



Obr. 9: Vytvoření mapy EyeMapC

$$EyeMapL = \frac{Y(x,y) \oplus g(x,y)}{Y(x,y) \otimes g(x,y) + 1} \tag{4.4}$$

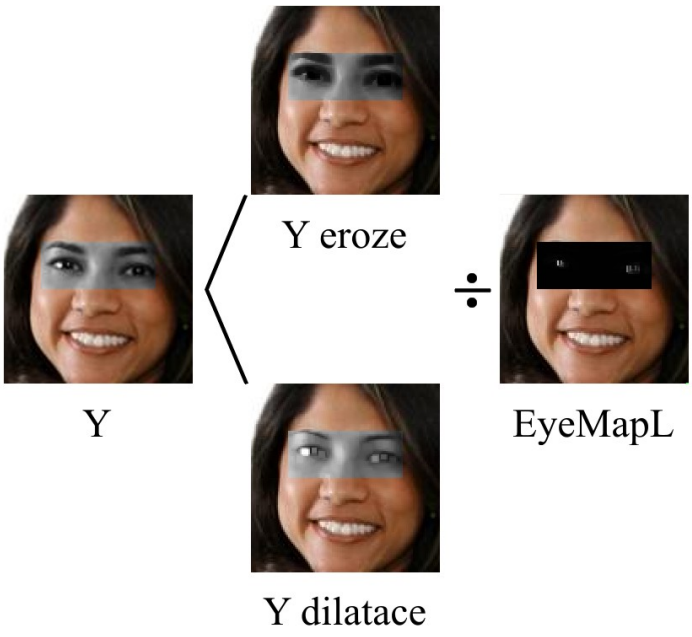
kde:

$\oplus$ - šedotónová dilatace

$\otimes$ - šedotónová eroze

$Y(x,y)$ - obrazová funkce, na kterou jsou morfologické operace aplikované

$g(x,y)$ - strukturální element



Obr. 10: Vytvoření mapy EyeMapL

Kombinací obou vytvořených map vzniká výsledná mapa EyeMap, která má hodnoty jednotlivých pixelů normalizované na hodnoty v intervalu <0.0;1.0>.

$$EyeMap = EyeMapC \cdot EyEMapL \tag{4.5}$$

Následným prahováním s hodnotou prahu získanou analýzou histogramu se mapa EyeMap převede na binární obraz. Pro zvýraznění příznaků očí je aplikována ještě dvojnásobná binární dilatace.

4.2.1 Šedotónová morfologie

Morfologie vychází z podstaty úpravy obrazu pomocí bodových množin [10]. Morfologická transformace je dána relací mezi obrazem (reprezentovaným bodovou množinou) a strukturním elementem (menší bodová množina). Tyto transformace se aplikují pomocí systematického posunu strukturního elementu po obraze a výpočtem příslušné operace [11].

$$H(i,j) \in R \qquad pro \qquad (i,j) \in E^2 \tag{4.6}$$

kde:

H	- strukturní element
R	- množina reálných čísel
E^2	- dvojrozměrný euklidovský prostor

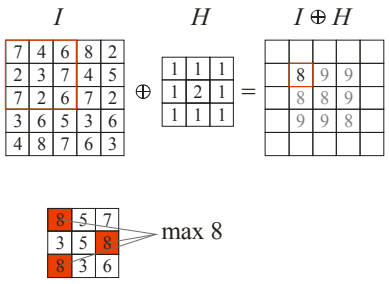
Dilatace

Šedotónová dilatace je definována jako maximum z hodnot získaných součtem hodnot strukturního elementu H s příslušnými hodnotami aktuálního okna v obraze I, je definována vztahem [11]:

$$(I+H)(u,v) = \max_{(i,j)\in H} I(u+i, v+j) \oplus H(i,j) \tag{4.7}$$

kde:

$\oplus$	- šedotónová dilatace
I	- aktuální podobraz
H	- strukturní element
max	- aritmetická funkce pro výpočet maxima



Obr. 11: Aplikace šedotónové dilatace

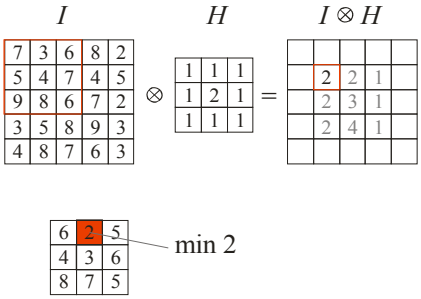
Eroze

Šedotónová eroze je definována jako minimum z hodnot získaných rozdílem příslušných hodnot aktuálního okna v obraze *I* a hodnot strukturního elementu *H*, je definována vztahem [11]:

$$(I \otimes H)(u,v) = \min_{(i,j)\in H} I(u+i, v+j) - H(i,j) \tag{4.8}$$

kde:

$\otimes$	- šedotónová dilatace
I	- aktuální podobraz
H	- strukturní element
max	- aritmetická funkce pro výpočet minima



Obr. 12: Aplikace šedotónové eroze

4.3 Detekce rtů

I detekce úst je založena na barevných mapách. Využít lze přímo složky RGB barevného prostoru, popř. detektory založené opět na barevném model YC_bC_r. Stejně jako u detekce lidské kůže se vychází z předpokladu, že i oblast rtů obsahuje výrazně větší podíl červené složky oproti ostatním (modré složky, resp. i zelené). Samozřejmě se u jednotlivých lidí liší rozdíl výraznosti rtů vůči okolní kůži, proto detektory založené na tomto principu nemusí vykazovat konstantně dobré výsledky u

velkého počtu zpracovávaných tváří (detekci u žen může zlepšit použitá rtěnka, u mužů pak mohou vousy pomoci k oddělení rtů od kůže, ale i zakrýt horní ret knírem).

4.3.1 Metody využívající RGB

Barevná transformace využívající Fisherovy lineární diskriminační [13] analýzy, upřednostňuje červenou složku RGB, která se pohybuje u rtů v okolí hodnoty 150 (zbylé dvě složky kolem 80). Okolní kůže má zpravidla větší podíl R složky, ale větší podíl G a B složek ji zesvětlují do narůžovělé podoby. Pro vytvoření výsledné mapy se získané hodnoty prahují, potenciální rty mají zápornou hodnotu.

$$FLDA = [-0{,}289 \quad 0{,}379 \quad 0{,}038] \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4.9}$$

Na stejném principu funguje i další detekce, která určuje podíl červené složky vůči všem složkám RGB. Vychází z *rg* transformace využívající normovaných složek RGB prostoru:

$$r = \frac{R}{(R+G+B)} \quad g = \frac{G}{(R+G+B)} \quad b = \frac{B}{(R+G+B)} \tag{5.0}$$

Součet těchto podílů musí být vždy roven jedné, proto je možné poslední složku vynechat, protože nenese žádnou informaci. Místo výskytu rtů lze tedy určit pouze ze složek *r* a *g*.

Tyto jednoduché postupy dokáží lokalizovat ústa zejména v menších oblastech, kde je možné předpokládat rty na základě předchozí práce s obrazem (hranice obličeje, pozice očí atd.).

4.3.2 C_r kanál z YC_bC_r

Detekování úst může být řešeno i pomocí kanálu C_r nesoucího červenou barevnou složku v modelu YC_bC_r [12].

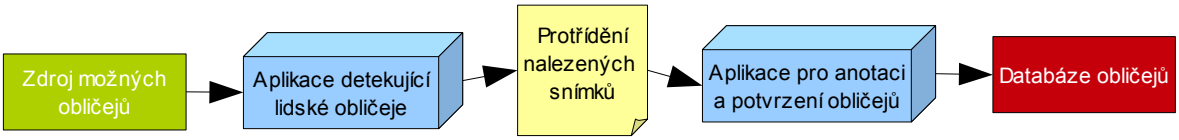
$$MouthMap = C_r^2 \tag{5.1}$$

5 Návrh a implementace

Pro implementaci programu umožňujícího pořízení rozsáhlé databáze lidských obličejů byl zvolen programovací jazyk Java s využitím knihovny OpenCV. Jazyk Java byl zvolen z důvodů snadné implementace uživatelského rozhraní (využívající knihovny Java Swing) i možnosti práce s obrazovými daty. Knihovna OpenCV je pro Javu dostupná ve verzi 1.0, je zaměřená na práci s obrazovými daty, zejména na detekci lidí. Pro tyto účely má zpracované rozhraní umožňující snadné používání webkamery, zpracování snímků a řadu detektorů týkajících se lidské postavy či tváře. Tyto detektory jsou založeny na algoritmu Viola-Jones (viz kap. 3.1). V aplikaci je knihovna využita pouze pro detekování obličejů z předložených snímků, k zpracování videa je použito JMF (Java Media Framework), který je součástí instalace JRE (Java Runtime Edition - programy a knihovny potřebné pro běh programů vytvořených v Javě).

Pořízení databáze vzhledem ke zvolenému řešení byla rozdělena na dvě části (viz diagram návrhu řešení Obr. 13). Jedna aplikace zajišťuje získání snímků s pravděpodobným výskytem obličejů, jejich lokalizace a extrahování. Druhá aplikace umožňuje zpracování těchto snímků a detekce částí lidského obličeje za asistence uživatele. Uživatel by měl v tomto případě sloužit jako kontrola vyhledaných dat a anotovaných oblastí tváří. Detekce potenciálních obličejů by měla probíhat plně automaticky a uživatel by měl jenom zadávat zdroje dat a spouštět samotné detekování.

Rozdělení na dvě části rovněž zvyšuje univerzálnost vytvořených aplikací. Umožňuje použití anotační části i na již dříve získané snímky obličejů z jiných zdrojů. Zároveň se tím tato část zbavuje nutnosti používání knihovny OpenCV (ta je využita jen při detekci) a je tedy možné zpracovávat snímky i u uživatelů, kteří ji nevyužívají a nemají nainstalovánu.



Obr. 13: Princip vytvoření rozsáhlé databáze lidských obličejů

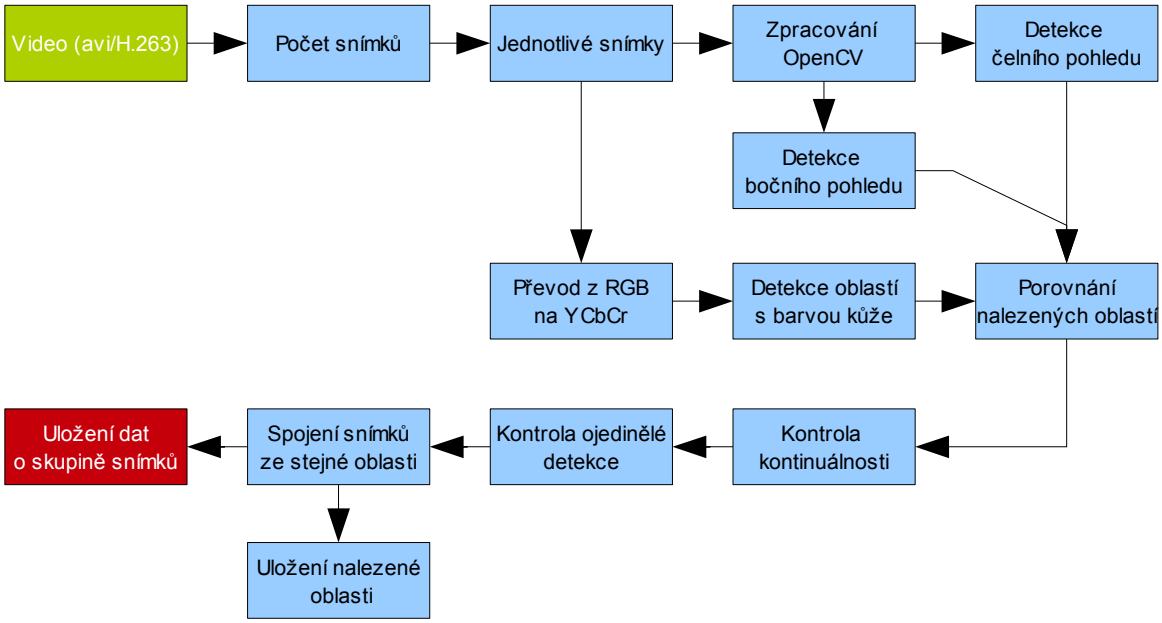
Záměrem je získat i netradiční záběry tváří, proto by neměl detektor propouštět jenom kandidáty, o kterých je pevně přesvědčen, že jsou lidskými obličeji, ale mezi kandidáty by měl řadit i oblasti s méně pravděpodobným výskytem tváří. Kvůli tomu ale může vzniknout celá řada falešně detekovaných snímků, které by bylo vhodné protřídit dříve, než se dostanou k uživateli anotujícího části tváří, protože samotné anotování může být poměrně časově náročné a kvalifikovaná obsluha by byla zbytečně zdržována snímky bez obličeje. Proto obě aplikace spojuje prvek protřídění databáze, který zajistí potvrzení pravdivých detekcí a zamítnutí těch falešných. Idea tohoto prvku je jednorázové rychlé rozhodnutí o velké skupině nalezených snímků, popř. hromadný přístup větší skupiny uživatelů podílejících se na kontrole snímků. V prvním případě by tedy bylo potřeba shlukovat nalezené kandidáty podle společných kritérií do skupin reprezentovaných jedním nebo několika málo kandidáty, kteří by byli prezentování k posouzení. V druhém případě se nabízí nezávislé zobrazování všech snímků více uživatelům, kteří by např. prostřednictvím webového rozhraní postupně rozhodovali o nalezených snímcích. Protože ne všechny snímky musí obsahovat jasně patrné obličeje a uživatelé by nemuseli u sporných snímků rozhodovat jednotně, vznikla by u kandidátů určitá pravděpodobnost možného obličeje, se kterou by mohlo být dále pracováno.

5.1 Získání potenciálních obličejů

Kvůli nutnosti velkého množství potřebných dat není v reálných lidských silách získat takto velký počet fotografií s lidskými tvářemi, a to jak jejich vlastní pořízení (vlastní snímání a fotografování), tak i získání ze sekundárních zdrojů (využití internetových galerií a sociálních sítí). Datové sady vycházející ze statických snímků se pohybují řádově v počtu stovek, rozsáhlejší databáze už využívají zařízení umožňující sérii snímků. Proto bylo pro získání rozsáhlé databáze zvoleno zaměření na zpracování videa, kdy jsou v každém snímku detekovány možné tváře. Pro získání tisíců tváří v různých podobách je tak potřeba záznam v řádu několika desítek minut. Je nutné pouze obstarat vhodné video s předpokladem četného výskytu lidí (nejlépe se zaměřením na jejich tváře).

Bohužel z pohledu práce s videem se Java neukázala jako zrovna nejvhodnější jazyk, na různé formáty videa reaguje značně odlišně. Nejspolehlivějším řešením je použití videa ve formátu AVI s kompresí H.263. Tato komprese byla vytvořena především pro videopřenosy se stálou bitovou rychlostí (bit rate). Nevýhodou stálé bitové rychlosti je, že v případě pohybujícího se objektu se sníží kvalita obrazu (rychle pohybující se objekty pixelizují, zatímco okolí zůstává ostré). Při použití pro detekci obličejů se ovšem problémy v kvalitě nalezených snímků výrazně neprojevovaly.

Princip aplikace představuje diagram na Obr. 14. Zdrojem dat je tedy video, u kterého je nejprve zjištěna jeho velikost v celkovém počtu snímků. Tyto snímky jsou postupně načítány a předkládány jednotlivým detektorům. Aby se zvýšila pravděpodobnost nalezení skutečných obličejů jsou použity tři různé detektory. Dva z nich zajišťuje knihovna OpenCV a měli by detekovat čelní pohled, resp. pohled z profilu. Třetí detektor je založený na metodě rozpoznání specifické barvy lidské kůže (viz kap. 3.2), nejprve převádí snímky do barevného modelu YC_bC_r, ze kterého následně získává mapu výskytu možných obličejů. Protože je velká pravděpodobnost, že dva nebo všechny detektory naleznou obličej na stejné (nebo hodně blízké) pozici, je nutné nalezené oblasti ze všech detektorů porovnat a sjednotit stejné lokalizace.



Obr. 14: Diagram návrhu detekce potenciálních snímků obličejů

Časové koherence je využito i při řešení problému efektivního protřídění falešných kandidátů ještě před zpracováním anotace, zmíněného v předešlé kapitole. Kontrola každého nalezeného snímku větším počtem lidí se jeví časově i organizačně náročná. Přesto možnost využití kontroly dat přes internet nebyla úplně zavržena, jen je použita kombinace s metodou jednorázového rozhodnutí o větším počtu snímků. Aplikace na základě kontinuálního výskytu potenciálních obličejů na stejném místě ve více snímcích za sebou sdružuje tyto lokace do skupin a generuje HTML dokumenty pro rozhodování o pravdivosti detekce těchto skupin.

Další činností aplikace je již jenom uložení skupin nalezených snímků a vytvoření výřezů možných obličejů ze snímků videa.

5.1.1 Zdroje dat

Základem vytvoření kvalitní a rozsáhlé databáze lidských obličejů pomocí videa je vhodný zdroj počátečních dat. Je zapotřebí obstarat videosekvence, u kterých lze předpokládat častý výskyt tváří i dostatečnou rozmanitost jednotlivých osob. Natáčení vlastního záznamu by nejspíš trpělo nedokonalostmi popsanými v kapitole o dostupných datových sadách (malé množství lidí, stejný typ tváří atd.). Proto se získávání vhodných dat zaměřuje na běžně dostupná videa. U takového videa lze také předpokládat výskyt netypických pozic tváří, jejich překrytí nebo deformace, která jsou obtížně simulovatelná při focení nebo laboratorním snímání videa.

Za hlavní zdroj dat byl zvolen webový archiv České televize, který se délkou i kvalitou záznamů jeví jako vhodnější než různé videoservery či sociální sítě. Rozsáhlost archívu zaručuje dostatečný výběr možných zdrojů dat. Vhodným zvolením pořadu lze ovlivnit zaměření výsledné databáze. Pro různorodost detekovaných obličejů je možné použít zpravodajské relace, pro databáze obsahující stejné osoby v různých polohách se nabízí využití diskuzních pořadů (je zde možné najít i netradiční videa v podobě záznamu pokerových turnajů apod.). Pro získání snímků obličejů v netradičních polohách či výrazech, popř. deformovaných nebo zakrytých je možné využít např. sportovní záznamy. U takových přenosů nelze očekávat kvanta nalezených objektů, ale můžou sloužit jako doplnění databáze zajímavými obličeji (množství výskytu snímků s tvářemi u těchto záznamů často kompenzuje jejich délka).

Archivace pořadů na webu byla zahájena v roce 2005, postupně byla rozšířena na všechny čtyři kanály ČT, přičemž se archivují pouze pořady z vlastní produkce. Ze zpravodajského kanálu ČT24 jde získat téměř nepřetržitý záznam celého dne.

5.1.2 Získání a předzpracování videa

Získání dat z webového archívu České televize bylo ještě donedávna možné přes některé volně dostupné přehrávače videa (konkrétně byl použit VLC media player), kterými se lze připojit na videostream a záznam nejenom přehrát, ale i nahrávat. Podobně lze postupovat při záznamu právě běžících pořadů ať už na webu, nebo prostřednictvím přístupu k televiznímu signálu (např. Kolej-netTV). Takto zaznamenaná data jsou uložena prostřednictvím kontejneru MPEG Transport Stream (TS), který je určen pro přenos v chybujícím prostředí, jako je právě DVB (standard digitálního televizního vysílání).

V současné době byla videa konvertována z formátů WMV (určená pro Windows Media Player) a RM (pro Real Media Player) na modernější H.264 a jsou distribuována pomocí Flash

přehrávače. Kodek H.264 umožňuje nízký i vysoký datový tok při různém rozlišení a lze jím kódovat i video v HD rozlišení. Pro získání takových dat existuje celá řada aplikací na stahování záznamů z nejrůznějších video serverů (např. YouTube Downloader ukládající videa do formátu MP4).

Kvůli již zmiňovaným problémům s formáty videa v Javě je nutné video převést do formátu, který spolehlivě zvládne zobrazovat jednotlivé snímky. Informace o schopnostech JMF zpracovávat různé druhy videa se značně liší a nejlepším způsobem jak vybrat vhodný formát se ukázal experimentální přístup. Nejspolehlivější funkčnost při získávání informací o počtu snímků ve videu a načítání jednotlivých obrazů vykazoval kodek H.263 s rozlišením 352 x 288 px v multimediálním kontejneru AVI. K převodu do toho formátu je k dispozici opět mnoho volně dostupných aplikací, které pro tyto účely zcela postačují.

	Zkratka	Název	Vlastnosti
Kontejnery	AVI	Audio Video Interlative	+ možnost volby kodeku + softwarová podpora - nedá se streamovat
	MPEG	Moving Picture Experts Group	+ lokální uložení i streamování + ISO/IES standart multiplexování audio a video streamů . několik vrstev, např. TS vhodné pro DVB
	ASF	Advanced Systems Format	- uzavřený formát, využívaný pouze programy Microsoft - nemožnost převádění na jiný formát
	RealMedia	Real Media Format	+ streamování po internetu + měnění datového toku v čase - nekvalitní obraz
	MP4	MP4	+ otevřený formát pro různá zařízení
Kodeky (ztrátové)	H.263	H.263	+ vysoká komprese . vhodný pro přenos obrazu
	H.264	H.264	+ jednoduchá implementace + vysoká účinnost komprese + vhodný pro HD - zatížení procesoru
	WMA	Windows Media Video	- při kompresi udržuje datový tok, ale zahazuje snímky, nebo je doplňuje

Tabulka 3: Přehled zmiňovaných multimediálních kontejnerů a video kodeků

5.2 Získání dat z videa

Pro zpracování videa je použito balíku *javax.media*, který je součástí knihovny JMF 2.0. Nejprve je zvolené video inicializováno a zjištěna jeho délka. Jelikož tento údaj neslouží jenom k informaci o právě zpracovávaném úseku filmu pro uživatele, ale je jím i určena navigace v načteném vstupním souboru, je časový údaj převeden na celkový počet snímků ve videu. Prostřednictvím vyhledání konkrétního snímku, nebo přeskočení určitého počtu snímků, lze zajistit nastavení počáteční pozice, protože je třeba předpokládat potřebu přeskočení úvodních částí některých záznamů (znělky, konec předchozích pořadů atd.).

Následně jsou postupně načítány snímky do bufferu a předkládány jednotlivým detektorům. Celý průběh od zpracování videa (načítání snímků), přes detekce možných obličejů až po ukládání získaných výřezů je řešen jako jedno vlákno, odvozené jako instance třídy *Thread* (z balíku *Java.lang*). Tato třída definuje základní metody jako je spuštění, zastavení a ukončení vlákna. Je tedy možné zdlouhavý proces pozastavit a následně pokračovat ve zpracování detekce.

5.2.1 Detekce OpenCV

Jednou z možností jak ve zpracovávaném videu nalézt potenciální oblast s lidským obličejem je použití detektorů knihovny OpenCV. Tato knihovna je založena na kaskádě klasifikátorů podobných Haarovým vlnkám. U použitého detektoru se vybraná kaskáda připojuje jako externí XML a součástí distribuované knihovny jsou i některé hotové kaskády pro detekci lidí nebo jejich obličejů.

Přehled dostupných kaskád:

- CASCADE_FRONTALFACE_DEFAULT - základní detekce čelního pohledu
- CASCADE_FRONTALFACE_ALT
- CASCADE_FRONTALFACE_ALT2 - další alternativy detekce čelního pohledu
- CASCADE_FRONTALFACE_ALT_TREE
- CASCADE_PROFILEFACE - detekce bočního pohledu
- CASCADE_FULLBODY - detekce lidského těla
- CASCADE_LOWERBODY - detekce nohou
- CASCADE_UPPERBODY - detekce trupu

Pro detektor zjišťující pohled zepředu byla experimentálně vybrána kaskáda CASCADE_FRONTALFACE_DEFAULT, protože je potřeba zachytit obličeje i z jiných úhlů byl vytvořen i detektor bočních pohledů využívajících kaskádu CASCADE_PROFILEFACE. Použití více kaskád detekujících přímí pohled se ukázalo jako zbytečné, detektory nacházejí stejné obličeje a selhávají ve stejných případech. Lokalizace celého těla nebo jeho horní poloviny a následný pokus o detekci hlavy na základě poměrů a umístění k tělu se při zpracování videa v použitém rozlišení jeví jako zbytečné, jelikož velikost nalezené hlavy by se pohybovala v hodnotách několika pixelů a nebylo by ji možné považovat za relevantní snímek hlavy.

Třídě vycházející z OpenCV je předán snímek načtený do bufferu, na základě rozměrů knihovna alokuje paměť a přebere si data. Následně je aktivována zvolená kaskáda a spuštěna samotná detekce. U té je možné nastavit přístup k vyhledávání objektu zájmu; ovlivnit lze velikost kroku zmenšování okna prohledávajícího vymezený prostor snímku a hledajícího obličej. Nastavit lze i hodnotu seskupování nalezených oblastí, procházení příznaků kaskády může vytvořit na jednom místě shluk kandidátů. Touto hodnotou je možnost tyto kandidáty seskupovat do jedné lokace. Dalším atributem je nastavení průběhu detekce, první alternativa nezmenšuje v krocích skenovací okno, ale mění velikost zkoumaného obrazu. Tuto možnost nelze kombinovat s následujícími. Druhá varianta zařazuje do procesu detekce Cannyho hranový detektor, prahové hodnoty jsou nastaveny pro detekci obličeje a vypouští tak oblasti s malým nebo velkým počtem hran. Tento hranový detektor by měl významně urychlit proces lokalizace obličejů. Třetí volba nachází ve snímku vždy největší oblast s potenciálním obličejem, postupuje tak od největšího okna skenování a jakmile vyhodnotí testovaný

objekt jako obličej, ukončí proces a vrací pozici pouze jednoho objektu. Poslední volbu je možné použít pouze vzájemně s předchozí, v podstatě upravuje hledání největšího objektu tak, že po nalezení jednoho kandidáta pokračuje ve vyhledávání dál, skenování využívá rozsáhlejšího okna a lokalizace obličeje je tak poněkud hrubější a nelze očekávat přesně ořezaný obličej, významně ale zrychluje proces detekce. Posledními dvěma atributy se nastavuje minimální šířka a výška skenovacího okna, při dosažení těchto hodnot se detekce u daného snímku ukončí a pokračuje se dalším.

U vytvořené aplikace je u obou detektorů zvolen poměr zmenšení okna o 20% (hodnota 1,2), hodnota seskupování byla ponechána defaultní. Styl procházení kaskády je nastaven kombinací druhé až čtvrté varianty (tj. použití Cannyho detektoru a styl průchodu největšího objektu rozšířenou i pro ostatní rozměry). Minimální rozměry obdélníku ohraničujícího potenciální obličej má možnost nastavit uživatel. Nalezené objekty ukládá detektor OpenCV hromadně do pole obdélníků, obdélník je reprezentován třídou *Rectangle* (součást balíku *java.awt*), výška a šířka obdélníku a x-ová a y-ová vzdálenost od nultého bodu (levý horní roh obrázku).

5.2.2 Detekce kůže

Třetí detektor v pořadí, aplikovaný na snímek v bufferu, je založený na odlišení barevného podání odstínů lidské kůže od okolí. Vstupní obraz je standardně reprezentovány barevným prostorem RGB. Zvolená detekce lidské kůže vychází z barevného modelu YC_bC_r a je tedy nutné provést převod dle vztahů 3.2 a 3.3 uvedených v kap. 3.2.1.

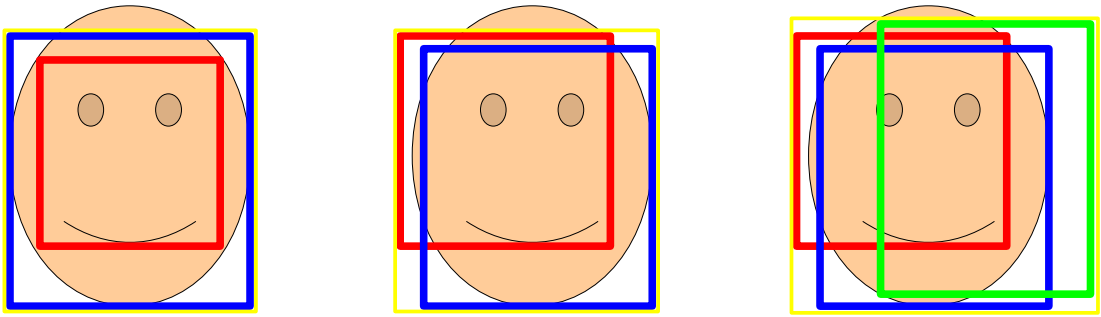
Rozdělení snímku na tři složky RGB zajišťuje přímo třída *BufferImage* (obsažena v balíku *java.awt.image*), složka Y nese pouze informaci o jasu obrazu a není tedy rozhodující při zobrazení jednotlivých barev, takže v této detekci nemá využití a není ji potřeba ani přepočítávat. Z důvodů výpočetní a časové náročnosti detekovacího procesu nebylo do detektoru zařazeno zpracování převedených kanálu C_b a C_r pomocí histogramu. Sestavení histogramu obou hodnot, následnou ekvivalizací histogramů a nalezení extrémů pro nastavení prahů by mohlo vést k lepším výsledkům detekce, ale toto řešení se ukazovalo jako příliš náročné a výpočetně neadekvátní v poměru k počtu a kvalitě nalezených tváří (zpracování histogramu je použito až u anotace obličejových části).

Pro vypočítané kanály C_b a C_r jsou proto hodnoty pro prahování stanoveny pevně. Testováním byly intervaly pro výskyt barvy lidské kůže určeny v rozmezí 100 – 115 u kanálu C_b a 145 – 165 u C_r. Z hodnot pixelů vstupního obrazu je sestavena pravdivostní mapa určující zda jednotlivé pixely leží ve stanovených intervalech obou kanálů. Pokud ano, je hodnota bodu v mapě nastavena jako *true*, v opačném případě jako *false.* Mapa je poté procházena skenovacím oknem, které vyhledává minimálně nadpoloviční většinu pravdivých buněk mapy. Pokud je nalezeno více takových oblastí je vybrána ta s největším počtem pravdivých buněk. Není-li nalezena žádná oblast s potřebným množstvím bodů, je okno pětinásobně zmenšeno a celý cyklus se opakuje až do zmenšení okna na zvolenou mez.

Pro uložení nalezených lokací je s ohledem na standardizované ukládání předchozích OpenCV detektorů zvolen stejný formát. K vytvoření pole obdélníků je vzhledem k dynamičnosti pole využito seznamu (třída ArrayList balíku *java.util*). Z důvodů dalšího zpracování nalezených obličejů je nalezená oblast rozšířena ve všech směrech o čtvrtinu rozměrů okna s výjimkou spodního okraje, který je navýšen o třetinu. Tímto krokem by měla narůst pravděpodobnost, že bude získán celý obličej, popř. i s částí hlavy a dalším zpracováním může být libovolně upraven dle požadavků.

5.2.3 Sdružování detekovaných oblastí

Protože je použití prvních tří detektorů navzájem nezávislé, mohou nastat (resp. by ve většině případů nastat měly) duplicitní detekce. Proto jsou data ze všech detektorů předána funkci starající se o sjednocení nalezených souřadnic. Za primární byl zvolen detektor využívající barvu lidské kůže. S údaji jím vytvořenými, uloženými v datové struktuře seznamu, jsou porovnávány jednotlivé položky polí s údaji z následujících dvou detektorů. Pokud seznam neobsahuje konkrétní prvek na stejné souřadnici ani jí blízké, je o tento prvek seznam rozšířen. Aby nedocházelo ke zdvojování detekcí, jsou sjednocovány snímky překrývající se alespoň z jedné třetiny v obou směrech x-ových a y-ových souřadnic. Je-li rozpoznáno překrytí lokalizačních oblastí, jsou porovnány vzdálenosti jednotlivých stran obdélníků a upraveny tak, aby výsledná oblast byla zvětšena na maximální detekované hranice ze všech detektorů současně.



Obr. 15: Princip rozšíření detekovaných hranic při sdružováni stejných lokací

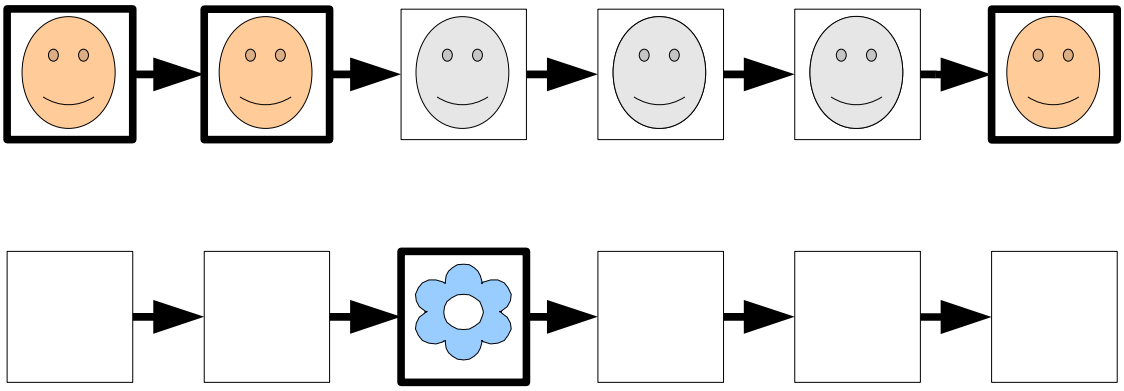
- *červená, modrá a zelená jsou výsledky lokalizace jednotlivých detektorů*
- *žlutá je nově vzniklá hranice výřezu*

5.2.4 Využití časové koherence

V případě, že jsou kontinuálně, v navazujících snímcích, detekovány obličeje na stejné lokaci, v několika snímcích je tato kontinuita přerušena, ale následné detekce se opět vyskytují ve stejné oblasti, uloží se do seznamu i výřezy snímků, u kterých detektor obličeje nezaznamenal.

Pozice detekovaných obličejů v jednotlivých snímcích uložených v seznamu nejsou přímo ukládány do souboru databáze obličejů. Seznam obdélníků nalezených detektory je vložen jako jeden prvek do vytvořeného pole o deseti pozicích, prvek se v poli posunuje a pole se postupně zaplňuje seznamy detekovaných obličejů z jednotlivých navazujících snímků (v poli je vždy deset seznamů se souřadnicemi obličejů). Porovnáním posledního prvku pole s jednotlivými prvky předešlými (resp. snímku se snímky následujícími) se kontroluje návaznost detekce v jednotlivých oblastech. Pokud vznikne u některých oblastí proluka, která ale končí v některém z následujících devíti snímků, jsou jednotlivé seznamy doplněny o souřadnice tohoto prvku (použije se hodnot z posledního prvku). Po případném doplnění o scházející obličeje je pole seznamů kontrolováno na ojedinělý výskyt možné tváře. Obsah předposledního prvku je kontrolován s okolními a pokud se v nich některá pozice nevyskytuje je odstraněna i ze seznamu v tomto prvku pole.

Vedle pole se seznamy nalezených oblastí je vytvořeno další o stejné velikosti, ve kterém se průběžně uchovávají načtené skenované snímky (buffery obrázků). Jakmile prvek na poslední pozici v poli seznamů toto pole opouští, je použit i poslední prvek pole snímků a podle souřadnic v seznamu jsou z načteného snímku vyříznuty jednotlivé oblasti s pravděpodobným výskytem lidského obličeje. Tyto oblasti jsou uloženy do samotných souborů ve formátu JPEG. Název souboru je tvořen pořadím snímku ze kterého byl získán a souřadnicemi počátečního bodu vyřezávané oblasti (levý horní roh). Název souboru 922_131_69.jpg tedy označuje, že nalezený kandidát je ze snímku 922 a počátek výřezu měl souřadnice 131 px na x-ové ose a 69 px na y-ové ose. Stejným názvem jsou pojmenovány položky v databázovém souboru.



Obr. 16: Využití časové koherence

- *doplnění nezdetekovaných oblastí mezi kontinuálně detekovanými tvářemi*
- *odstranění ojediněle zdetekovaného objektu*

5.2.5 Generování HTML formuláře

Nalezené snímky nejsou do databáze zapisovány současně s ukládáním samotných snímků, ale je vytvořen nový seznam sdružující názvy souborů podle souřadnic nalezených oblastí do dalšího seznamu (hodnotami x a y je určena skupina, do které jsou jednotlivé názvy souborů ukládány v seznamu). Z tohoto seznamu jsou názvy zapsány v jednom řádku do databázového souboru až v době, kdy je přerušena kontinuita nalezených kandidátů na daném místě ve snímku. Takto složité přeuspořádání nalezených dat slouží především pro možnost rychlého protřídění nesprávně detekovaných snímků, kdy je předložen uživateli k posouzení jenom jeden kandidát a na jeho základě rozhodne o zamítnutí celé skupiny.

K tomuto účelů může být vygenerován *HTML* soubor s přehledem jednotlivých skupin. Jako reprezentant znázorňující obsah skupiny je zvolen snímek uprostřed seznamu. Tím je možné předejít nesprávnému vyhodnocení uživatelem, kdy první snímky v seznamu mohou částečně obsahovat video předěly mezi scénami a u prolínaných snímků nemusí být lidská tvář na první pohled jasně patrná. *HTML* soubor je koncipován jako jeden formulář, kde je pod každým obrázkem kandidáta reprezentujícího danou skupinu snímků umístěno zaškrtávací tlačítko (*checkbox*). Uživatel zatrhne políčka u snímků na stránce, které neobsahují lidský obličej a pomocí tlačítka (submit) odešle údaje z formuláře vygenerovanému php souboru, který uloží změny do databázového souboru. Tuto kontrolu

lze provádět opakovaně a více lidmi, tím se zvyšuje pravděpodobnost správného rozdělení. V počátečním stavu jsou v souboru před každým řádkem skupiny snímků hodnoty nastaveny na nulu. K těmto hodnotám *php* skript zpracovávající formulář přičítá (resp. odečítá) jedničku v případě, že políčko nebylo zaškrtnuto (resp. bylo zaškrtnuto). Pro snadnější zpracování a přehlednost je každý obrázek společně se zaškrtávacím tlačítkem umístěn v samostatném DIVu (formátovací značka jazyka HTML určující blokový element) pomocí jazyka *javascript* jsou oba prvky propojeny. K nastavení hodnoty checkboxu tedy postačí označení daného obrázku kurzorem. Současně se i změní pozadí bloku ze zelené na červenou, aby měl uživatel jasnější přehled o vytříděných snímcích.

Při testování se ukázalo, že je poměr vyřazených a ponechaných skupin docela vyrovnaný a v některých případech je třeba označit více obrázků než jich ponechat. Není to až tolik způsobeno mírou falešných detekcí, ale především tím, že chybné detekce se nevyskytují v tak velkých skupinách a jsou více ojedinělé, vytvářejí skupiny jako velké bloky s množstvím lidských obličejů. Nabízí se proto možnost označovat správně detekované skupiny. Vzhledem k dalšímu zpracování je ale přijatelnější propustit i některé nevhodné kandidáty než vyloučit podstatné snímky. Z důvodů, že člověk spíše přehlédne některé snímky, než že je chybně označí bylo ponecháno označení pro vylučování skupin. Pro lepší přehlednost jsou formuláře rozděleny, každý z nich zobrazuje volbu pro stovku skupin nalezených kandidátů.

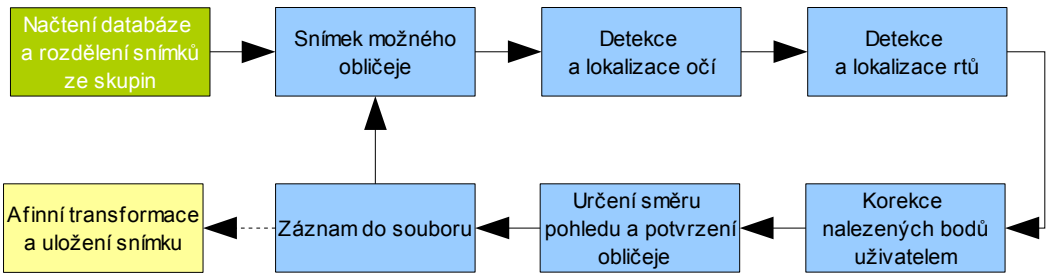
Obr. 17: HTML formulář pro rychlou redukci nalezených snímků

5.3 Anotování dat

Uživatelská část pro akceptování detekovaných snímků a zadání souřadnic jednotlivých rysů obličeje je řešena jako samostatná aplikace. Jsou na ni kladeny požadavky na rychlou kontrolu nalezených snímků a případné určení souřadnic důležitých bodů v obličeji. Ruční anotace desítek tisíců obličejů je velmi časově náročná a u takto rozsáhlých databází se prakticky nevyskytuje. Úkolem této aplikace je provést značnou část označování objektů na tvářích prostřednictvím dalších

detektorů. Uživatel by měl v procesu figurovat jako kontrolní prvek správné lokalizace a případné chybné lokace poopravit.

Návrh aplikace přibližuje diagram na Obr. 18. Využívá se souboru v němž jsou uloženy databáze skupin získané aplikací pro detekci lidských obličejů ve videu (soubor mohl být upraven php skriptem z webového rozhraní). Data ze souboru jsou načítána po řádcích a následně dělena podle znaků mezery. Pokud je hned první hodnota záporná (tzn. že prostřednictvím formuláře byla tato skupina označena za neobsahující obličeje), řádek se přeskakuje. Pokud je nulová nebo kladná jsou prvky řádku načteny do seznamu jako struktura obsahující název souboru, druh pohledu, pozice obou očí a pozice dvou koutků úst.



Obr. 18: Diagram návrhu potvrzování a anotace obličejů

Podle názvu souboru je načten příslušný obrázek obsahující potenciální obličej detekovaný předchozí aplikací. Druh pohledu není určován automaticky a je jen na rozhodnutí uživatele, do které kategorie obličej zařadí. Základním rozdělením je čelní, pravý a levý směr pohledu. Díky tomu, že detektory jsou schopny nalézt specifické a těžko zařaditelné pohledy (překrytý nebo přetočený pohled, neúplné tváře, rozmazané snímky atd.), které by pro budoucí zpracováním při strojovém učení mohly představovat problém, nebo se naopak výrazně podílet na učení specifických detektorů, je tyto netypické pohledy možné zařadit do kategorie ostatních pohledů (příklady na Obr. 19). Jako zkratky označující vybraný pohled při uložení v souboru nové databáze byly zvoleny F pro čelní pohled, L a P pro levý, resp. pravý pohled (myslí se směr natočení tváře snímané osoby z pohledu uživatele) a O pro všechny ostatní, obtížně zařaditelné pohledy.

Pokud jsou označeny oči, ústa i směr pohledu potvrzuje uživatel, že daný snímek obsahuje skutečně zaznamenaný lidský obličej a jméno souboru obrázku je společně se značkou pohledu a souřadnicemi umístění očí a úst zaznamenán v novém souboru. V každém řádku je zaznamenán jeden snímek obličeje.

Do souboru jsou data o anotovaných snímcích ukládána se zpožděním pěti snímků. Tato rezerva umožňuje návrat ke zpracovanému snímku a jeho případnou opravu.

Součástí aplikace je i možnost alternativního uložení snímků ze zpracované databáze. Protože je video zpracováváno v poněkud netradičním rozlišení s poměrem stran 1,2 a nejčastějším poměrem stran u televizního vysílání je v dnešní době 1,6, vycházejí výsledné snímky obličejů částečně zkreslené. Tuto deformaci lze eliminovat při ukládání výsledných snímků obličejů. Metodou afinních transformací (v Javě implementovány třídou *AffineTransform* v balíku *java.awt.geom*) jsou výstupní data upravena do jednotné podoby, kdy osoby na snímcích mají oči srovnány do jedné roviny, velikost je upravena podle vzdálenosti očí a poměr stran upraven podle zvoleného koeficientu vztahujícího se na původní poměr stran videa.

Obr. 19: Obtížně zařaditelné typy tváří

a - prolínání obrazů
b - neúplné obličeje
c - dva obličeje na jednom snímku
d - rozmazané obličeje

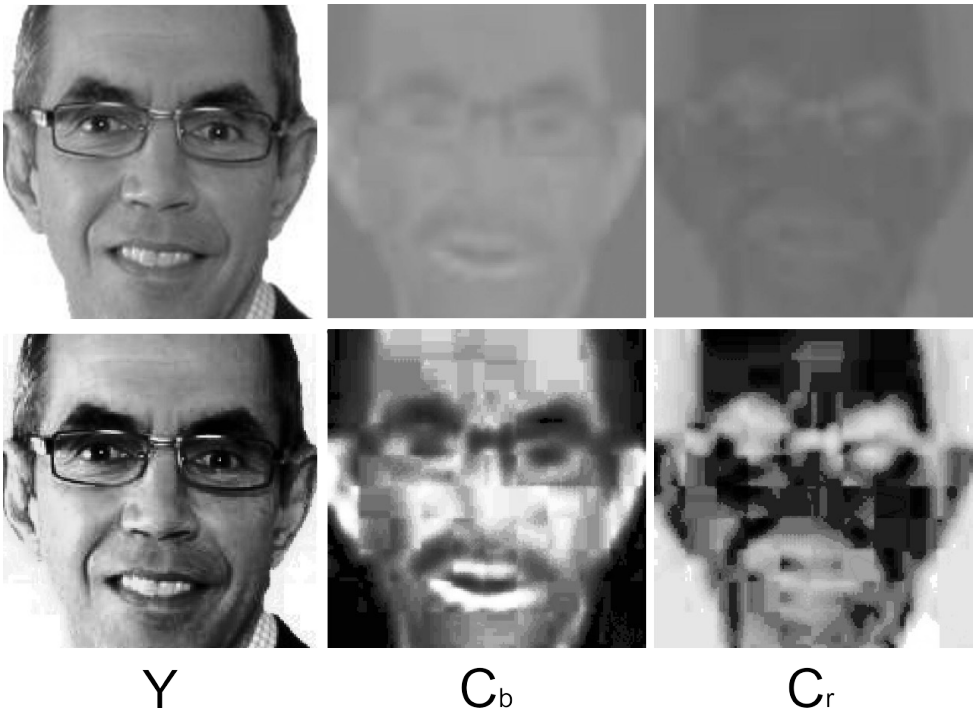
5.3.1 Lokalizace očí

Původní záměr automatické detekce očí se neukázal jako nejvhodnější řešení. Při testování na větších snímcích s vyšší úrovní detailů pracoval implementovaný detektor na základě vztahu (4.5) poměrně dobře. Pokud se ještě zúžila oblast vyhledávání možných očí na pravděpodobný výskyt očí na snímku (na základě předpokladu, že oči se na detekovaném snímku nacházejí v horní polovině a určité vzdálenosti od okrajů), byla úspěšnost ještě vyšší. Problém nastal u detekovaných obličejů z videí, která nemají příliš velké rozlišení. U malých obrázků, kde je oko reprezentováno jen několika pixely, detekce funguje dobře jen u kvalitních čelních pohledů. U obličejů pořízených v pohybu a za ne úplně ideálního osvětlení detekce často chybuje a upíná se na řasy nebo okraje očí. Nutno podotknout, že u některých snímků je přesná lokalizace očí (ideálně očních panenek) obtížná i pro samotného uživatele.

Přesto od detekce pomocí modelu YC_bC_r nebylo upuštěno. Je využito toho, že při detekci obličejů jsou v databázi ukládány snímky do skupin a stejně tak i načítány pro předložení uživateli při anotaci. Po sobě tak nenásledují detekované objekty podle pořadí, jakým byly vybrány z jednotlivých snímků. Uživateli jsou postupně předkládány objekty tak, jak na sebe navazovaly ve snímcích na stejné pozici. Pokud tedy byly v jednom snímku lokalizovány dvě osoby a vyskytovaly se společně i v řadě následujících snímků, je při anotaci zpracovávána nejprve jedna osoba ze všech snímků, až poté následuje druhá (v případě řazení podle pořadí detekce by se tyto osoby neustále střídaly).

Jelikož se série snímků jedné osoby vyskytuje přibližně na stejném místě je pravděpodobné, že i pozice očí bude vždy v následujícím snímku poblíž pozice ve snímku předchozím. Detektor pro pozici očí tak pracuje jen v malém okolí pozice z předchozího snímku a jen koriguje zadanou lokalizaci. Uživatel v ideálním případě označí pozici očí vždy jenom u prvního snímku ze série a detektor by měl zajistit udržení označené pozice neustále na místě očí.

Pro detekci očí je opět využito převedení načteného obrázku z barevného RGB do modelu YC_bC_r. Všechny tři kanály jsou upraveny pomocí ekvalizace histogramu (podle vztahu 4.2). Nejprve jsou všechny pixely sdruženy v seznamu do skupin podle hodnoty dané složky, následně jsou skupiny pomocí rekurzivního řazení přeskupeny od nejmenších hodnot po největší. Z podílu četnosti výskytu k celkovému počtu pixelů a umístění v pořadí seznamu je určen nový podíl jasových složek jednotlivých skupin pixelů. Na základě tohoto podílu jsou hodnoty přepočteny do intervalu <0;255> a přiřazeny jednotlivým pozicím v obrázku na základě porovnání s původní hodnotou (výsledky viz Obr. 20).



Obr. 20: Ekvalizace histogramu složek YC_bC_r

- *nahoře: původní hodnoty*
- *dole: hodnoty upravené pomocí histogramu*

Podle vztahu 4.4 je vytvořena EyeMapL, kanál Y upravený histogramem je podroben šedotónové dilatace a erozi (viz kap. 4.2.1). Do jednoduchého pole je sestaven strukturní element, který je postupně aplikován na úseky načtené složky a konkrétní hodnota pixelu je u dilatace nahrazena maximální (resp. minimální u eroze) hodnotou procházeného okolí. Následně jsou obě složky poděleny, tzn. že jasná místa na dilatovaném obrázku zůstanou jasná pokud jsou konfrontována s tmavými místy eroze. Jak je vidět na Obr. 21 zůstane zvýrazněný jenom odlesk očí (v tomto konkrétním případě i část brýlí). Bohužel právě tyto detaily můžou u obrazů v menším rozlišení chybět.

Obr. 21: Ekvalizovaná složka Y, dilatace, eroze, EyeMapL

Druhá část detektoru (EyeMapC podle vztahu 4.3) je jednoduchou aplikací uvedeného vztahu. Následně jsou obě mapy (EyeMapC a EyeMapL) vynásobeny (vztah 4.5). Jak dokládá Obr. 22, v některých případech je detekce očí úspěšnější jen za pomocí EyeMapL.

Výsledná mapa je procházena v okolí bodů, ve kterých byly v předchozím snímku označeny pozice očí, okolí je dáno 1/15 celkového rozměru obrázku v daném směru. Jsou nalezeny nejkrajnější světlé body v této lokaci (nejvzdálenější světlé body ve čtyřech směrech středu) a z nich je vypočítána nová pozice oka.

Obr. 22: EyeMapL, EyeMapC, EyeMap

5.3.2 Lokalizace rtů

U detekce rtů se postupuje způsobem popsaným v kap. 4.3, jen nebyl použit detektor využívající složku C_r protože při testování nedosahoval dobrých výsledků. Jak je patrné i z kanálu C_r na Obr. 20 není ani po ekvalizaci histogramu jasně patrná kontura rtů a snad výhodnější by bylo použít invertovaný kanál C_b.

Výsledný detektor je kombinací přepočtu RGB do FLDA (vztah 4.9), u které byl experimentálně stanovený práh na hodnotu -8 a *rg* transformace (vztah 5.0), ze které je však využita jen naprahovaná složka r (práh je stanoven na 0,5). Pro zacelení výsledné mapy je aplikována dilatace. Výsledná mapa rtů je složena z výsledků obou těchto detektorů, které jsou vzájemně vynásobeny (viz Obr. 23).

Lokalizace koutků úst probíhá podobně jako v případě očí. Je procházena oblast mezi koutky úst označenými u minulého snímku (rozšíření o 1/15 ve třech směrech od středu, přilehlé strany bodů jsou spojeny). Počáteční dva body jsou nastaveny na střed úst a vyhledávají nejvzdálenější světlé

body ve svém horizontálním směru (jeden vlevo, druhý vpravo), v těchto bodech jsou pixely procházeny ještě vertikálně a podle vzdálenosti světlých bodů je určen střed nových pozic úst.

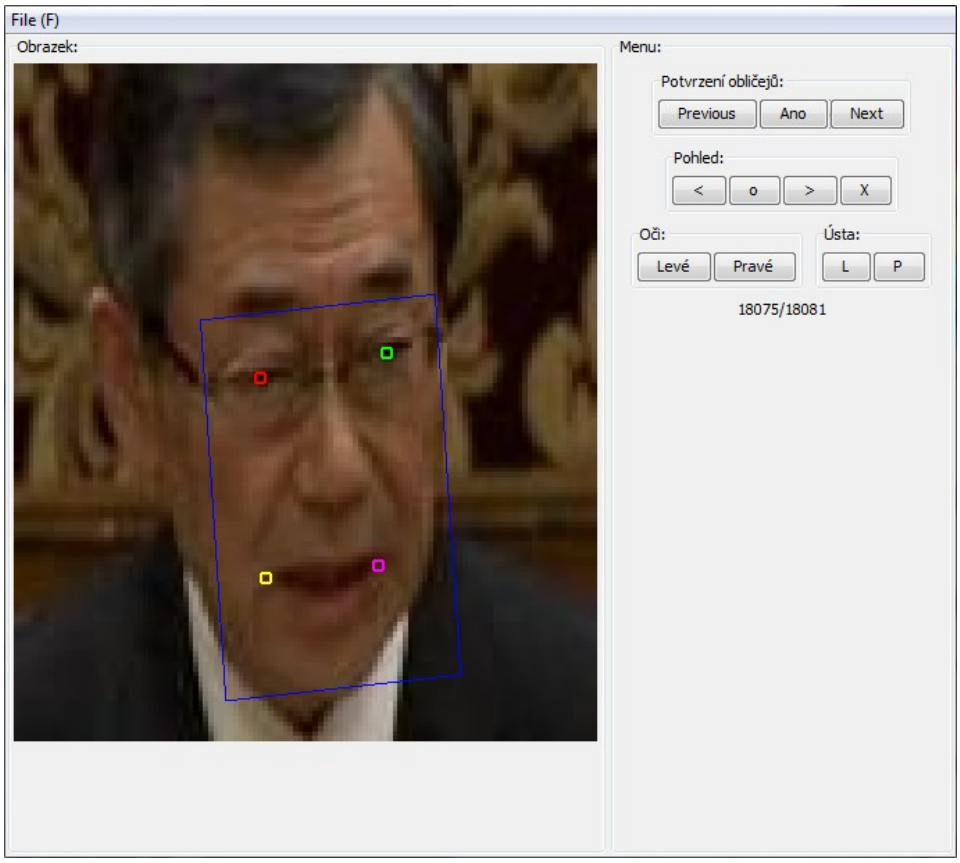
Obr. 23: Detekce rtů

5.3.3 Ovládání

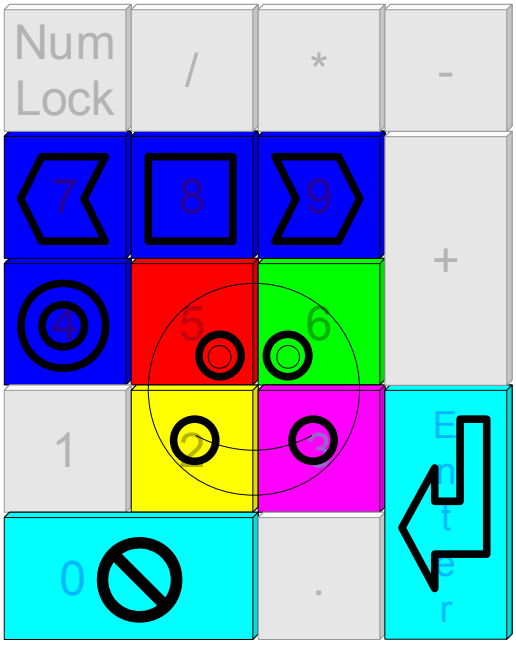
Uživatelské rozhraní je vytvořeno pomocí knihovny *java.Swing* obsahující třídy vhodné k tvorbě tlačítek a ovládacích prvků. Z důvodu zrychlení práce s aplikací jsou ale ovládací tlačítka jenom alternativou využitou spíš jen pro demonstraci aplikace. Zaškolený uživatel anotující větší množství dat má možnost ovládat aplikaci pomocí numerického bloku klávesnice.

Práce s aplikací by měla probíhat tak, že uživatel prostřednictvím menu *File* vybere databázi vygenerovanou při zpracování videa nebo vybere dříve rozpracovanou databázi anotační aplikace. Po načtení snímku kliká do obrazu myší. Místo posledního kliknutí je vždy označeno růžovou kružnicí. Prostřednictvím jedné z kláves numerické bloku (2, 3, 5 nebo 6) přiřadí zvolenou lokaci příslušnému objektu obličeje (5 – levé oko (červená), 6 – pravé oko (zelená), 2 – levý koutek úst (žlutá), 3 – pravý koutek úst (purpurová)), přičemž se změní barva kružnice na barvu příslušného objektu. Klávesami o řadu výš v num. bloku je zadáván pohled osoby na snímku (7 – pohled vpravo, 8 – čelní pohled, 9 – pohled vlevo). Usnadnit uživateli orientaci může modrý obdélník lemující tvář podle zvolených pozic očí a úst. Pokud pohled nespadá ani do jedné z těchto skupin, měla by být zvolena kategorie „ostatní“ (klávesa 4). Nakonec zbývá jenom potvrdit klávesou *Enter*. Nejedná-li se o obličej je možné přeskočit na další snímek klávesou 0. Databázi není potřeba ukládat, průběžně je ukládána do souboru.

Rozmístění ovládacích kláves (viz Obr. 25) je vedeno snahou minimalizovat pohyb při anotování, jednou rukou uživatel používá myš a opravuje automaticky detekované pozice, zatímco druhou rukou tyto pozice jenom potvrzuje a pohybuje se v pořadí snímků (potřebuje-li se uživatel k některému snímku vrátit, může až 5x stisknou levou šipku).



Obr. 24: Ukázka aplikace pro anotování obličejů



Obr. 25: Ovládání pomocí num. bloku

6 Získané databáze

Pomocí vytvořených aplikací má být pořízena rozsáhlá databáze lidských tváří. Kvalita získaných snímků ovšem nezáleží jenom na implementovaných detektorech, neméně důležitý je výběr vstupních dat. Vybrané video musí mít předpoklady pro výskyt požadovaného typu nebo množství obličejů

Aplikace byla testována na třech rozsáhlejších video záznamech. Dva byly získány z archívů České televize (viz kap. 5.1.1) a jeden z nahrávky živého vysílání. Vybrán byl jeden zpravodajský pořad (*Události* ze dne 26.4.2011), jeden dokumentární pořad (*Rozmarná léta českého filmu* z 5.3.2011) a jeden sportovní přenos (třetina utkání v hokeji Slovensko – Rusko hraného 3.5.2011).

6.1 Použité zdroje dat

Zvolené pořady byly vybrány podle odlišných kritérií. V případě večerních zpráv je předpoklad nalezení velkého množství čelních pohledů u moderátorů pořadu a velkou variabilitu obličejů v jednotlivých zpravodajských spotech, ve kterých by se měli objevovat především lidé v nejrůznějších pozicích. Velkou část spotů tvoří rozhovory nebo komentáře se záběry hlav osob v dostatečné velikosti pro získání kvalitních snímků. Součástí zpravodajství zpravidla bývají i záběry z Poslanecké sněmovny, z jejichž schůzí jsou pravidelně pořizovány záznamy a do budoucna by bylo možné uvažovat o těchto záznamech jako o možném zdroji dat.

Dalším zvoleným záznamem je pořad o historii českého filmu, v němž filmový tvůrci a herci popisují vznik daného filmu v konkrétním roce. Dokument je koncipován tak, že mezi filmové ukázky jsou vloženy stručné komentáře zúčastněných osob. Počet lidí komentující filmové dění je poměrně velký a střídají se v rychlém sledu po pronesení několika vět, navíc jsou při projevu zabíráni podobně jako moderátoři u zpráv. Takže lze i u tohoto pořadu očekávat dostatečný počet různých osob v dostatečně podrobném snímku. Mimo to se tímto pořadem otestuje detekce ve filmových záběrech, ve kterých nelze očekávat tolik detailů na tváře lidí, ale scény jsou dynamičtější a osoby se nacházejí v různorodějších pozicích.

Třetím pořadem byl zvolen sportovní přenos. U záznamu, který je většinu času pořizován ze značné vzdálenosti od ledu nelze předpokládat získání velkého počtu nalezených obličejů. Očekávání od tohoto typu programu je nalezení specifických obličejů, které jsou obtížně získatelné jakýmkoliv simulováním v laboratorních podmínkách. Během přerušení hry se kamery soustředí na detailní záběry hráčů nebo publika (na rozdíl např. od fotbalu kde není tolik detailů), v těchto momentech by měly detektory nalézt záběry osob s očima krytýma hledím přilby, diváky za plexisklem, emoce ve tvářích hráčů, trenérů i diváků. Součástí přenosu po skončení třetiny je i debata odborníků ve studiu a rozhovory s hráči ve výstroji. V této části už kamery opět zabírají především hlavy osob a detektory by případně měly zaznamenat alespoň tyto obličeje.

Pořízená videa byla převedena do formátu AVI s kodekem H.263 (viz 5.1.2) o rozměrech 352 x 288 px, s frekvencemi snímků 25 (u hokeje a dokumentu) a 30 (u zpráv) snímků za sekundu a datovým tokem v rozmezí 1633 až 3984 kbps. Záznam zpravodajského pořadu má délku 24:51 minut (tj. 44 730 snímků), dokument má délku 49:53 minut (74 847 snímků) a hokejový zápas trvá 42:30 minut (63 756 snímků).

6.2 Kvalita a rychlost zpracování

Úvodní detekcí prošly kompletně všechna videa v plné délce. Hokejový zápas byl zpracováván bez přerušení, zpravodajství bylo jedenkrát přerušeno, zpracování dokumentu proběhlo dokonce v pěti úsecích. Důvodem přerušení bylo vytížení výpočetní techniky pro jiné účely a výpadek proudu. Díky možnosti nastavení počátečního snímku není problém pokračovat v rozpracovaném videu.

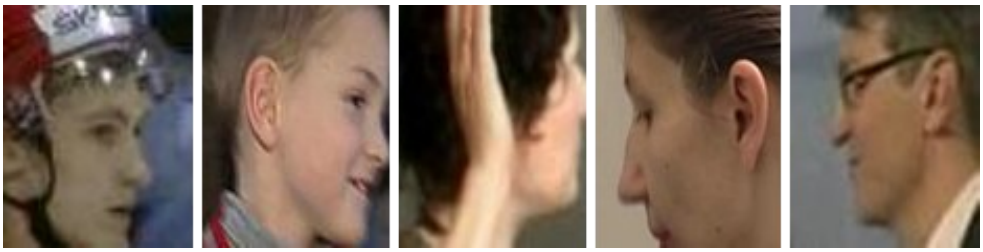
Úplné anotaci byly podrobeny pouze obličeje nalezené v záznamu hokejového zápasu. Celkový počet nalezených obličejů v tomto videu je sice nejmenší, ale z hlediska testování aplikace pro poloautomatické anotování jsou tato data nejzajímavější. Obličeje nalezené ve zbylých záznamech byly pomocí anotační aplikace podrobeny kontrole, zda se skutečně jedná o obličeje. Všechny snímky ve výsledné databázi lze brát za lidské obličeje.

6.2.1 Získaná data

Prostřednictvím detekce obličejů z videa bylo u *Událostí* detekováno 37 499 možných obličejů, rychlým tříděním prostřednictvím webového rozhraní byl tento počet redukován na 20 707. Tyto snímky byly předloženy anotační aplikace, ze které vzešlo celkem 17 709 snímků obličejů. Z dokumentárního pořadu bylo získáno 50 627 kandidátů, rychlou redukcí byl tento stav snížen na 39 627, po potvrzení anotační aplikací zůstalo celkem 34 743 obličejů. V hokejovém utkání bylo nalezeno pouze 7 570 potenciálních obličejů, rychlým výběrem byl počet snížen na 6 623, ze kterých bylo 6 466 skutečných obličejů. V průměru je přes 65% zdetekovaných snímků z videa skutečně obličej. Rychlou redukcí se vyřadí průměrně 26% snímků.

Mezi detekovanými daty lze najít velké množství kvalitních čelních pohledů v rozlišení převyšujícím 100 px na stranu (viz Obr. 27). Velikost i kvalita vyřezaných fotek z videa je ale velmi proměnlivá. Závisí na nastavené minimální velikosti skenovacího okna i na probíhající scéně. Je pochopitelné, že malé výřezy se nemůžou podrobností rovnat větším objektům, ale jako materiály pro strojové učení se předkládají snímky o velikosti 16 x 16 px, čemuž vyhovují všechny nalezené snímky obličejů (minimální délka strany byla nastavena na 50 px).

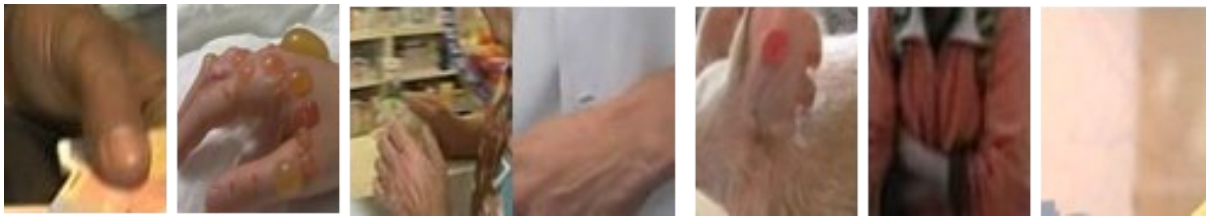
Při detekování se příliš neprojevovala druhá kaskáda z OpenCV, která by měla nacházet především profilové pohledy. Většina tváří zabíraná z boku byla detekována buď detektorem lidské kůže nebo čelním detektorem (viz Obr. 26). Výsledek mohl ovlivnit výskyt čistě profilových záběrů, kterých přece jenom není v televizních záběrech tolik (většinou se jedná o poloprofil). Přesto získané databáze obsahují i tyto snímky a i několik snímků otáčejících se osob, kdy je zachycena část obličeje ve chvíli, kdy už je osoba otočena zády.



Obr. 26: Ukázka detekovaných profilových pohledů

Obr. 27: Ukázka detekovaných čelních pohledů

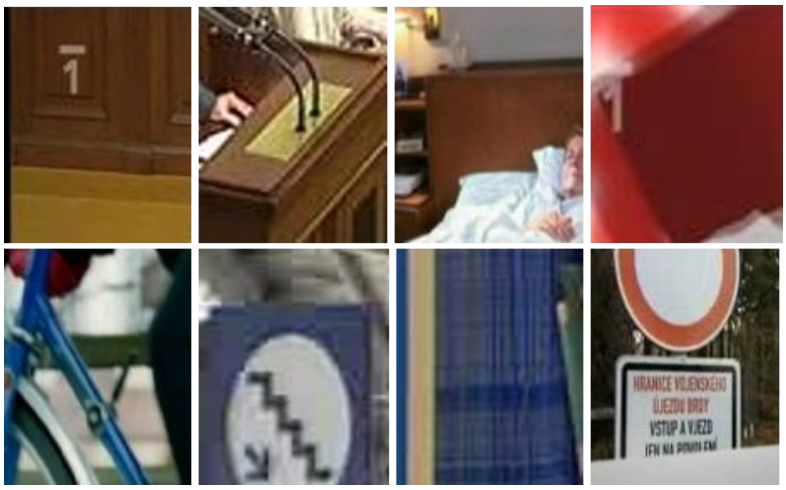
Nezanedbatelné jsou ovšem i chybné detekce. Největším úskalím detektoru lidské kůže jsou růžová, oranžová a načervenalá pozadí, výstřihy, ruce a jakékoliv narůžovělé objekty (viz Obr. 28). Bohužel s tímto se u detektoru založeného na barevném podání musí počítat a jedinou obranou je spojit více typů detektorů do kaskády podobně jako fungují použité detektory z OpenCV.



Obr. 28: Problémové objekty

Dalšími problémovými objekty pro detektor kůže byly hnědé objekty (různé dveře a zejména lavice v Poslanecké sněmovně). Tento problém by šel vyřešit změnou prahů pro určení barvy kůže, ale mohlo by se stát, že by detektor vynechával některé osoby s tmavší barvou pleti a ztratil by tak na univerzálnosti. Takovouto úpravu by bylo možné provést v případě použití detektoru na konkrétní typ záznamu (např. právě záznamy schůzí PSP). Určitých chyb se dopouští i použité kaskádové detektory. Na rozdíl od detektoru kůže jsou chyby mnohem méně předvídatelné a nevyskytují se pouze u oválných objektů, které by mohly připomínat obličej (viz Obr. 29). Podstatnou vadou při

detekci obličejů je i jejich špatné ořezání (viz Obr. 30). Tímto vznikne místo snímku vhodného pro uložení do databáze falešná detekce. Částečně lze předejít tomuto problému zvětšením výřezového okna, ale u jiných snímků by vznikaly příliš obecné snímky obličeje a širokého okolí.



Obr. 29: Hrubé chyby detekce

- *nahoře: chyby detektoru* YC_bC_r
- *dole: chyby detektoru OpenCV*

Obr. 30: Chybné výřezy obličejů

6.2.2 Doba zpracování databáze

Doba zpracování detekce z videozáznamu se ukázala pravděpodobně jako největší problém. Jeden snímek je v průměru zpracováván téměř 1,5 sekundy, u 45 minutového záznamu s 25 snímky/s se tak doba zpracování protáhne přes 24 hodin. Výhodou je, že uživatel po tu dobu nemusí do procesu vůbec zasahovat a doba zpracování nevyžaduje žádnou lidskou práci. Zpracování Událostí trvalo 18:29 hodiny, v průměru tak jeden snímek trval 1,48 s. S ohledem na celkový počet snímků byla nejrychleji zpracována *Rozmarná léta českého filmu.* Zpracování trvalo 28:27 hodiny a jeden snímek byl tak v průměru zpracováván jen 1,36 s. Nejpomaleji probíhala detekce u hokejového utkání, kdy zpracování narostlo na 28:54 hodiny a jeden snímek tak aplikaci zabral 1,63 s.

K úplné anotaci dat bylo uživateli předloženo 6 623 kandidátů, výsledkem je 6 466 snímků obsahující lidské tváře, ke kterým jsou uloženy informace o pohledu osob na jednotlivých snímcích, souřadnice obou očí a koutků úst. Zpracování všech snímků trvalo uživateli bezmála 13 hodin práce,

v průměru lze tedy zvládnout označit 500 snímků za hodinu, což představuje 3,6 s práce na jednom snímku. Za necelé 4 s by pravděpodobně člověk nebyl schopen se zaměřit na 4 body v obraze, označit je kliknutím a potvrdit směr pohledu zobrazené osoby. U mnoha snímků bylo využito automatického zachycení označení očí a úst. Po získání více zkušeností a cviku uživatele by se mohla doba zpracování ještě zkrátit.

Snímky z dalších videozáznamů byly jenom potvrzovány v případě výskytu tváře v obraze. Tato metoda je výrazně rychlejší a řazením snímků detekovaných ve stejné lokaci v navazujících snímcích lze rychle projíždět rozsáhlé bloky dat bez zdlouhavého zkoumání, které by si vyžádala častá změna předkládaných snímků. Zpracování snímků ze zpravodajského pořadu trvalo 2:18 hodiny, předloženo bylo 20 707 snímků a průměrná rychlost procesu je 2,5 snímku/s. Databáze vytvořená ze záznamu dokumentárního pořadu byla zpracována za 3:33 hodiny. Při celkovém počtu předložených snímků v počtu 39 627 vychází rychlost zpracování na 3,1 snímků za sekundu. Toto už je téměř hraniční rychlost zpracování, která se blíží rychlosti zpracování obrazových dat.

6.3 Hodnocení výsledků

Srovnání vytvořené databáze s dostupnými alternativami je docela obtížné. Přístup k jejímu vytvoření je zcela odlišný od běžně pořizovaných datových sad. Při vytváření takové sady má tvůrce představu jaký typ dat chce vytvořit a jakou úlohu potřebuje strojově učit. V případě dat z videa je vždy trochu překvapením jaké budou ve skutečnosti výsledné snímky a kolik jich bude vhodných pro učení. Vzniklá datová sada ale netrpí jednoúčelovostí a pro učení systému používaných v reálném světě může přinést užitečná data.

Tato databáze se nemůže srovnávat kvalitou snímků s databázemi vytvořenými digitálními fotoaparáty, které pořizují snímky v mnohem větší kvalitě a rozlišení. Naopak vytvoření takto rozsáhlé databáze ze statických obrazů je velmi časově náročné.

Srovnáním specifických snímků nalezených ve videu s dostupnými sadami vychází použití videa jednoznačně úspěšněji. Snaha o pitvoření se do objektivu nebo maskování brýlemi se nevyrovná reálným lidským reakcím na emotivní události (viz Obr. 31) a rovněž způsoby částečného překrytí obličeje může obsahovat nepřeberné množství variací.

Obr. 31: Emoční fotky

Databáze obsahuje i snímky, které by v laboratorních podmínkách vznikaly jen těžko. Detektory vyhodnotily jako tvář i záběry na obličej, který je již prolínán další scénou. Dalšími specifickými snímky mohou být černobílé záběry, nebo snímání starých fotografií či obrazů. Jako lidský obličej detektory vyhodnotily i hlavy soch, osoby schované za foťákem či kamerou a mnoho dalších. Několik příkladů je uvedeno na Obr. 32.

Obr. 32: Netradiční snímky obličejů

Závěr

Úkolem této práce bylo pořízení rozsáhlé databáze lidských obličejů. Bylo nutné se seznámit s metodami pro detekci a rozpoznávání tváří, které jsou založeny především na metodách strojového učení. Pro trénování těchto algoritmů by měla být výsledná databáze určena.

V dnešní době existuje mnoho sad vhodných pro tato trénování. Většinou se ale jedná o data pořízená v laboratorních podmínkách a nemusejí proto vyhovovat pro učení robustních algoritmů fungujících v reálném světě. Zároveň jsou dostupné databáze jednosměrně zaměřené. Buď neobsahují dostatek typů lidských tváří (databáze obsahují obličeje od lidí stejné národnosti) nebo obsahují omezené množství pohledů a situací v kterých se snímaní lidé nacházejí. Databáze vytvořená z veřejně dostupných videozáznamů odstraňuje některé tyto nedostatky a dovoluje v relativně krátkém čase pořízení rozsáhlé databáze (až v řádu desítek tisíců) obličejů. Rozmanitost a velikost výsledné databáze jsou částečně ovlivněny i použitými zdroji videa. Některé alternativní zdroje pro snadné získání záznamu jsou v práci rovněž zmíněny.

Pořízení nové databáze by nemělo využívat složitě trénovaných algoritmů. K získání dat by mělo být použito jednodušších přístupů, které získají naprostou většinu potřebných snímků i s tím, že množina nesprávných snímků bude poměrně velká. Jako vhodné algoritmy detekce obličejů v obraze se jeví metody založené na barvě lidské kůže nebo kaskádový klasifikátor založený na algoritmu Viola-Jones. Na těchto principech byly vytvořeny detektory, které ze zpracovávaného videa získávají snímky lidských obličejů v nejrůznějších polohách a výrazech, které by byly v laboratorních podmínkách jen stěží simulovatelné. Z hodinového záznamu lze získat až 35 tisíc tváří, na jejichž anotování prostřednictvím vytvořené aplikace by bylo potřeba asi 70 hodin lidské práce. Vznikne tak ale rozsáhlá databáze s lokalizovanou pozicí očí a rtů u každého snímku obličeje. V rámci práce byla vytvořena takováto databáze o velikosti téměř 6 500 obličejů z netradičního prostředí. Dalších více jak 50 000 snímků obličejů bylo detekováno a schváleno jako snímky obsahující lidskou tvář.

Další vývoj aplikace by měl směřovat ke zrychlení detekce tváří ve videu, pravděpodobně implementováním rychlejšího algoritmu lokalizace nalezených obličejů po detekci na základě barvy kůže. Vylepšením mohou být i implementace dalších detektorů, ať už pro samotnou detekci obličejů, nebo pro detekci očí a úst. Alternativně by mohly být přidány i detektory rozpoznávající další části lidského obličeje.

Literatura

[1] Honzík, P.: Strojové učení., 2006. 85 s.

[2] Duda R., Hart, P., Stork, D.: Unsupervised Learning and Clustering, Ch. 10 in Pattern classification: Wiley. 2001. 571 s. ISBN 0-471-05669-3, 2001.

[3] Ethem, A.: Introduction to Machine Learning. [s.l.] : [s.n.], 2004. 445 s. ISBN 978-0-262-01211-9.

[4] Přinosil, J., Krolikowski, M.: Využití detektoru Viola-Jones pro lokalizaci obličeje a očí v barevných obrazech. Elektrorevue - Internetový časopis (http://www.elektrorevue.cz), 2008, roč. 2008, č. 31, s. 1-16. ISSN: 1213-1539.
Dostupný z WWW:
<http://www.elektrorevue.cz/file.php?id=200000238-97bc998b67>

[5] Sung, K.K.; Poggio, T.: Example-Based Learning for View-Based Humna Face Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 1, Jan. 1998. 39-51 s.

[6] Štanclová, J.: Template matching, Strukturální rozpoznávání I. Praha: KSI MFF UK Praha, 23.4.2009. 23 s.

[7] Singh K., Chauhan, S., Vatsa, M.: A Robust Skin Color Based Face Detection Algorithm Tamkang Journal of Science and Engineering. 2003, vol. 6, no. 3

[8] Žára, J.; Beneš, B.; Sochor, J.; Felkel, P.: Moderní počítačová grafika, Computer Press, 2004, ISBN 80-251-0454-0.

[9] Yang, G.Z.; Huang, T.S.: Human face detection in a complex background, Pattern Recognition, 1994. 53-63 s.

[10] Blázsovits, G.: Interaktívna učebnica spracovania obrazu. Bratislava: Fakulta matematiky, fyziky a informatiky Univerzity Komenského v Bratislavě, Katedra aplikovanej informatiky, první vydání, 2006. URL: <http://dip.sccg.sk/>

[11] Burger, W.; Burge M.: Digital Image Processing:An Algorithmic Introduction Using Java. Springer, 2007.

[12] Campadelli, P.; Cusmai, F.; Lanzarotti R.: A Color-Based Method For Face Detection, Dipartimento di Scienze dell‘ Informazione, Universita degli Studi di Milano.

[13] Chaloupka, J: Rozpoznávání akustického signálu řeči s podporou vizuální informace, dizertační práce Technická univerzita v Liberci, 2005.

Seznam obrázků

Seznam tabulek