VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

Semestrální projekt

2006 **Bc. Pavel Hendrych**

Prohlašuji, že jsem tento semestrální projekt vypracoval samostatně pod vedením Ing. Miroslava Švuba. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal. Zároveň bych na tomto místě rád poděkoval vedoucímu projektu Ing. Švubovi za cenné připomínky a odborné vedení.

V Brně dne 2. ledna 2007 Bc. Pavel Hendrych

Abstrakt

Tato práce se teoreticky zabývá současnými přístupy k detekci a lokalizaci obličeje v obraze. Popisuje možné přístupy k řešení tohoto problému, detailněji se zabývá lokalizací pomocí neuronových sítí a operacemi před vlastní detekcí a následné správné reprezentaci výsledků. Obsahuje též návrh systému k detekci obličeje, který bude implementován v rámci diplomové práce.

Klíčová slova

Počítačové vidění, lokalizace obličeje, perceptron, neuronová síť, PCA, obličejová databáze

Abstract

This thesis issues with possible methods for face detection and localization according to the state of the art. It describes various approaches and it is aimed at localization by neural networks and at necessary operations that have to be done before detection and after it for correct results representation. It contains also a detection system design, which will be implemented later as a part of my diploma thesis.

Keywords

Machine Vision, face localization, perceptron, neural network, PCA, principal component analysis, face database

Obsah

Úvod ... 6
Neuronové sítě ... 7
Umělý neuron ... 7
Učení perceptronu ... 8
Zapojování neuronů do sítí ... 9
Vybrané metody zpracování obrazu pro potřeby detekce obličeje ... 11
Ekvalizace histogramu ... 12
Normalizace ... 13
Změna měřítka ... 14
Filtrace obrazu ... 14
Analýza hlavních komponent ... 16
Současné přístupy k detekci obličeje ... 18
Základní požadavky na systém ... 18
Metody založené na znalostech ... 19
Příznakové invariantní metody ... 20
Metody založené na šablonovém porovnávání ... 21
Metody založené na vzhledu ... 22
Návrh systému pro detekci obličeje ... 24
Motivace ... 24
Předzpracování vstupních dat ... 24
Návrh klasifikátoru ... 25
Neuronová síť ... 25
Trénovací množina ... 25
Zpracování výsledků ... 26
Eliminace duplikátů ... 26
Blokové schéma systému ... 27
Očekávaný výstup systému ... 27
Závěr ... 28
Použitá literatura a odkazy ... 29

Úvod

Původně byly počítače vytvořeny jako stroje pro automatizované zpracování dat. Během své existence prošly velmi rychlým vývojem a již po mnoho let nejsou jen užitečnými kancelářskými pomocníky, ale přidáváním nových, důmyslných funkcí se z nich vyvinuly velmi vážní konkurenti lidské pracovní síly. V některých oblastech dokonce člověka porážejí. Není to způsobeno pouze nárůstem výkonu, ale také důmyslnějším procesem zpracování dat, od jejich pořízení do vykonání příslušné akce. Zdrojem dat již není jen klávesnice, data mohou být pořízena z nejrůznějších senzorů, jako jsou například pohybová čidla, sonary atd.
V posledních několika letech zaznamenaly značný vývoj kamery. Již v 70. letech 20. století, kdy byly počítače schopné zpracovávat 2D obrazy, se začal vyvíjet nový obor – Počítačové vidění (angl. Computer Vision), avšak dnes se v souvislosti s průmyslovými roboty vybavenými různými senzory používá pojem Machine Vision. Computer vision se zaměřuje spíše na rekonstrukci 3D informace z kamer (avšak tyto obory se značně překrývají). Je tedy přirozená snaha využít možnosti počítačových systémů o tyto optické senzory. Naneštěstí jsou takto získaná data dvourozměrná a velmi často zatížená šumem. To klade velký důraz na následné zpracování získaného obrazu, kde už si s klasickými přístupy často nevystačíme.

Tato práce se zabývá jednou z fází zpracování obrazu, konkrétně detekcí obličeje ve snímku. První a druhá část rozebírá základní pojmy z oblasti umělé inteligence a počítačového vidění. Třetí část shrnuje současný stav použitelných metod. Z nich jsou vyzvednuty ty, které se zabývají neuronovými sítěmi. Závěrečná kapitola popisuje návrh systému pro detekci, který bude implementován v rámci diplomové práce. Zabývá se různými praktickými aspekty detekčního systému a formuluje požadavky na implementaci.

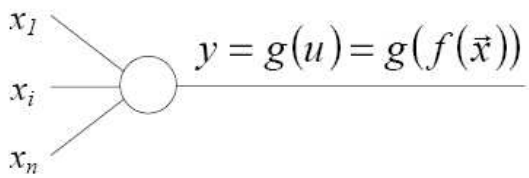
Práce je ukončena sumarizací poznatků a nástinem možného rozšíření implementovaného systému.

Neuronové sítě

Umělé neuronové sítě patří do oblasti umělé inteligence zvané soft-computing. Ten, na rozdíl od konvenčních algoritmů, toleruje data, která mohou obsahovat nepřesnosti, částečnou pravdivost či aproximace, za účelem dosažení řešitelnosti daných problémů. Neuronové sítě jsou jedním z výpočetních modelů soft-computingu, stejně jako fuzzy logika, evoluční algoritmy a teorie chaosu. Model může být implementován jak softwarově, tak i hardwarově. Výhodou těchto přístupů je, že při vhodné aplikaci na některé problémy poskytnou řešení rychleji, byť někdy jen částečné, a to právě v případech, kdy klasické algoritmy selhávají zcela. Důvodem selhání je většinou nedostatek paměti nebo času.

Umělý neuron

Inspirací pro vznik umělých neuronových sítí byly znalosti z biologie. Matematicky byl poprvé umělý neuron popsán v roce 1943, ovšem princip činnosti byl znám již z konce 19. století. Činnost umělého neuronu je podobná jako v případě biologického, pouze je abstrahováno od pro funkci umělého neuronu nepodstatných vlastností, např. synapse je modelována číslem představujícím váhu, nikoliv množstvím proteinů, iontů a nervových přenašečů, jak je tomu u biologického neuronu. Nejjednodušším modelem neuronu je perceptron, poprvé představeným Frankem Rosenblattem v roce 1958. Perceptron má následující strukturu:



Obrázek 1: Schéma perceptronu

Vstupy neuronu jsou označeny x_i, váha vstupu x_i je w_i. Funkce f(**x**) se nazývá bázová funkce. Lineární bázová funkce je definována rovnicí 1, radiální bázová funkce má tvar dle rovnice 2. Samozřejmě lze použít i jiné typy bázových funkcí, ale s lineární a radiální se můžeme setkat nejčastěji.

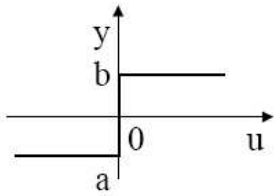
$$u = \sum_{i=i}^{N} w_i x_i \qquad (1)$$

$$u = \left\| \vec{x} - \vec{w} \right\| = \sqrt{\sum_{i=i}^{N} (x_i - w_i)^2} \qquad (2)$$

Výstup neuronu je definován aktivační funkcí g(u) = g(f(**x**)). Tabulka 1 znázorňuje některé z používaných aktivačních funkcí. Aktivační funkce může být bipolární, v tom případě nabývá hodnot {–1,1}, nebo binární, jejímž výstupem jsou {0,1}, a v případě spojitých funkcí jsou pak výstupem hodnoty z intervalu <0,1>, respektive <-1,1>.

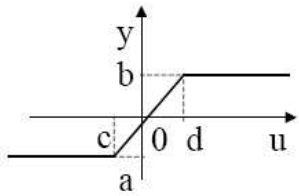
Skoková aktivační funkce

$$y^{new} = \begin{cases} a & pro\ u < \Theta \\ b & pro\ u > \Theta \\ y^{old} & pro\ u = \Theta \end{cases}$$



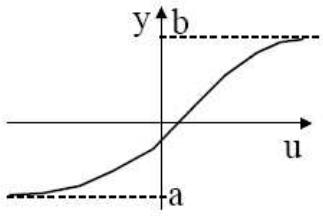
Po částech lineární aktivační funkce

$$y^{new} = \begin{cases} a & pro\ u < c \\ b & pro\ u > d \\ a + \dfrac{(b-a)(u-c)}{d-c} & pro\ c \le u \le d \end{cases}$$



Sigmoidální funkce

$$y = a + \frac{b-a}{1+e^{-\lambda u + c}}$$



Tabulka 1: Definice a průběhy často používaných aktivačních funkcí

Učení perceptronu

Základní pravidlo učení perceptronu má tvar dle rovnice 3. Výstup neuronu je označen y, d je požadovaná hodnota výstupu pro vektor **x**. Koeficient učení p má nejčastěji hodnotu 0,5, nebo může být adaptivně snižován během cyklu učení. Proměnná k značí k-tý vstupní vektor.

$$\vec{w}(1) = random$$
$$\vec{w}(k+1) = \vec{w}(k) + p(d(k) - y(k))\vec{x}(k) \tag{3}$$
$$pro\ k = 1,2..N, \quad (\vec{x}(k), d(k)) \in T$$

Perceptron dokáže klasifikovat vstupní obrazy do dvou tříd. Tyto třídy musejí být lineárně oddělitelné, tj. v prostoru obrazů musí existovat nadrovina, která obrazy dokáže rozdělit. Nadrovina je utvar v n-rozměrném prostoru o dimenzi n-1. Dá se matematicky dokázat, že buď tato nadrovina neexistuje, nebo existuje nekonečně mnoho dělících nadrovin. V případě radiálních bázových funkcí je pak dělícím útvarem hypersféra. Při vhodně zvoleném koeficientu učení p vstupní váhy neuronu konvergují k jednomu z řešení.

Zapojování neuronů do sítí

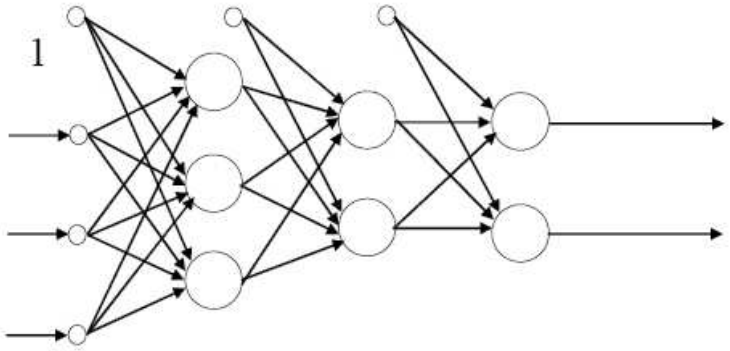
Neuron ve smyslu samostatné výpočetní jednotky nepřináší žádná praktická zlepšení výpočtu. Toho však dosahujeme spojováním většího počtu neuronů do struktur zvaných neuronové sítě. Síť dovoluje modelovat mnohem složitější separační oblasti. Dělícím útvarem není nadrovina (hypersféra), ale několik nadrovin (hypersfér). Každá síť se skládá z několika vrstev. Jde o vrstvu vstupní (ovšem i v odborné literatuře neexistuje shoda, zda se má díky své jednoduchosti také nazývat vrstvou), několik vrstev skrytých, které nejsou viditelné uživatelem, a vrstva výstupní, která slouží k prezentaci výsledků. Neuronové sítě lze klasifikovat podle několika různých hledisek, např. podle stupně propojení, směru propojení mezi jednotlivými vrstvami apod. Pro každý typ sítě lze najít specifickou oblast použití a hlavně odpovídající algoritmus učení. Typickým příkladem učícího se algoritmu je zpětné šíření chyby (angl. Back Propatation, BP), navržený pro stejně pojmenovanou dopřednou plně propojenou neuronovou síť. Jde o učení s učitelem, které je odvozeno z myšlenky minimalizace energetické funkce definované rovnicí 4, podoba vzorce pro učení má pak tvar dle rovnice 5.

$$E = \frac{1}{2}\sum_{i=1}^{N}(y_i - d_i) \tag{4}$$

$$\Delta^l w_i = \mu^l \ \delta_j^l x_i \tag{5}$$

kde w_i je váha i-tého neuronu, μ je koeficient učení a ${}^l\delta_j = -\frac{\partial E}{\partial\, {}^l u_j}$

Možnost nasazení neuronových sítí je poměrně široká v oblastech spadajících do soft-computingu. Lze je použít jako asociační paměti, v různých oblastech rozpoznávání obrazů, či například shlukování. Možnosti nasazení neuronových sítí v jedné z oblastí rozpoznávání – detekci lidského obličeje ve fotografiích, jsou popsány v následující kapitole. Neuronové sítě jsou použity jako klasifikátory, které klasifikují vstupní data do tříd obličej/neobličej. Neuronové sítě jsou však pouze jednou z možností, jak vytvořit klasifikátor, nikoli však jedinou. Alternativou k neuronovým sítím je například Support Vector Machines, které jsou krátce zmíněny dále.



Obrázek 2: Dopředná neuronová síť se dvěmi skrytými vrstvami

Vybrané metody zpracování obrazu pro potřeby detekce obličeje

Jak jsem již uvedl v úvodu této práce, pojem „počítačové vidění" je velmi široký. V této práci se budu zabývat počítačovým viděním z pohledu zpracování obrazu jako dvourozměrného signálu a následným rozpoznáváním.

Zpracování obrazu vyžaduje specifické modely zpracování. Nelze použít postupy známé z klasického zpracování informací, kde data jsou získána jako jednoznačné hodnoty ze senzorů nebo vstupem z klávesnice od uživatele. Před vlastním zpracováním je nutno provést některé kroky pro transformaci vstupního obrazu, jako je například filtrace. V případě rozpoznávání nebo jakéhokoli zpracování obrazu obličeje je velmi důležitou částí sytému detekce místa v obraze, kde se obličej nachází. Detekce je velmi náročný úkol, protože vstupní obrazy vykazují značnou variabilitu např. ve výrazu obličeje, velikosti, orientaci apod. Detekce pracuje s libovolným obrazem. Cílem detekce je určit, zda se v obraze nacházejí nebo nenacházejí nějaké obličeje. V případě pozitivního nálezu je výstupem systému umístění obličeje ve vstupním obraze a jeho rozměr [2].

Variabilita vstupních dat způsobuje jisté problémy, které lze rozdělit do několika kategorií, převzatých z [2]. Následující aspekty mají značný vliv na výkonnost detektoru a při návrhu detektoru je potřeba na ně brát ohled.

- Pozice vzhledem ke kameře: Teoreticky může být obličej natočen ve všech třech osách relativně k pozici kamery. Hlava na snímku může být předkloněna nebo zakloněna, ukloněna na obě strany nebo otočena podle středy obličeje. To může způsobovat absenci některých obličejových rysů jako např. očí, což může činit některé metody detekce nepoužitelnými.
- Variabilita obličejových rysů: Může se jednat například o presenci či absenci brýlí nebo vousů. V případě vousů, kníru a vlasů je potřeba vzít v úvahu častou změnu tvaru, délky a v případě vlasů též barvu.
- Výraz obličeje: Změna nálady může vyvolat změnu tvaru některých obličejových částí, především úst a oblasti kolem očí.
- Částečné zakrytí: Některé části obličeje mohou být zakryty jinými objekty, v případě skupinové fotografie též jinými obličeji.

- Podmínky při snímání obrazu: Obraz obličeje ve snímku je ovlivněn různými světelnými podmínkami (intenzita, směr světla atd.) nebo nastavením objektivu kamery (ohnisková vzdálenost, polarizační filtr atd.)

Výčet faktorů, které mohou daný obraz ovlivnit, je poměrně široký a je potřeba před návrhem systému vyšetřit, které z nich mohou nastat, a zvolit vhodnou metodu pro detekci. Používané metody detekce jsou popsány dále. Přesto je vhodné před vlastní detekcí provést předzpracování obrazu, kterým zajistíme maximální uniformitu zpracovávaného obrazu. Do předzpracování lze zahrnout následující metody:

- Ekvalizace histogramu: Ekvalizací myslíme úpravu kontrastu ve snímku nebo jeho části tak, aby nebyla změněna globální intenzita celého snímku.
- Normalizace: Hodnoty barev jsou převedeny do intervalu <0,1>.
- Změna měřítka: Snímek je zmenšen či zmenšován tak, aby jeho velikost přesně vyhovovala požadavkům použité klasifikační metody.
- Filtrace obrazu: K rozpoznávání je použita pouze část frekvenčního spektra snímku.

Další fází zpracování obrazu je vlastní detekce obličeje. Souhrn metod použitelných pro detekci je popsán dále. Výsledek těchto metod však nemusí vždy odpovídat požadavkům, proto je často nutné výsledky detekce vhodně prezentovat. Možností, jak správně prezentovat dosažené výsledky, může být například eliminace duplikátních nálezů.

Ekvalizace histogramu

Jedná se o úpravu kontrastu barev ve vstupním snímku. Používá se především pro obrazy reprezentované stupni šedi, ale je možno tuto techniku použít i na barevné obrazy. Pomocí rozložení barev v histogramu dokáže lépe distribuovat frekventované hodnoty intenzit, čímž lze dosáhnou zvýšení kontrastu v málo kontrastních oblastech, aniž by došlo ke změně globálního kontrastu vstupního snímku. Je tedy velmi vhodné ji použít v případech, kdy barva pozadí i barva objektu v popředí je v obou případech velmi světlá, nebo velmi tmavá. Dokáže tedy zlepšit kvalitu přeexponovaných nebo podexponovaných snímků v důsledku změny světelných podmínek nebo nesprávného nastavení kamery. Pro snímek ve stupních šedi je pravděpodobnost výskytu bodu s hodnou šedi definována jako

$$p(x_i) = \frac{n_i}{N}\ ,\quad i=0\ ..\ L-1 \tag{6}$$

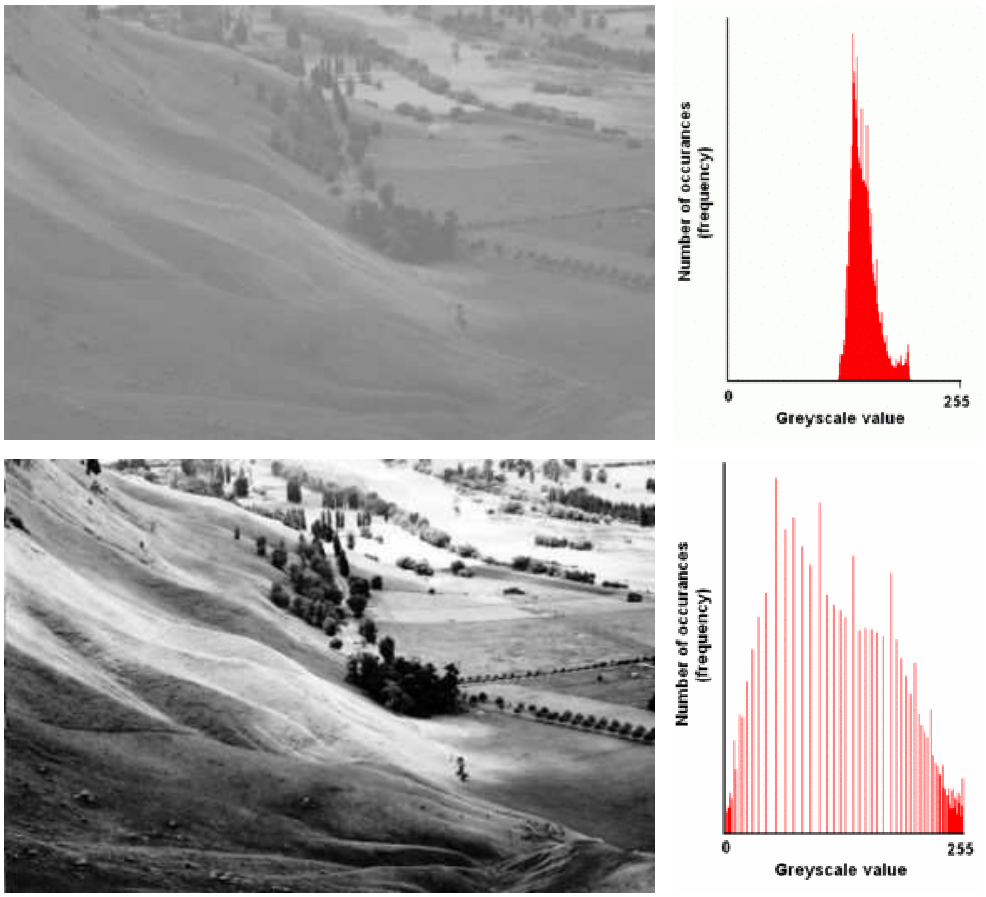
kde L je počet stupňů šedi v obraze a n je celkový počet bodů obrazu. Kumulativní pravděpodobnostní funkce je pak definována jako

$$c(i) = \sum_{j=0}^{i} p(x_j) \tag{7}$$

kde c je normalizovaný histogram obrazu. Ekvalizace histogramu je pak definována následujícím způsobem

$$y_i' = y_i(\max - \min) + \min \tag{8}$$

V případě potřeby ekvalizovat barevné snímky, zle ekvalizaci provést zvlášť pro všechny barevné složky.



Obrázek 3: Ukázka ekvalizovaného a nekvalizovaného snímku, včetně histogramu. Převzato z http://en.wikipedia.org/wiki/Histogram_equalization

Normalizace

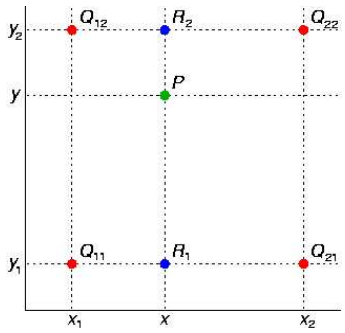
Jedná se o změnu kódování, zejména pro potřeby neuronových sítí. Barvy vstupního obrazu jsou reprezentovány reálnými čísly v intervalu <0,1>. Změna kódování se provede

dělením hodnoty barvy bodu maximálním číslem, které je možno zobrazit na daném bitovém rozsahu.

Změna měřítka

Vzhledem k tomu, že většina metod je navržena pro detekci obličeje určitých rozměrů, je nutno postupně zmenšovat vstupní snímek. Obličej je pak hledán na snímku v několika měřítkách, aniž by různá absolutní velikost obličeje vyžadovala různá nastavení systému. Tak je to prezentováno například v [3].

Naneštěstí postupné vynechávání řádků či sloupců obrazu vede ke značné degradaci kvality obrazu. Proto je nutno použít jiné techniky, které ve výsledném zmenšeném obraze interpolují barvy původních, vynechaných bodů. Barvy bodů, které mají v obraze zůstat, se změní na základě váženého průměru z okolních vynechaných bodů. Váhou může být vzdálenost středu bodů od přesně vypočítaného místa, kde se má vzorkovat. Barvu bodu P pokud známe barvu bodů Q_{ij} vypočítáme např. dle postupu uvedeného na v rovnici 9. Nejjednodušší interpolací je bilineární, složitější pak bikvadratická a bikubická.



$$f(R1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{1,1}) + \frac{x - x_1}{x_2 - x_1} f(Q_{2,1}) \quad kde\ R = (x, y_1)$$

$$f(R2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{1,2}) + \frac{x - x_1}{x_2 - x_1} f(Q_{2,2}) \quad kde\ R = (x, y_2) \qquad (9)$$

$$f(P) = \frac{y_2 - y}{y_2 - y_1} f(R1) + \frac{y - y_1}{y_2 - y_1} f(R2)$$

Obrázek 4: Výpočet barvy pomocí bilineární interpolace, převzato z http://en.wikipedia.org/wiki/Bilinear_interpolation

Filtrace obrazu

Ne vždy je vhodné zpracovávat obraz ve formátu, v jakém byl pořízen z kamery, protože zpravidla obsahuje velké množství informací, které jsou pro daný účel nepodstatné a mohou stěžovat výběr informací podstatných. Proto se často obraz před vlastním zpracováním často filtruje. Výsledný obraz je konvolucí vstupního obrazu a vlastního filtru. Každý pixel výsledného obrazu je vypočítán jako součet součinů bodu obrazu se středem

filtru a jeho sousedů s příslušnými sousedy v matici filtru. Právě kvůli zajištění existence středu filtru má jeho matice liché rozměry, nejčastěji 3x3 nebo 5x5 bodů.

Filtrace je vhodná například pro zaostření obrazu či detekci hran v oblasti prostoru barev zejména v případech, kdy je velikost filtrační matice malá. V případech, kdy je zapotřebí použít filtr větších rozměrů, operace konvoluce značně zvyšuje výpočetní náročnost, proto se často používá filtrace ve frekvenční oblasti. Často používanou operací je Fourierova transformace, pro zpracování obrazů pak zejména rychlá Fourierova tranformace (FFT), která vypočítá frekvenční spektrum. Ve frekvenčním spektru je pak možno aplikovat požadované filtry, například horní propust (HPF, High Pass Filter) pro detekci hran nebo dolní propust (LPF, Low Pass Filter) pro efekt rozmazání obrazu. Po rekonstrukci obrazu pomocí inverzní Fourierovy transformace pak dostaneme žádaný filtrovaný obraz.

Často používaným filtrem ve frekvenční oblasti v oblasti počítačového vidění je Gaborův filtr. Používá se zejména k segmentaci obrazu. Jedná se o lineární tzv. pass band filtr, který zesiluje frekvence v daném frekvenčním zespoda i shora ohraničeném pásmu definovaný jako:

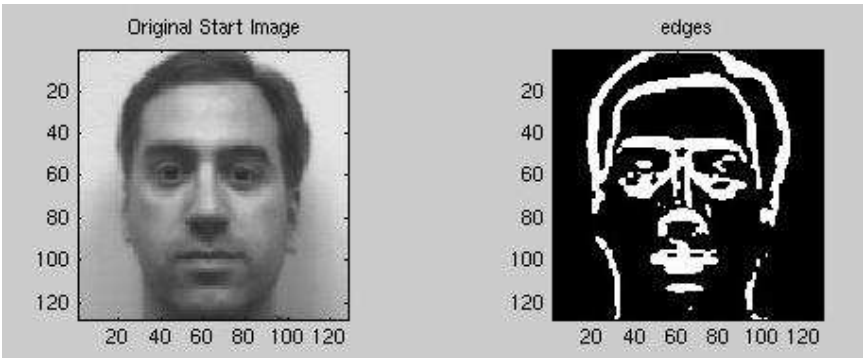
$$g(x, y : \lambda, \Theta\psi, \sigma, \gamma) = \exp\left(\frac{-x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \psi\right)$$

kde (10)

$$x' = x * \cos\Theta + y\sin\Theta$$

$$y' = -x * \sin\Theta + y\cos\Theta$$

kde λ reprezentuje vlnovou délku kosinového faktoru, Φ orientaci normály na paralelní pruhy Gaborovy funkce ve stupních, ψ fázové posunutí ve stupních, γ prostorový koeficient. Aplikací Gaborových funkcí získáme Gaborovy vlnky, které jsou zpravidla navrženy a předpočítány pro různé úhly otočení a šířku filtru.



Obrázek 5: Ukázka detekce hran z původního snímku, převzato z http://www.owlnet.rice.edu/~elec539/Projects97/morphjrks/moredge.html

Analýza hlavních komponent

Analýza hlavních komponent (angl. Principal Component Analysis – PCA) je často používanou metodou z oblasti statistiky pro redukci dimenze vstupních dat. Přesněji řečeno, jedná se o lineární transformaci, která transformuje data z jednoho souřadného systému do jiného tak, že v druhé dimenzi leží na první souřadnici dimenze s největší variancí hodnot v prostoru prvním, na druhé souřadnici dimenze s druhou největší variancí v prostoru prvním atd. V praxi se tak PCA použije k výběru podmnožiny složek vstupních dat tak, aby ztráta informace byla co nejmenší.

Cílem výpočtu PCA je získání vlastních vektorů a jim příslušných vlastních čísel ze vstupní matice. Vlastní čísla a vlastní vektory získáme posloupností následujících operací.

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{11}$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2} \tag{12}$$

Vstupní data jsou uložena v matici **X**, která má následující strukturu:

$$X = \begin{pmatrix} obrázek\ 1 \\ obrázek\ 2 \\ \vdots \\ obrázek\ N \end{pmatrix} \tag{13}$$

Rovnice 11 ukazuje způsob výpočtu střední hodnoty, rovnice 12 pak směrodatné odchylky. Dále předpokládáme normalizovaná data se střední hodnotou 0.

$$C^{m\times n} = (c_{i,j}, c_{i,j} = Cov(Dim_i, Dim_j)$$

kde

$$Cov(X,Y) = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{N-1} \tag{14}$$

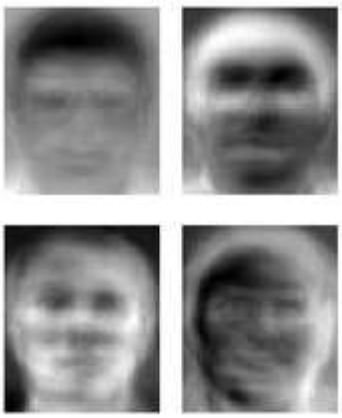
Rovnice 14 ukazuje výpočet kovarianční matice **C**. Další operací je výpočet vlastních čísel a vlastních vektorů matice X. Pro vlastní vektory platí rovnice 15, kde **A** je operátor řádu n a **x** je vlastní vektor a λ je jemu příslušné vlastní číslo. V našem případě je operátorem kovarianční matice **C**.

$$A\vec{x} = \lambda\vec{x} \tag{15}$$

Pro nalezení vlastních čísel kovarianční matice **A** převedeme matici na soustavu rovnic dle 16.

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0 \tag{16}$$

Po vyřešení této soustavy rovnic tedy známe vlastní vektory a vlastní čísla matice **C** a můžeme říci, že hlavní komponentou je vektor s největším vlastním číslem. Vlastní vektory kovarianční matice se nazývají vlastní tváře (angl. Eigenfaces). Seřadíme-li vlastní vektory podle velikosti jejich vlastních čísel, dostaneme složky seřazené podle rozptylu. Pro redukci počtu dimenzí pak stačí vzít tolik prvních složek, kolik uznáme za vhodné.



Obrázek 6: Ukázka vlastních tváří. Pro potřeby zobrazení nedošlo k redukci dimenze. Převzato z http://en.wikipedia.org/wiki/Eigenface.

Současné přístupy k detekci obličeje

Základní požadavky na systém

Detekcí obličeje rozumíme určení pozice a otočení všech obličejů nacházejících se na vstupní fotografii, nebo žádnou odezvu v případě, že se na vstupním snímku žádný obličej nevyskytuje. Ideální systém je nezávislý na invarianci ve vstupních datech, a to ve všech výše jmenovaných bodech. Detekce je většinou pouze první částí systému pro rozpoznávání, je proto nutno návrh obou částí systému synchronizovat. Například velmi precizní systém pro rozpoznávání nepotřebuje přesný detektor, ale poradí si i s větším množství neobličejových dat chybně detekovaných jako obličejové. K ohodnocení výkonnosti dané metody budeme používat náročnost na výpočetní prostředky. Úspěšnost je možno hodnotit pomocí míry chybně klasifikovaných obličejů jako neobličejová data (angl. false negatives) a chybně klasifikovaných neobličejových data jako obličeje (false positives). Pro detekci se v současné době používají různé modifikace metod, které lze shrnout do čtyř základních přístupů, které jsou popsány na několika následujících stranách. Bylo představeno přibližně 150 různých přístupů k detekci, které často využívají znalosti a postupy z více než jedné z níže popsaných kategorií. Všechny se primárně zabývají detekcí obličejů ze statických snímků (tedy nikoliv z videa), předpokládají snímky ve škále šedých odstínů a předpokládají též, že obličeje jsou natočeny přímo na kameru a normála roviny obličeje je opačná oproti normále snímku (angl. frontal upright face detection). U některých metod je diskutováno možné rozšíření i na detekci otočených obličejů. Následujících několik stran poskytuje základní přehled metod, přičemž jsem čerpal převážně z [2]. Je nutno podotknout, že hranice mezi nimi nejsou pevné a některé se vzájemně překrývají.

- Metody založené na znalostech (knowledge-based methods): Metoda předpokládá zakódování lidských znalostí o prvcích, ze kterých se obličej typicky skládá, a popisu jejich vzájemných relací.
- Příznakové invariantní metody (feature invariant approaches): Metoda se soustředí na nalezení strukturálních příznaků. Tyto příznaky se vyskytují v obličeji nezávisle na jeho otočení či světelných podmínkách.
- Šablonové porovnávání (template-matching methods): Metoda předpokládá existenci šablon, které popisují obličej jako celek nebo jeho jednotlivé části. Detekce se pak provádí výpočtem korelace mezi vstupním obrazem a šablonami. Tuto metodu lze použít i následně ve vlastním rozpoznávání.

- Metody založené na vzhledu (appearance-based methods): Metoda je velmi podobná jako předchozí, avšak modely (šablony) nejsou předem dodány. Znalosti o modelech jsou získány (naučeny) z tréninkové množiny vstupních obrazů, u nichž je dodána správná klasifikace. Úspěšnost metody je pak silně ovlivněna složením testovací množiny.

K jednotlivým metodám bylo vypracováno množství studií, jejichž stručný přehled lze nalézt v [2].

Metody založené na znalostech

Jak již název metody naznačuje, metoda předpokládá množství apriori znalostí o lidském obličeji. Nepracuje se šablonou, ale se sadou pravidel, které obličej obecně popisují. Vstupní snímek se pak prohledává pomocí okna dané velikosti a v každém okně se hledají oblasti, které vyhovují zadaným pravidlům. Pravidla mohou mít tvar například:

- střed obličeje má podobné hodnoty intenzity
- obličej se skládá z dvou očí, které jsou symetrické, nosu a úst, které se nacházejí v prostoru pod očima, může být též definována průměrná relativní vzdálenost mezi těmito obličejovými prvky [2].

Je poměrně jednoduché zavést takováto pravidla, ovšem vzhledem k vysoké variabilitě obličejů není již tak jednoduché takto zavedená pravidla správně aplikovat. Nejprve je tedy nutno zvolit vhodný kompromis mezi výkonností systému a precisností pravidel. Obecně platí, že precisně zavedená pravidla poskytují nižší míru neobličejových částí detekovaných jako obličeje, ale některé obličeje nemusí systém vůbec detekovat.

Zvýšení výkonnosti metod založených na pravidlech přináší vícestupňová aplikace pravidel. Jednou možností je podvzorkovávání a aplikací různých pravidel na různém stupni podvzorkování. Tento hierarchický model přináší kromě snížení počtu chyb též snížení nároků na výkon počítače. Jinou alternativou je zjištění přítomnosti obličejových rysů například pomocí statistických metod. K tomu může sloužit histogram rozložení intenzity v právě zpracovávaném okně a použití pravidel pouze k validaci, zda takto rozložená intenzita je náhodná a odpovídá pozadí nebo zda signalizuje přítomnost obličeje. Pravidla určená pro validaci používají již velmi specifické obličejové prvky, jako je například poloha obočí. Metoda selhává při příliš komplexním pozadí nebo u fotografií vyššího počtu osob. Nevýhodou je též nemožnost rozšíření na detekci pootočených obličejů.

Příznakové invariantní metody

Na rozdíl od metod založených na znalostech, metody založené na příznacích, se nesnaží hledat pomocí předem známých informací strukturální příznaky, které potvrdí existenci obličeje. Naopak se snaží jednoduchými cestami nalézt množinu příznaků a vyšetřit, zdali se může jednat o obličej. Nalezení příznaků tedy není řešením, ale výchozím bodem. Vychází z metod umělé inteligence pro příznakové rozpoznávání obrazů a předpokladu, že pokud člověk dokáže různé objekty v různých pozicích jednoznačně identifikovat, musí existovat nějaká invariantní charakteristika těchto objektů. Nalezení těchto příznaků jako například obočí není složitou záležitostí a lze ji provést například detekcí hran pomocí filtrace v barevném (intenzitním) nebo frekvenčním spektru snímku.

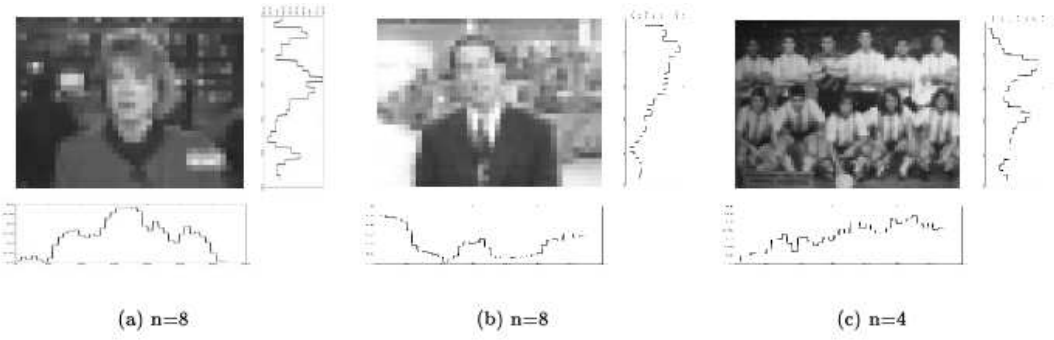
Opět bylo představeno několik konkrétních metod pracujících na základě příznaků. Všechny lze rozdělit do dvou fází. První fází je nalezení příznaků, které mohou být částí obličeje. Jednou z metod je filtrování pomocí pásmových filtrů. Pásmové filtry odstraňují frekvence vyšší i nižší než je dané pásmo, pro které je filtr navržen.

Další z možností je vytváření tzv. mapy hran. Systém nejprve nalezne hrany a pak postupně odebírá ty, které nemohou být součástí obličeje na základě seskupování. Pravidla pro jednotlivé skupiny jsou předem známa.

Vzhledem k možnosti zašumění vstupních dat jako jsou například nepříznivé světelné podmínky, je nutno před vlastní aplikací metod tyto nepříznivé jevy potlačit, například pomocí ekvalizace histogramu, který je definován výše. Obecně nejsou tyto metody příliš často nasazovány, a i když je úspěšnost těchto metod poměrně vysoká pro nějakou konkrétní databázi obličejů, ukazuje se, že lze nalézt a upravit jiné metody, které pro daný účel vykazují úspěšnost vyšší.

Velmi zajímavá je též možnost vyjít ze specifické barvy lidského obličeje a tuto informaci použít při detekci. Metoda předpokládá snímky získané při kontrolovaných podmínkách, které nemají vliv na případné barevné zkreslení snímku. Používá se specifické schéma prohledávání snímku oknem a po aplikaci filtrů hledání oblastí obsahujících barvu obličeje a vlasů. Pro rozhodnutí, zda se v oblasti nachází obličej, lze použít neuronovou síť nebo shlukovací algoritmus, který žádané shluky vytvoří z trénovací množiny, u níž je známa správná klasifikace. Velkou výhodou těchto metod je skutečnost, že dokáží lokalizovat i obličeje otočené a dokáží si do jisté míry poradit i s tvářemi obsahujícími například brýle či vousy. Nevýhodou však je již zmíněná citlivost na světelné podmínky při snímání obrazu a různá barva vlasů, která též může způsobovat komplikace.

Další rozšíření nabízí přístup kombinující hledání několika různých rysů v obraze. Jako prvotní nástroj pro výběr kandidátních oblasti lze použít např. metody využívající barvy společně s metodami pracujícími s jednoduchými příznaky. Pro vyloučení či potvrzení kandidátů lze pak použít metody složitější, například hledání detailnějších obličejových prvků.



Obrázek 7: Detekce obličeje pomocí barvy [2]

Metody založené na šablonovém porovnávání

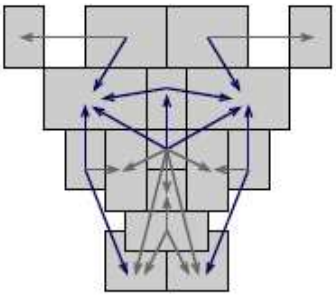
Metody jsou založené na jednoduchém principu, že obličej se dá obecně popsat pomocí jednoduchých geometrických útvarů. Vlastní šablona je pak takto popsaná sada obličejových prvků včetně jejich rozmístění. Šablona je většinou vytvořena člověkem na základě znalostí o obličeji. Vlastní detekce pak probíhá výpočtem korelace mezi vstupním snímkem (případně právě zpracovávaným oknem) a šablonou. Pokud je korelace dostatečně velká, je vstupní obraz klasifikován jako obličej.

Jako i ostatní metody však tento základní přístup není dostatečně přesný pro praktické použití. Vzhledem k vysoké variabilitě obličejů je často nutno použít ne jednu globální šablonu, ale hierarchii šablon. Na nejvyšší úrovni jsou ze vstupního snímku vybírání kandidáti splňující minimální požadavky na šabloně nejvyšší úrovně. Postupně jsou aplikovány další zpřesňující šablony jak pro celý obličej, tak pro jeho jednotlivé oblasti.

Problémem však zůstává, jak jednoznačně namapovat šablonu na vstupní obraz. Jedním z řešení je detekce hran ve snímku a následný pokus vyhledat v seznamu hran takové, kterým je možno opsat ovál představující obličej. Další možností je například pro nalezení kandidátní oblasti použití nějaké jednoduché příznakové metody, hledající pro obličej specifické rozdíly kontrastu ve vstupním snímku.

Dalším problém, který stěžuje možnost nasazení metod založených na šablonách, je skutečnost, že tyto metody se jen velmi špatně vypořádávají s transformovanými obrazy.

Různé velikosti zkoumaných obličejů, stejně jako jejich různé otočení, znemožňuje nasazení šablonových metod v systémech, kde nelze zaručit nízkou variabilitu vstupních dat. Přesto existují i metody, které dokáží pracovat i s takto částečně transformovanými obrazy. Jednou z nich může být například šablona popsaná parametrickými funkcemi. Až po nalezení prvních příznaků a na základě znalostí relativních vzdáleností mezi jednotlivými obličejovými rysy se vytvoří šablona s příslušnou velikostí, případně i natočením. U systémů s nízkou variabilitou vstupních dat je velkým plusem metod založených na šablonovém porovnání jejich vysoká úspěšnost.



Obrázek 8: Ukázka šablony pro oblasti obličeje včetně jejich vzájemných vztahů [2]

Metody založené na vzhledu

Na rozdíl od metod založených na šablonách, metody založené na vzhledu nepočítají s existencí uživatelem zadaných pravidel, ale „šablony" jsou získány učením z trénovací množiny. Právě toto spolehnutí se na výběr znaků typických pro obličej učícím algoritmem za podpory statistických metod umožňuje metodám založeným na vzhledu dosáhnout poměrně vysoké efektivnosti.

Tyto metody pracují se vstupními daty jako s náhodnou proměnnou. Vstupem klasifikátoru je pak celý obraz (okno) nebo část jeho prostoru. Může pak být klasifikován například pomocí Bayesovského klasifikátoru. Jiným přístupem je hledání nějaké funkce, která dokáže oddělit třídu obličejů od neobličejových dat. Taková funkce může mít tvar například rozdělující nadroviny a implementována může být neuronovou sítí, pomocí Support Vector Machines, ale existují i jiné způsoby implementace. Vzhledem k tomu, že dimenze vstupních dat je poměrně velká, jsou často před vlastní klasifikací aplikovány na tato data metody, které jejich dimenzi snižují. Mezi takovéto metody patří typicky výše zmíněná PCA. Právě PCA lze použít k výpočtu tzv. Eigenfaces, což je podmnožina vlastních vektorů kovarianční matice. Klasifikace se pak děje na základě výpočtu vektoru vlastních vektrorů

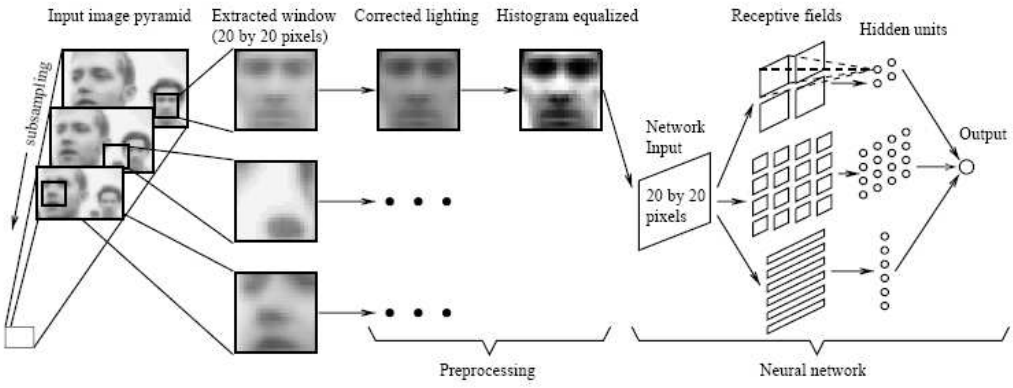
(Eigenface) pro klasifikovaný obraz a následným porovnáním vzdálenosti mezi shlukem vektorů obličejů a neobličejových dat.

Shlukování je jednou z dalších použitelných metod. Obecně nemusí docházet k redukci obrazového prostoru, pokud jsou obrazy dostatečně malé (často používanou velikostí jsou výřezy snímku oknem o velikosti kolem 20x20 bodů). V čase trénování se rozdělí trénovací množina do několika shluků pro obličeje a pro pozadí. Klasifikace probíhá na základě porovnání vzdáleností od jednotlivých shluků, případně na základě složitějších metrik. Samozřejmě i zde je možné použít PCA a prostor tak zredukovat.

Další častou architekturou pro výstavbu klasifikátoru jsou neuronové sítě. Vstupem neuronových sítí mohou být výřezy snímků, Eigenfaces či obrazy různě filtrovaných snímků. Neuronové sítě pro klasifikaci se skládají obvykle z několika skrytých vrstev. Mezi nejvýznamnější aplikace neuronových sítí patří bezesporu systém představený autory Rowley, Baluja, Kanade v [3]. Pro klasifikaci je použita několikavrstvá neuronová síť se specifickým zapojením jednotlivých vrstev. Následně byla síť rozšířena o detekci otočení obličeje, která tak umožnila detekovat i otočené obličeje.

Výhodou neuronové sítě je poměrně vysoká schopnost korektně reagovat na vysokou variabilitu vstupních dat, nevýhodou může být nutnost manuálně nastavovat její parametry. Ty totiž velmi významně ovlivňují celkovou detekční schopnost.

Dalším možným přístupem ke klasifikátoru je pomocí Support Vector Machines (SVM). Jedná se o klasifikační nástroj, který během učení hledá dělící nadrovinu mezi testovanými obličeji a pozadím tak, aby od shluků obou tříd měla co největší vzdálenost. Výpočet této nadroviny však může být velmi náročný v případech, kdy trénovací data jsou rozsáhlá a mají velký počet dimenzí.



Obrázek 9: Systém pro detekci obličeje představený v [3]

Návrh systému pro detekci obličeje

Motivace

V minulé kapitole jsem představil základní kategorie metod, které mohou být použity k detekci. Celkový výčet metod je poměrně široký a vzhledem k množství kombinací nejsou zdaleka popsány všechny metody. Dále jsem již zmínil důležitost detekce obličeje pro další fáze zpracování a hlavně rozpoznávání. Téma této práce vzniklo na základě obou těchto skutečností a úkolem je navrhnout a implementovat systém, který nevychází z některého jiného úspěšného systému jako například [3], ale umožní otestovat a zdokumentovat jeden z méně známých postupů. Tato poslední kapitola popisuje návrh systému krok po kroku, včetně zdůvodnění, proč jsem jednotlivé komponenty do návrhu zařadil.

Pro implementaci je zvolen v zadání jazyk C++, což je rozumná volba vzhledem k přenositelnosti systému, dostupnosti knihoven pro implementaci neuronové sítě, jakož i možnosti ladit systém na serverech fakulty.

Předzpracování vstupních dat

Vstupní snímek bude postupně projížděn oknem o velikosti 19x19 bodů. Velikost není volena náhodně, jedná se o rozměry snímků z trénovací databáze obličejů. Použití právě této databáze umožní srovnat výsledky dosažené tímto systémem se systémy jinými, jejichž výsledky jsou velmi často prezentovány na testovacích množinách této databáze. Po projetí celého snímku oknem dojde vždy ke zmenšení snímku, přičemž výchozí koeficient zmenšení je odhadnut na 1,2. Při vytváření nového zmenšeného snímku budou barvy jeho bodů interpolovány z původních pomocí bilineární interpolace.

V další fázi bude ze vstupních oken vypočítán histogram a okna budou ekvalizována. Ekvalizace oken namísto ekvalizace celého původního snímku by měla přinést zlepšení kontrastu pro jednotlivá okna. Pro potřeby neuronové sítě budou barvy bodů převedeny do normalizované formy.

Za účelem snížení dimenze obrazu a následné snížení výpočetní náročnosti neuronové sítě jsem se rozhodl použít výše zmíněnou analýzu hlavních komponent (PCA), která má za úkol snížit původní dimenzi 361. V případě PCA se nejedná o výběr příznaků, ale skutečně pouze o snížení dimenze a ignorování dat s nesoucích malé množství informace. Při trénování systému budou počítány vlastní tváře z celé trénovací množiny. Použitím PCA bych chtěl

dosáhnout snížení dimenze alespoň o jednu třetinu. Výsledný počet dimenzí bude nastaven na základě testování během implementace.

Při trénovaní systému je nutno si zapamatovat výsledky PCA. Při testování systému budou z testovaného snímku ponechány pouze ty dimenze, které byly určeny jako nejvíce relevantní ve fázi trénování. Ostatní dimenze budou ignorovány.

Návrh klasifikátoru

Neuronová síť

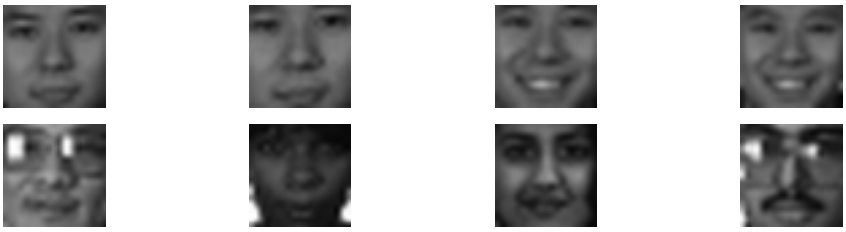
Jako klasifikátor jsem vybral neuronovou síť. Neuronové sítě mají obecně velmi slušnou výkonnost a pro účely detekce jsou vhodné. Z množství v oblasti softcomputingu používaných sítí jsem zvolil síť Back Propagation, jejímž algoritmem učení je zpětné šíření chyby. Jedná se o plně propojenou dopřednou síť. Plně propojená architektura je vhodná pro tyto účely, nejsou tedy potřeba žádné apriori manuální optimalizace propojení. Síť je schopná veškeré nastavení provést sama na základě dat z trénovací množiny.

Po skončení tréninkové fáze může být aplikováno tzv. prořezávání (angl. Prunning) neuronové sítě. Jedná se o odstranění spojení či celých neuronů, pokud toto odstranění nemá příliš velký vliv na úspěšnost systému (rozdíl chyb je pod toleranční hranicí). Typicky se jedná o odstranění spojení s vahou blížící se nule a neuronů, které jsou málo aktivní jak při přiložení obličejových tak i neobličejových dat na vstup systému.

K implementaci bude pravděpodobně použita některá dostupná knihovna pro neuronové sítě s vhodnou licencí. Tato volba bude upřesněna až při implementaci systému na základě konzultace s vedoucím diplomové práce.

Trénovací množina

Existuje větší množství dat přístupných pro trénování. Po dohodě s vedoucím jsme zvolili databázi CBCL dostupnou na webových stránkách MIT [5]. Databáze obsahuje 2.429 obličejů a 4.548 neobličejových dat pro trénování a 472 obličejů a 23.573 neobličejových dat pro testování systému. Jsou ve formátu PGM, což je jednoduchý formát pro uložení obrázků ve stupních šedi. Velikost obrázků pro trénování je zmíněných 19x19 bodů. Není tedy nutná žádná manuální příprava snímků.

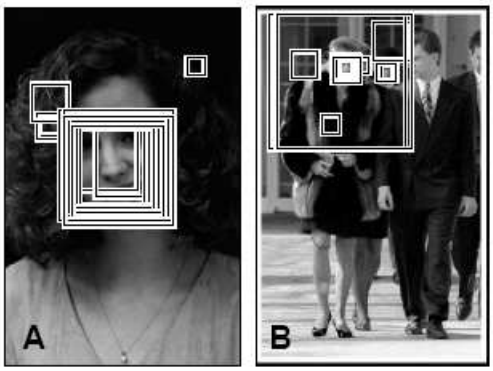


Obrázek 10: Ukázka obličejů z databáze CBCL (obrázky zvětšeny)

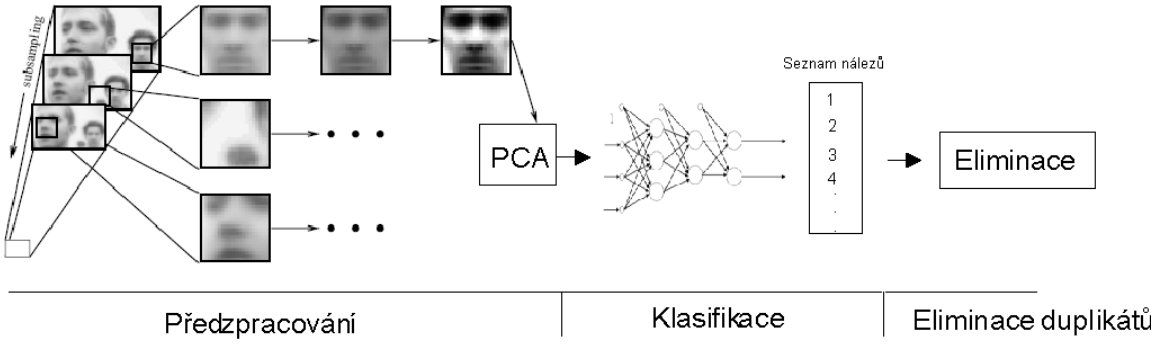
Zpracování výsledků

Eliminace duplikátů

Obrázek 11 ukazuje situaci, kdy použitá metoda lokalizuje jeden obličej na několika místech ve snímku. Nejedná se zpravidla o chybu metody, která takto nízko nastaveným prahem pro rozhodnutí, zda se ve zkoumané oblasti nachází obličej, zvyšuje počet správně nalezených obličejů. Naopak, ve skutečnosti lze na snímku s právě jedním obličejem nalézt větší množství výřezů tohoto snímku, na kterých je obličej stále dobře rozeznatelný. Dále pak navržená metoda pracuje s fixní velikostí zpracovávané oblasti obrazu, tzv. okna. Pro detekci obličejů i na snímcích, kde se mohou nacházet různé velikosti obličejů, je tedy nutné zpracovávat snímek několika průchody. Sloučení nálezů pak je provedeno na základě prahu, pokud se plocha nálezů překrývá o více než je stanovený práh, přičemž jako výsledek může být průnik nebo sjednocení oblastí nálezů. Pokud se oblasti překrývají méně nebo se jedná o nálezy na v obraze s velmi rozdílným měřítkem, jsou nálezy prezentovány jako dva různé detekované obličeje.



Obrázek 11: Několikanásobná detekce, převzato z [3]



Obrázek 12: Blokové schéma navrženého systému

Očekávaný výstup systému

Očekávaným výstupem neuronové sítě je rozhodnutí, zda se v daném okně nachází obličej či nikoli. Výstupem celého systému je však zanesení pozice nalezeného obličeje ve vstupním snímku. Z nálezu bude vytvořen seznam a atributy nálezů budou pozice a aktuální zmenšení vstupního snímku. Protože koeficient zmenšení je poměrně malý a okno bude posouváno po jednom bodě, očekávám množství detekcí, které se budou překrývat. Bude tedy použit postup popsaný výše ke sloučení takovýchto nálezů. Ke zhodnocení systému pak bude použita úspěšnost jak při slučování vícenásobných nálezů pomocí sjednocení tak při slučování pomocí průniku.

Výstup systému může vypadat podobně jako v [3]. Jeden z příkladů je znázorněn na obrázku 13.

Obrázek 13: Příklad výstupu systému pro detekci obličeje, převzato z [3]

Závěr

V minulé kapitole jsem představil systém pro detekci a lokalizaci obličeje v podobě, v jaké bude následně implementován. Použití neuronových sítí v oblasti detekce vykazuje obecně dobré výsledky a použití PCA by mělo procentuální úspěšnost ještě zvýšit. Avšak vzhledem k poměrně malé velikosti vstupního obrazu a složitosti problému nemusí být optimálním řešením a je zapotřebí tuto metodu porovnat s metodami jinými, jako je například filtrace Gaborovým filtrem ve fázi předzpracování. Velmi užitečným rozšířením by mohlo být roztřídění trénovací množiny do několika, obsahujících podobné obličeje. Vlastní tváře by pak mohly být počítány zvlášť pro obličeje s brýlemi, s vousy atd. Další možností zlepšení je nalezení optimální konfigurace neuronové sítě. Systém by měl tedy být implementován tak, aby mohlo dojít k výměně komponenty pro předzpracování dat a nastavení parametrů neuronové sítě. Výše popsané možnosti zvýšení výkonnosti budou součástí diplomové práce jen zčásti a to dle konzultací s vedoucím. Neimplementovaná rozšíření pak mohou být námětem pro pokračování v této práci v rámci jiných projektů

Použitá literatura a odkazy

[1] Šnorek, M., Jiřina M.: Neuronové sítě a neuropočítače, Skriptum ČVUT, vydavatelství ČVUT 1996.

[2] Rowley, H., Baluja, S., Kanade, S. : Neural Network-Based Face Detection, PAMI, 1998.

[3] Yang, M. H., Kriegman D., Ahuja, N.: Detecting Faces in Images – A Survey, IEEE Transaction on Pattern Analysis and Machine Intelligence, leden 2002

[4] Smith, L. I.: A tutorial on Principal Component Analysis, únor 2002, dokument dostupný na http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf (kontrolováno 20. 12. 2006)