

Technická univerzita v Liberci

Fakulta mechatroniky, informatiky a mezioborových studií

Studijní program: N2612 - Elektrotechnika a informatika

Studijní obor: 1802T007 - Informační technologie

Identifikace osob v obraze pomocí HMM

Human Face Identification from Image by HMM

Diplomová práce

Autor:

Bc. Zdeněk Šulc

Vedoucí diplomové práce:

Ing. Josef Chaloupka, Ph.D.

Konzultant:

Ing. Jindřich Žďánský, Ph.D.

V Liberci dne 28. 5. 2009

Místo pro vložení zadání

Prohlášení

Byl jsem seznámen s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 o právu autorském, zejména § 60 (školní dílo).

Beru na vědomí, že TUL má právo na uzavření licenční smlouvy o užití mé DP a prohlašuji, že **s o u h l a s í m** s případným užitím mé diplomové práce (prodej, zapůjčení apod.).

Jsem si vědom toho, že užít své diplomové práce či poskytnout licenci k jejímu využití mohu jen se souhlasem TUL, která má právo ode mne požadovat přiměřený příspěvek na úhradu nákladů, vynaložených univerzitou na vytvoření díla (až do jejich skutečné výše).

Diplomovou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím diplomové práce a konzultantem.

Datum

Podpis

Poděkování

Na tomto místě bych chtěl poděkovat zejména vedoucímu mé diplomové práce, Ing. Josefu Chaloupkovi, Ph.D. za mnoho podnětných rad a připomínek k mé práci.

Zároveň děkuji také mé rodině, která mě morálně i finančně podporovala při studiu i při tvorbě této práce, a také mým kolegům a kamarádům studentům za kritické zhodnocení této práce.

Abstrakt

Tato práce se zabývá možnostmi a reálným využitím skrytých Markovských modelů při identifikaci mluvčích v získaném obraze. První část pojednává o vlastnostech této statistické metody a její základní charakteristice. Druhá část se zabývá možnostmi získávání příznakových vektorů z obrazu vzhledem k invarianci vůči různým podmínkám snímání. Poslední část obsahuje rozbor vytvořeného programu a různé aspekty a postřehy získané při jeho tvorbě včetně popisu používaných obrázků a jejich případných úprav.

Klíčová slova: identifikace osob, skryté Markovské modely, příznakové vektory

Abstract

This work describes the possibilities of human face identification using hidden Markov models. The first part concludes characterization of this statistic method and its using. The second part considers possibilities of feature vectors extraction towards various illumination conditions. The last part describes the created program and its testing, including test results.

Keywords: face identification, hidden Markov models, feature vector

Obsah

Úvod.....	7
1 Skryté Markovské modely (HMM)	8
2 Rešerše z oblasti identifikace osob pomocí HMM	10
2.1 Rozhraní pro detekci a rozpoznávání obličeje pomocí HMM [10]	10
2.2 Rozpoznávání obličeje pomocí vnořených HMM [4]	10
2.3 Identifikační systém využívající jeden HMM a jediný vzorek obrazu na osobu [5].	13
2.4 Využití HMM a vlnkové transformace při identifikaci osob [6]	16
2.5 Optimalizace HMM a výběru trénovacích obrazů pro identifikaci osob [7].....	18
2.6 Kombinování obličejových klasifikátorů pro identifikaci osob [9].....	22
2.7 Shrnutí rešerše.....	23
2.7.1 HMM.....	23
2.7.2 Příznakové vektory	24
2.7.3 Trénování modelů	24
3 Příznakové vektory	25
3.1 Charakteristické příznaky v obrazech.....	25
3.2 Tvorba příznakových vektorů	25
3.2.1 Dvourozměrná diskretní kosinová transformace 2D-DCT	26
3.2.2 CTAG.....	26
3.2.3 Jasová intenzita	26
3.2.4 Detekce hran	26
3.2.5 Dvourozměrná Fourierova transformace (2D-FFT)	28
3.2.6 Vlnková transformace (Wavelet Transform, WT).....	28
4 Testování.....	30
4.1 ROI.....	30
4.2 Počet vzorků.....	31
4.3 Metody extrakce příznaků.....	31
4.3.1 Vlnková transformace	32
4.3.2 2D-FFT	32
4.3.3 2D-DCT	33
4.4 Jasová závislost.....	34
4.5 Shrnutí testů	34
5 Program Identifikátor	37
Závěr	39
Seznam použité literatury	40

Úvod

V posledních letech se detekce a identifikace lidských obličejů z pořízených fotografií, či častěji videa, stala vyhledávanou aplikací počítačového zpracování obrazu. Využívá se nejen na letištích a v hromadné dopravě, ale prakticky na všech místech, kde je k dispozici kamerový systém a je potřeba provádět identifikaci osob.

Skryté Markovské modely se využívají v mnoha různých úlohách, například při rozpoznávání řečových signálů, vyhledávání slov v rozmluvě, identifikaci mluvčího, rozpoznávání rukopisných znaků, analýze EEG a EKG, analýze DNA sekvencí nebo mobilní robotice. Tedy zejména ke zpracování jednorozměrného signálu.

Tato práce se zabývá možnostmi aplikace HMM na obrázky, tedy dvourozměrné signály. Cílem bylo vytvoření programu, který by dokázal zpracovat vstupní obraz takovým způsobem, aby mohl být výstupní vektor použitý jako jednorozměrný signál a zpracovaný pomocí další části programu, kde dojde k identifikaci osoby v původním obraze. Za použití sady programů se vytváří skryté Markovské modely a je umožněno také rozpoznávání vložených signálů.

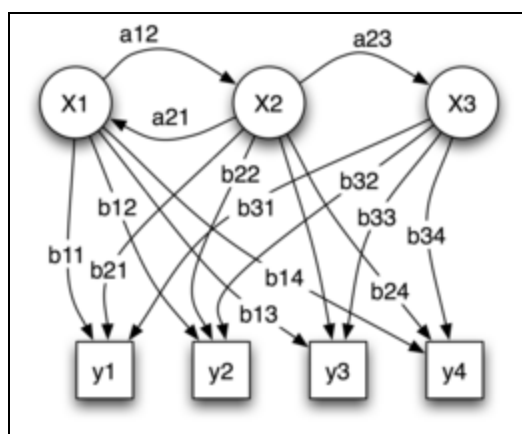
Cílem práce bylo především navrhnout metodu pro extrakci příznaků z obrazu, která by byla výpočetně jednoduchá a zároveň dostatečně robustní vůči rušivým vlivům, jako je například variabilní osvětlení osob v obraze. Příznaky by pak měly být řazeny do vektorů takovým způsobem, aby byla co nejvíce zachována dvourozměrnost obrazu. Tato část práce vyžaduje velké množství testů a pokusů.

Některé tyto metody by bylo možné převzít z vytvořené rešerše, případně použít kombinaci několika metod pro zlepšení výsledků identifikace.

1 Skryté Markovské modely (HMM)

Ve své podstatě jsou HMM zvláštním případem stochastických konečných automatů. Jak už název napovídá, stav, ve kterém se automat nachází, zůstává neznámý, dá se na něj však usuzovat pomocí výstupních proměnných. Mezi jednotlivými stavy se přechází pomocí přechodových funkcí, které mají různou pravděpodobnost. Tyto pravděpodobnosti jsou vypočítány na základě pozorovatelných parametrů zkoumaného jevu, tzv. příznakových vektorů. Výpočet pravděpodobností přechodu je nazýván „trénováním“ modelu, které je prováděno na skupině příznaků extrahovaných z fotografií dané osoby.

S přidáním každé nové osoby do databáze musí být vytvořen nový model. Každá osoba má totiž přiřazen samostatný model. Při rozpoznávání neznámé osoby prochází příznakový vektor všemi modely, ten s největší pravděpodobností shody určuje identitu osoby. Jedná se vlastně o ekvivalent lidského učení – čím vícekrát je daná osoba pozorována (počet snímků v databázi), tím lépe je osoba rozpoznána. Naopak čím více osob poznáváme, tím těžší je mezi nimi rozlišovat, zejména pokud jsou si navzájem podobné.



Obr. 01* : Příklad třístavového Markovského modelu se čtyřmi výstupními hodnotami

Formální popis skrytého Markovského modelu by byl takovýto [6]: Na skrytý Markovský model λ může být nahlíženo jako na Markovský model, jehož stavy nemůžou být explicitně pozorovány. Každý stav má přiřazenu pravděpodobnostní funkci,

* Převzato z [3]

modelující pravděpodobnost vyslání symbolu z daného stavu. HMM je definován následujícími entitami:

- $S = \{ S_1, S_2, \dots, S_N \}$... konečná množina možných skrytých stavů
- přechodová matice $\mathbf{A} = \{ a_{ij}, 1 \leq j \leq N \}$ reprezentující pravděpodobnost přechodu ze stavu S_i do stavu S_j ,

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], 1 \leq i, j \leq N, \quad (1.1)$$

kde $a_{ij} \geq 0$ a zároveň $\sum_{j=1}^N a_{ij} = 1$

- výstupní matice $\mathbf{B} = \{ b(o/S_j) \}$, vyjadřující pravděpodobnost vyslání symbolu o když je systém ve stavu S_j ; $b(o/S_j)$ je reprezentováno Gaussovským rozložením, tedy

$$b(o/S_j) = \mathfrak{N}(o/\mu_j, \Sigma_j) \quad (1.2)$$

- $\pi = \{ \pi_i \}$, počáteční pravděpodobnostní rozložení stavů, reprezentující pravděpodobnost, že se model na počátku nachází ve stavu i , tedy

$$\pi_i = P[q_1 = S_i], 1 \leq i \leq N, \quad (1.3)$$

kde $\pi_i \geq 0$ a zároveň $\sum_{i=1}^N \pi_i = 1$.

HMM lépe modelují jednorozměrné signály. Při zpracování obrazu jsou často využívány pseudo-dvourozměrné modely viz obr. 2, které umožňují zachování informací v horizontálním i vertikálním směru. Software HTK, který bude použit v této práci, však tvorbu takových modelů nepodporuje, bylo tedy třeba vytvořit vektory takovým způsobem, aby co nejlépe aproximovaly dvourozměrný signál a ztráty informací byly co nejmenší. Toho je dosaženo rozdělením obrazu na kvadranty, které jsou zpracovány samostatně a výsledky poté sloučeny zpět do jednoho vektoru. Více o této metodě je uvedeno v kapitole 4.

2 Rešerše z oblasti identifikace osob pomocí HMM

K identifikaci osob v obraze se kromě HMM používají i další metody. Veskrze se jedná o použití algoritmů, které je třeba nejprve natrénovat na daný problém, zejména se využívá Bayesovské statistiky a neuronových sítí, které mají schopnost se učit dokonce samostatně.

Jelikož identifikace osob je moderní a rozvíjející se obor, většina postupů je patentovaná nebo dokonce tajná. Je tedy velký problém vytvořit na toto téma aktuální rešerši, většina společností si své algoritmy dobře chrání a často není možné ani zjistit, jakou metodu k rozpoznávání používají. Tato rešerše proto čerpá zejména z výzkumných projektů a prací různých institucí, které se zabývají spíše teoretickou stránkou, než praktickým využitím v komerčních produktech.

2.1 Rozhraní pro detekci a rozpoznávání obličeje pomocí HMM [10]

Metoda využívá šedotónové obrazy bez pozadí a jednorozměrné HMM s 6 stavy, kdy každému stavu je přiřazena část obličeje – vlasy, čelo, oči, nos, ústa, brada. Každý obraz je rozdělen do horizontálních řezů, které se částečně překrývají (překryv je menší než výška řezu) a odpovídají jednotlivým částem obličeje.

Použitím dvourozměrné diskrétní kosinové transformace jsou v jednotlivých řezech nalezeny nejnižší frekvence. Ve čtvercovém okénku v jejich okolí se soustředí většina energie řezu. Tyto okénka jsou použity jako příznakové vektory. Před jejich složením do jednoho vektoru je však ještě provedena Karhunen-Loevova transformace. Velikost vektorů se tak dramaticky sníží a stejně tak i složitost systému.

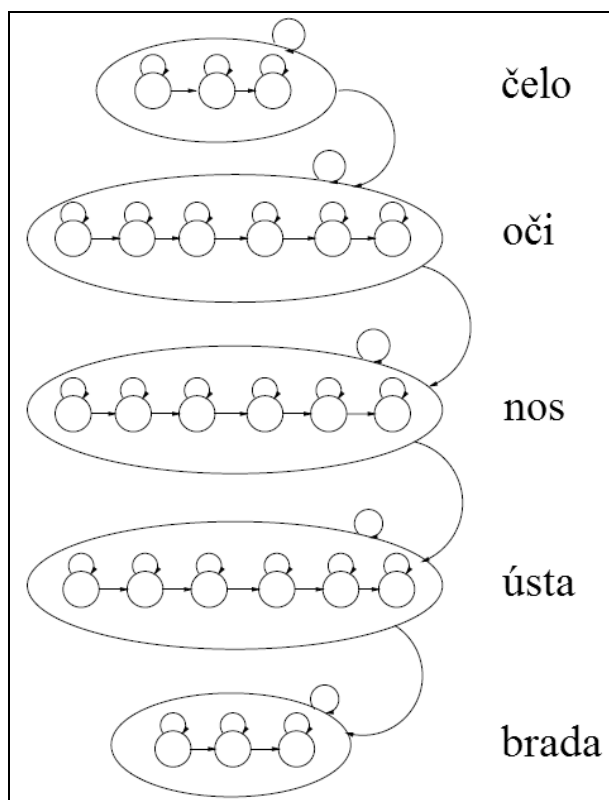
Přesnost rozpoznávání tohoto systému je 86 % za 250 ms na 10 fotografiích o rozměrech 92x112 pixelů, s různými účesy, výrazy, brýlemi i bez nich.

2.2 Rozpoznávání obličeje pomocí vnořených HMM [4]

Popisovaný HMM dokáže modelovat dvourozměrná data lépe než jednorozměrný HMM a výpočetně je méně náročný než dvourozměrný HMM. Takovýto model je vhodný pro obličejové snímky, jelikož využívá důležité rysy obličeje: ty mají tu samou strukturu jako superstavy* – odshora dolů, a také zleva doprava – jednotlivé stavy v superstavech.

* Superstav, z anglického „superstate“ – stav automatu, který obsahuje vnořené stavy.

Jednorozměrný HMM může být zobrazen tak, aby vzbuzoval dojem dvourozměrné struktury. Toho se docílí tak, že každý stav v celkovém modelu je sám modelem. Tímto způsobem se tedy HMM skládá z množiny *superstavů* a množiny *vnořených stavů*. Superstavy mohou být použity k modelování dvourozměrných dat podél jedné osy, vnořené HMM k modelování podle druhé osy.



Obr. 3* : Systém vnořených HMM pro identifikaci obličeje

Prvky daného HMM jsou následující:

- Počet superstavů N_0 a množina superstavů S_0 ,

$$S_0 = \{S_{0,i}\}, 1 \leq i \leq N_0. \quad (2.1)$$

- Počáteční pravděpodobnostní rozložení superstavů Π_0 ,

$$\Pi_0 = \{\pi_{0,i}\}, \quad (2.2)$$

kde $\pi_{0,i}$ je pravděpodobnost, že model se na počátku nachází ve stavu i .

- Přejchodová matice A_0 ,

$$A_0 = \{a_{0,ij}\}, \quad (2.3)$$

kde $a_{0,ij}$ je pravděpodobnost přechodu ze superstavu i do superstavu j .

- Parametry vnořených HMM Λ , které zahrnují:

* Přejzato z [4]

- Počet vnořených stavů $N_1^{(k)}$ v k -tém superstavu a množina vnořených stavů

$$S_1^{(k)} = \{S_1^{(k)}, i\}. \quad (2.4)$$

- Počáteční pravděpodobnostní rozložení stavů Π_1 ,

$$\Pi_1^{(k)} = \{\pi_{1,i}^{(k)}\}, \quad (2.5)$$

kde $\pi_{1,i}^{(k)}$ je pravděpodobnost, že model vnořený do superstavu k se na počátku nachází ve stavu i .

- Přejchodová matice A_1 ,

$$A_1^{(k)} = \{a_{1,jk}^{(k)}\}, \quad (2.6)$$

kde $a_{1,jk}^{(k)}$ je pravděpodobnost přechodu ze stavu k do stavu j .

- Výstupní matice B ,

$$B^{(k)} = \{b_i^{(k)}(O_{t_0,t_1})\} \quad (2.7)$$

pro sadu příznaků, kde O_{t_0,t_1} reprezentuje vektor pro řádek t_0 a sloupec t_1 .

Necht' $\Lambda^{(k)} = \{\Pi_1^{(k)}, A_1^{(k)}, B_1^{(k)}\}$ je množina parametrů definující k -tý superstav. S využitím zjednodušeného zápisu je vnořený HMM definován trojicí $\lambda = (\Pi_0, A_0, \Lambda)$, kde $\Lambda = \{\Lambda^{(1)}, \Lambda^{(2)}, \dots, \Lambda^{(N_0)}\}$.

Struktura stavů obličejového modelu a nenulové přechodové pravděpodobnosti vnořených HMM jsou znázorněny v obr. 3. Každý stav v celkovém HMM (odshora dolů) je přiřazen vnořenému HMM (zleva doprava).

Vstupní obraz je parametrizován rozdělením na čtvercové bloky, které se částečně překrývají. Pro každý blok je provedena 2D-DCT (dvourozměrná diskretní kosinová transformace), čtvercový výřez 2x3 kolem nejnižších frekvencí obsahuje většinu energie bloku a je použit jako příznakový vektor.

Pro každou osobu v databázi je vytvořen samostatný model. Trénování modelu probíhá následovně:

1. Data jsou nejprve rovnoměrně rozložena, čímž se vytvoří počáteční podmínky parametrů modelu. Příznaky celkového HMM jsou rovnoměrně vertikálně rozloženy v N_0 superstavech, poté jsou data odpovídající daným superstavům rovnoměrně rozložena zleva doprava do $N_1^{(k)}$ stavů.
2. V další iteraci je rovnoměrné rozložení nahrazeno dvojitě vnořeným algoritmem Viterbiho rozložení. To je nejprve aplikováno na všechny řádky obrazu, a jsou vypočítány pravděpodobnosti superstavů

$$P(O_{t_0,1} \dots O_{t_0,T_1}, q_{1,1}^{(t_0)} \dots q_{1,T_1}^{(t_0)} | \lambda^{(k)}), 1 \leq k \leq N_0, \quad (2.8)$$

kde $q_{1,t_1}^{(t_0)}$, $1 \leq t_1 \leq T_1$ reprezentuje stav uvnitř superstavu přiřazený příznaku O_{t_0,T_1} . Pravděpodobnosti superstavů, spolu s pravděpodobnostmi přechodu superstavů \mathbf{A}_0 a počátečními pravděpodobnostmi superstavů $\mathbf{\Pi}_0$ slouží k vykonání Viterbiho rozložení odshora dolů a k určení

$$P(O_{1,1} \dots O_{1,T_1}, \dots, O_{T_0,1} \dots O_{T_0,T_1}, q_{0,1} \dots q_{0,T_0} | \lambda), \quad (2.9)$$

či zjednodušeně $P(O, Q | \lambda)$. q_{0,t_0} , $1 \leq t_0 \leq T_0$ jsou superstavy odpovídající řádce t_0 .

3. Parametry modelu jsou odhadnuty rozšířením segmentového k-means algoritmu do druhé dimenze.
4. Iterace se zastaví, když rozdíl pravděpodobností Viterbiho rozložení sousedních iterací je menší, než nastavený práh.

Pravděpodobnost příznakového vektoru, daná vnořeným HMM obličejového modelu, je vypočítána pomocí dvojitě vnořeného Viterbiho rozpoznávače. Je vybrán model s největší pravděpodobností a ten určuje identitu rozpoznávané osoby. Testována byla databáze 40 osob, 10 snímků na osobu. Polovina snímků byla použita pro trénování, zbytek pro testování. Databáze obsahuje fotografie obličejů osob obou pohlaví, různého věku, s rozličnou mimikou a účesy. Přesnost tohoto systému je 98 % oproti 85% běžného HMM [4], ale výpočetní náročnost je řádově vyšší. Předčí však o tři až 8 procent i pseudo-dvourozměrný HMM, a to s menší výpočetní náročností.

2.3 Identifikační systém využívající jeden HMM a jediný vzorek obrazu na osobu [5]

Z hlediska identifikace osob se jedná o zcela novátorský přístup. Všechny osoby v databázi sdílí společně jediný HMM, který slouží jako prostředek k vyvážení dvojice příznaků. Každá osoba přispívá do modelu pouze jediným snímkem. K extrakci příznaků z obrazu je použita Haarova vlnková transformace. Tím jsou signifikantně zmenšeny rozměry příznakových vektorů. Identifikace rozpoznávané osoby je založena na nejvyšším výsledku, který je dán sumou pravděpodobností všech příznakových sekvencí extrahovaných z obrazu jak ve vertikální, tak i horizontální dimenzi.

Sekvence příznakových vektorů je generována následujícím způsobem:

- Obraz je nejprve rozdělen do překrývajících se vertikálních řezů, zleva doprava. Každý řez je označen pořadovým číslem. Maximální překryv sousedních řezů je o 1 pixel menší, než šířka řezu.
- Každý řez je poté rozdělen do překrývajících se bloků o délce strany rovné šířce řezu. Maximální překryv sousedních bloků je o 1 pixel menší, než šířka bloku. Nové příznakové bloky jsou vypočítány jako rozdíl mezi sousedními bloky.
- Poté jsou příznakové bloky normalizovány podle vztahu

$$b_i = \frac{B_i - \mu}{\sigma^2}, \quad (2.10)$$

kde μ je střední hodnota a σ^2 odchylka příznakového bloku B_i .

- Normalizované příznakové bloky b_i jsou poté řazeny po sloupcích, čímž vytvoří příznakové vektory.

K redukování rozměrů příznakových vektorů je na obraz nejprve aplikována Haarova vlnková transformace. Místo extrakce rozměrných bloků (8x8, 16x16 pixelů) s použitím malého překryvu lze na transformované obrazy použít menší bloky s větším překryvem. Dle potřeby lze na obrazy použít několikanásobnou vlnkovou dekompozici.

Rozpoznávání probíhá následujícím způsobem: Rozpoznávaný obraz projde nejprve Haarovou vlnkovou transformací a poté je rozdělen do překrývajících se vertikálních řezů. Tyto řezy jsou poté považovány za samostatné obrazy. Identifikace obrazu je založena na identifikaci jeho jednotlivých řezů. Každý řez je nezávisle zpracováván takto:

- Příznakové vektory jsou extrahovány stejným způsobem jako při trénování.
- Výpočet pravděpodobnosti podobnosti mezi neznámou osobou a osobou v databázi je založen na příznakových vektorech extrahovaných z daného řezu.

Nechť je pozice neznámého příznakového vektoru v řezu označena i_u a pořadové číslo řezu, ze kterého vektor pochází je j_u . Neznámý příznakový vektor je hledán v seznamu kódů oproti vektorům, indexovaným jako $\{i, j, r\}$:

$$\left\{ \begin{array}{l} i \in \{i_u, i_u \pm 1, i_u \pm 2, \dots, i_u \pm n_v\} \\ j \in \{j_u, j_u \pm 1, j_u \pm 2, \dots, j_u \pm n_h\} \\ r \in \{1, 2, \dots, R\} \end{array} \right\} \quad (2.11)$$

Parametry n_v a n_h limitují vyhledávání v seznamu kódů. Rozsah vyhledávání by mohl být rozšířen s ohledem na rotaci či posun obličeje v obraze, vedlo by to však k větší výpočetní

náročnosti. Jsou vypočítány Euklidovské vzdálenosti mezi vektory a ten s minimální hodnotou je označen jako nejlepší shoda.

Nechť je nejshodnější vektor indexován takto: $\{i_f, j_f, r_f\}$. Výpočet pravděpodobnosti shody neznámé osoby a osoby $\{r: 1 < r \leq R\}$ v databázi založený na příznakovém vektoru z daného řezu je prováděn dle těchto kroků:

1. Příznakový vektor je označen jako o_{if} , pokud $r = r_f$, jinak jako o_0 . To znamená, v případě že je bráno v potaz porovnání s osobou r , že příznakový vektor, který pravděpodobně patří k jiné osobě, nemá žádný význam.
2. Ty příznakové vektory, které mají stejný index j_f vytvoří příznakovou sekvenci O^{r,j_f} .
3. Sekvence O^{r,j_f} s délkou $T > 2$ jsou rozloženy na subsekvence délky 2, $\{O_1^{r,j_f}, O_2^{r,j_f}, \dots, O_{r-1}^{r,j_f}\}$, kde

$$O_t^{r,j_f} = \{o_t, o_{t+1}\}, 1 \leq t \leq T-1. \quad (2.12)$$
4. Každá subsekvence O^{r,j_f} je použita jako vstupní testovací příznaková sekvence HMM pro výpočet pravděpodobnosti. Čím vyšší je pravděpodobnost, tím spíše náleží daná subsekvence do řezu j_f dané osoby r .
5. Suma pravděpodobností přes všechny subsekvence extrahované z řezu j_u testovaného obrazu udává hodnotu shody pro osobu r .

Popsaná procedura může být považována za vertikální průběh HMM identifikace. Stejným způsobem probíhá i horizontální identifikace, ale na transponovaném obraze. Součet pravděpodobností všech subsekvencí v obou dimenzích udává výslednou hodnotu shody pro každou osobu. Výběr identifikované osoby je dán maximální hodnotou shody.

Testování bylo prováděno na databázi obsahující 120 osob. Ty byly nafoceny dvakrát, s dvoutýdenní prodlevou. V průběhu každého focení bylo zachyceno 13 různých podmínek, jako je variabilní výraz obličeje, osvětlení a okluze. Celkem bylo tedy použito 1320 obrázků. Před použitím fotografií bylo odstraněno pozadí, obrazy byly bilineární transformací upraveny tak, aby vzdálenost očí byla konstantní a poté byly ořezány na shodné rozměry.

Poté, co byly obrazy upraveny tak, aby obsahovaly pouze obličejovou část, byly tyto převedeny na šedotónové a předzpracovány histogramovou ekvalizací. Jako trénovací byly použity obrazy bez variabilních podmínek. Pro testování byl použit zbytek obrázků. Příznakové vektory jsou extrahovány z transformovaných obrázků za použití bloků

o velikosti 5x4 pixely. Krok pro vertikální i horizontální skenování je jeden pixel. Parametry n_h a n_v definující rozsah vyhledávání seznamu kódů jsou nastaveny na 3. Počet skrytých stavů v modelu je 5.

Výpočetní systém pracuje v prostředí Matlab, na platformě PC/WIN2000 používající PIII 1 GHz. Výpočetní čas pro natrénování HMM pro všechny osoby v databázi byl 0,2 s a čas potřebný pro rozpoznání jednoho obrazu 0,9 s. Tento čas zahrnoval i vlnkovou transformaci, převádějící obraz o velikosti 224x184 na matici parametrů 28x23. Rozpoznávací skóre systému se pohybovalo od 90 % do 100 % v závislosti na chybách fotografií. Těmito výsledky předčil systémy založené na PCA, LDA i LFA, které se nedokázaly přizpůsobit zhoršené kvalitě fotografií.

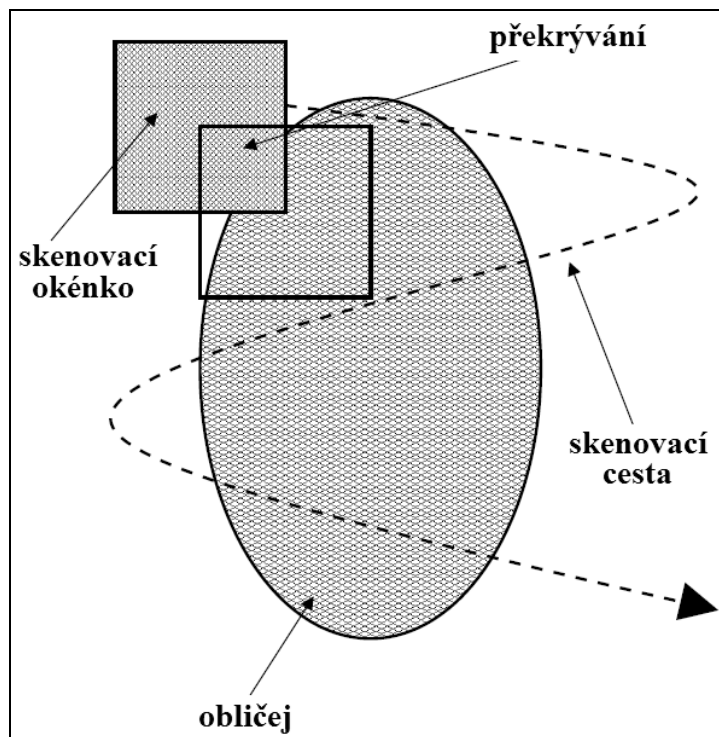
2.4 Využití HMM a vlnkové transformace při identifikaci osob [6]

V tomto přístupu jsou HMM používány standardním způsobem: pro každou třídu objektů (osobu) je vytvořen samostatný model použitím Baum-Welchova algoritmu a identifikace je prováděna pomocí maximální pravděpodobnosti, tj. neznámá fotografie je přiřazena té osobě, jejíž model vykazuje nejvyšší pravděpodobnost.

Postup vedoucí k získání příznakových vektorů z obrazu se sestává ze dvou kroků. V prvním kroku je z obrazu získána sekvence podobrazů* shodných rozměrů. Ty jsou získány posouváním čtvercového okénka obrazem, jak je znázorněno v obr. 4. Jednotlivé podobrazy se částečně překrývají. Druhým krokem je aplikování vlnkové transformace na jednotlivé podobrazy. Pouze několik klíčových koeficientů je uloženo. Po aplikaci na všechny podobrazy vznikne matice příznaků o velikosti $D \times T$, kde D je počet klíčových koeficientů a T je počet podobrazů.

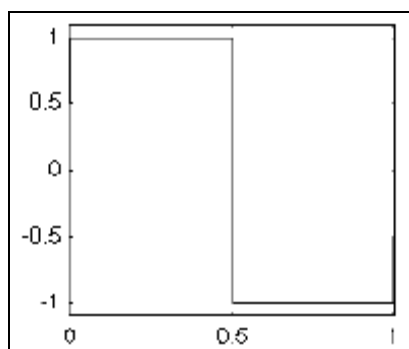
Vlnková transformace je metoda používaná v posledních letech v množství užitečných aplikací, zejména na poli obrazové komprese. Kódování založené na této transformaci přináší výrazné zvýšení kvality obrázků i při vyšším kompresním poměru. V horizontu několika posledních let bylo vyvinuto a implementováno mnoho postupů pro kompresi obrazu pomocí vlnkové transformace. Z důvodu velkého množství výhod jsou všechny kompresní algoritmy pro standart JPEG-2000 založeny právě na této metodě.

* Podobraz, z anglického „sub-image“ – výřez z celkového obrazu



Obr. 4* : Generování sekvence podobrazů

Vlnky mohou být definovány jako matematický nástroj pro hierarchické dekompoziční funkce. Vlnková transformace je zaměřena na popis funkce, jako je celkový tvar, ale také detaily, od velkých až po malé. Více formálně, vlnky jsou funkce definované nad konečným intervalem a mají nulovou střední hodnotu. Hlavní myšlenkou je reprezentovat jakoukoliv libovolnou funkci $f(t)$ pomocí superpozice sady elementárních (bazických) funkcí. Tyto bazické funkce, neboli *baby wavelets*, jsou získávány z jediného prototypu, zvaného *mateční vlnka* (*mother wavelet*) pomocí dilatace nebo kontrakce a posunu.



Obr. 5* : Haarova mateční vlnka

* Převzato z [2]

Pro extrakci příznaků je použita vlnková transformace používající Haarovy [2] mateční vlnky. K Haarovi se váží první zmínky o WT vůbec. Proto je Haarova vlnka tvarově nejjednodušší (obr. 5). Je nespojitá a její průběh připomíná skokovou funkci. Použitím nestandardní dekompozice, která propojuje řádkové a sloupcové zpracování, je docíleno efektivnějšího výpočtu koeficientů. Popisovaný algoritmus vypočítává koeficienty reprezentující obraz pomocí normalizované dvourozměrné Haarovy vlnky a řadí je v pořadí podle snižujícího se významu. Ponecháno je jen prvních M koeficientů, což nahrazuje ztrátovou kompresi obrazu. Počet ponechaných koeficientů určuje rozměry příznakové matice, přičemž jeho délka je dána počtem podobrazů.

Výsledky popsané metody byly porovnávány s jiným přístupem získávání koeficientů, a to pomocí diskrétní kosinové transformace (DCT). V testech byla použito databáze 40 osob s 10 obrazy pro každou. Pět obrazů bylo použito pro natrénování modelů, zbylých pět pro testování. Fotografie obsahovaly obličeje v různých posunutích a natočeních, s různou mimikou a navíc s variabilním osvětlením. Velikost podobrazů byla nastavena na 16x16 pixelů, počet ukládaných koeficientů byl variabilní (4, 8, 12) a stejně tak i úroveň překrytu podobrazů (50 % a 75 %).

Úroveň identifikace s použitím DCT a vlnek se nijak výrazně nelišila. Prokázalo se, že nejlepšího skóre je dosaženo použitím 12 koeficientů a překryvem 50 %. Při tomto nastavení je rozpoznávací skóre obou metod 100 %. To dokazuje, že HMM dokáže pracovat i s obrazy, které prošly ztrátovou kompresí JPEG-2000.

2.5 Optimalizace HMM a výběru trénovacích obrazů pro identifikaci osob [7]

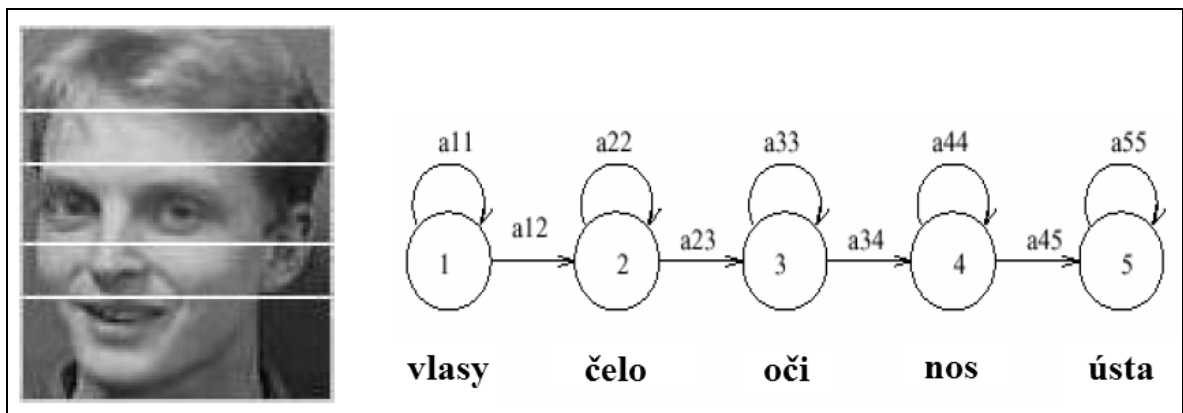
Techniky pro identifikaci obličejů mohou být rozděleny do tří kategorií:

- a) Metody využívající geometrické rysy obličeje
- b) Metody využívající vzory v obličejí
- c) Metody založené na modelech

Metody založené na modelech jsou méně citlivé na změny osvětlení, účesů, mimiky a posuny či rotace obrazu. Tato robustnost je dána statistickou povahou modelů.

Pro frontální obrazy obličejů jsou signifikantní obličejové oblasti (vlasy, čelo, oči, nos, ústa) řazeny v přirozeném pořadí odshora dolů, a to i v případě mírné rotace kolem středu nebo vertikální osy obrazu. Každé takové oblasti je přiřazen jeden stav HMM,

v návaznosti zleva doprava. Struktura stavů obličejového modelu, přechodové pravděpodobnosti a oblasti obličeje jsou uvedeny v obr. 6.



Obr. 6*: HMM pro identifikaci obličeje

Pro úlohu detekce obličeje jsou obrazy obličejů použity pro natrénování jediného modelu, avšak pro identifikaci obličeje musí být pro každou osobu v databázi natrénován samostatný HMM.

Obraz obličeje je vertikálně skenován odshora dolů, čím je vytvořena sada bloků. Z každého bloku jsou pomocí DCT získány koeficienty, které po sloučení vytvoří příznakový vektor. Po extrakci vektorů jsou tyto použity pro natrénování HMM.

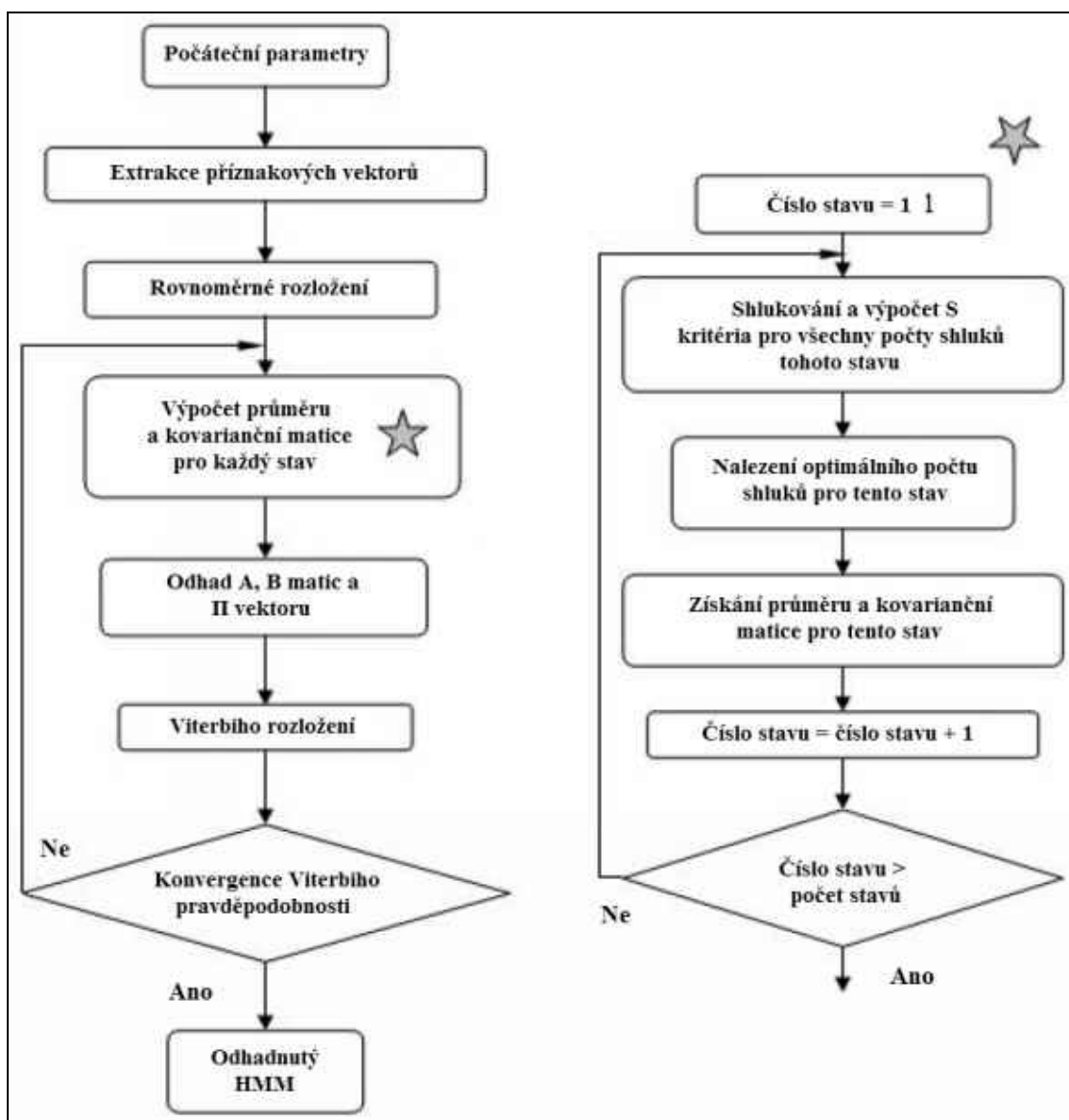
V průběhu trénování je obraz nejprve rovnoměrně rozdělen do 5 stavů odshora dolů a poté jsou inicializovány parametry HMM. V následující iteraci je rovnoměrné rozložení nahrazeno Viterbiho rozložením. Iterace je zastavena, když pravděpodobnost Viterbiho rozložení dosáhne požadovaného prahu. Konečných parametrů HMM je dosaženo pomocí Baum-Welchova rekurzivního algoritmu. Ve fázi identifikace je pro příznakový vektor vypočítána pravděpodobnost všech HMM. Model s největší pravděpodobností určuje identitu rozpoznávané osoby.

Průměrné rozpoznávací skóre pro jednorozměrné HMM se pohybuje okolo 85 % [10, 11]. Ve snaze vytvořit jednorozměrný model s lepšími výsledky byla provedena tato 3 vylepšení:

1. Nahrazení Baum-Welchova algoritmu K-means shlukovým algoritmem.
2. Přidání „S“ kritéria do shlukového algoritmu a nalezení nejvhodnějšího počtu shluků pro jednotlivé stavy.
3. Automatický výběr nejvhodnějších trénovacích obrázků.

* Převzato z [7]

Hlavní část trénovacího procesu sestává z opakování za sebou jdoucích Viterbiho a K-means algoritmů. V tomto případě specifikuje Viterbiho algoritmus optimální cestu v každém opakování a poté, v závislosti na této cestě, jsou výpočtem kombinací normálových funkcí K-means algoritmem odhadnuty parametry modelu (kovarianční matice, průměry, A a π matice). Trénovací fáze pokračuje, dokud není dosaženo prahové hodnoty Viterbiho pravděpodobnosti. Za předpokladu, že rozložení DCT koeficientů je normální, hustota pravděpodobnosti je vypočítána sečtením normálních kombinací pro každý příznakový vektor v každém stavu. V praxi, v porovnání s klasickým HMM, je složitost operací snížena nahrazením Baum-Welchova algoritmu K-means algoritmem.



Obr. 7*: Optimalizovaný algoritmus trénování HMM

* Převzato z [7]

„S“ kritérium může být použito k ohodnocení výkonu shlukového algoritmu. Je definováno jako poměr celkového rozptylu shluků ku minimální vzdálenosti mezi středy shluků. Ve druhém kroku optimalizace je přidáním S kritéria do shlukového algoritmu vypočten optimální počet shluků pro každý stav HMM. Počet shluků se pohybuje mezi dvěma až pěti. Nejlepší počet shluků odpovídá nejmenší hodnotě S.

Relativně nízké rozpoznávací skóre klasického jednorozměrného HMM je v postatě způsobeno nevhodným výběrem trénovacích obrázků a tím pádem nedostatkem důležitých informací. Tyto informace jsou nezbytné pro adekvátní natrénování modelu.

V popisovaném řešení jsou obrazy pro trénování vybrány automaticky ze sady dostupných obrazů dané osoby. V tomto případě (ORL databáze) se jedná o výběr 5 obrazů z 10 dostupných. Nejvhodnější trénovací sada by měla očividně obsahovat fotografie zachycené za různých podmínek. K vytvoření trénovací sady byly použity obrazy, jejichž DCT koeficienty se od sebe navzájem nejvíce odlišují. Tak lze z daného setu fotografií extrahovat maximální množství informací o dané osobě. Algoritmus pro výběr nejvhodnějších fotografií je následující:

1. Vyber náhodně jeden trénovací snímek.
2. Spočti rozdíl mezi DCT koeficienty daného snímku a všech zbylých snímků.
3. Jako druhý trénovací snímek vyber ten, který se nejvíce odlišuje od prvního.
4. Pro zbývající snímky vypočti souhrnný rozdíl mezi DCT daného snímku a DCT vybraných trénovacích snímků dle rovnice (2.13).
5. Jako další trénovací snímek urči ten s největším rozdílem.
6. Pokud ještě zbývá vybrat trénovací snímky, jdi zpět na 4.
7. Konec.

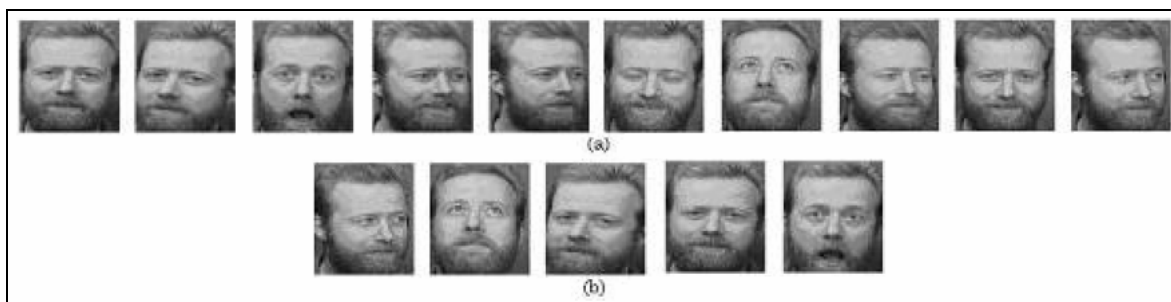
Po aplikaci tohoto algoritmu se dá očekávat, že poslední snímek bude obsahovat některé informace, které nebyly obsaženy v předchozích snímcích. Na obr. 8 je patrné, že vybrané snímky se od sebe maximálně odlišují.

Rovnice použité pro výpočet rozdílů DCT jsou následující:

$$D_{i,j}^2 = \sum_{n=1}^m (d_i(n) - d_j(n))^2 \quad (2.13)$$

$$\text{Index následujícího trénovacího snímku} = \arg \text{Max}_i (\text{Min}(D_{i,j})), \quad (2.14)$$

kde m je délka obrazového vektoru (počet sloupců x počet řádků).



Obr. 8^{*}: a) Snímky dostupné pro trénování, b) Snímky vybrané algoritmem.

K testování bylo použito následujících parametrů:

- Počet stavů HMM: 5
- Počet trénovacích snímků: 5
- Velikost snímků: 112x92 px
- Počet DCT koeficientů: 39

Aplikací popsaných úprav bylo rozpoznávací skóre zvýšeno z 85 % až na 96 %. Navýšení nebylo dosaženo zvýšením komplexnosti modelu, právě naopak, výpočty byly zjednodušeny díky nahrazení Baum-Welchova algoritmu K-means shlukovým algoritmem.

2.6 Kombinování obličejových klasifikátorů pro identifikaci osob [9]

Jako poslední je zde prezentován přístup kombinující 3 typy klasifikátorů do jediného systému s vylepšenou úrovní rozpoznávání. Jedná se o kombinaci profilového klasifikátoru, HMM klasifikátoru a Eigenface klasifikátoru. Profilový klasifikátor určuje totožnost osoby z binární černobílé fotografie profilu osoby. HMM klasifikátor je běžný klasifikátor, popsaný např. v [10]. Eigenface klasifikátor je další metodou z celé řady obličejových klasifikátorů a zabývat se jím podrobněji není náplní této rešerše.

Existuje několik možností, jak zkombinovat výsledky z klasifikátorů:

- **Hlasování** – Každý klasifikátor je považován za experta s jedním hlasem, to znamená, že volí svou nejpravděpodobnější osobu. Výsledné rozhodnutí je závislé na majoritě hlasování. Pokud skončí hlasování nerozhodně (každý klasifikátor identifikoval jinou osobu), je kombinační klasifikátor neschopný rozhodnout o přiřazení osoby k rozpoznávané fotografii a ta je zamítnuta.
- **Řazení** – Nejjednodušším přístupem je vypočítat sumu pravděpodobností ze všech tří klasifikátorů pro každou osobu z databáze. Osoba s nejvyšším výsledkem je určena jako identifikovaná.

^{*} Převzato z [7]

- **Vyhodnocování** – Podobné jako řazení, ale pravděpodobnosti jednotlivých klasifikátorů mají přiřazenu svoji váhu, v závislosti na vyhodnocení jejich úspěšnosti rozpoznávání zjištěné během testování jednotlivých klasifikátorů. Tento přístup přináší nejlepší výsledky, jelikož zohledňuje schopnosti jednotlivých klasifikátorů.

Co se týče úspěšnosti této metody, rozpoznávací skóre je 99,7 %, což je naprosto dostatečné pro většinu aplikací identifikace osob. Tento výsledek je dán povahou jednotlivých klasifikátorů, každý má své nevýhody i výhody, účelem jejich kombinace bylo posílit výhody a potlačit nevýhody. Jak bylo zjištěno, i prostá kombinace 3 výsledků bez zohlednění vhodnosti daného klasifikátoru zvýšila signifikantně úroveň rozpoznávání. Nevýhodou zůstává, že každá fotografie se musí de facto rozpoznávat třikrát, což přináší trojnásobné výpočetní časy, ale klasifikace se dá úspěšně provádět i při malém počtu trénovacích snímků.

2.7 Shrnutí rešerše

2.7.1 HMM

Skryté Markovské modely se používají v celé řadě úloh rozpoznávání. Jejich původním určením bylo rozpoznávat jednorozměrné signály, jako je např. zvuk. K tomuto účelu se hodí nejlépe, protože i ony mají jednorozměrnou strukturu. Přesto, jak bylo zjištěno, se dají použít i při rozpoznávání dvourozměrných signálů, tedy např. obrázků. V případě jejich použití jako jednorozměrných modelů vykazují okolo 85 procent úspěšnosti rozpoznávání. Proto zde byla snaha přizpůsobit je více dvourozměrným signálům, tedy vytvořit dvourozměrné modely.

Mezi dvourozměrné HMM se řadí např. vnořené HMM, které mají stavy řazené vertikálně, uvnitř každého z těchto stavů je vnořen další HMM s vertikálním řazením stavů. Tímto způsobem se dá částečně aproximovat dvourozměrný model. Dvourozměrné modely lépe zachovávají strukturu obrazu, jejich rozpoznávací skóre se pohybuje od 90 do 98 procent v závislosti na metodě extrakce příznaků. Tohoto zvýšení je dosaženo zvýšením složitosti modelu a tedy i výpočetní náročnosti.

Ve většině aplikací se pro každou osobu v databázi vytvoří samostatný HMM, po průchodu neznámého obrazu všemi modely se vybere ten, který má největší

pravděpodobnost. Ten určuje identitu rozpoznávané osoby. Existují však i metody, které používají pro všechny osoby v databázi jediný společný HMM.

2.7.2 Příznakové vektory

Standardní metodou pro vyhledávání příznaků v obraze je kosinová transformace, někdy i dvourozměrná kosinová transformace. Získané koeficienty nebývají použity všechny, jako příznaky jsou vybrány jen ty, které mají nejvyšší informační hodnotu (koeficienty kolem nejnižších frekvencí). Novější metodou je vlnková transformace.

Příznaky bývají zřídka extrahovány z celého obrazu najednou. Obraz je většinou rozdělen na několik menších částí, ze kterých jsou příznaky vybírány samostatně. Jejich sloučením pak vznikne příznakový vektor, který obsahuje více relevantních informací, než by bylo možné získat z celého obrazu. Čím větší je segmentace obrazu, tím více příznaků je získáno, od jistého prahu však tyto příznaky nenesou žádné nové informace. Tento práh je třeba nalézt, nejlépe testováním různých měřítek segmentů. Větší množství příznaků přináší lepší rozpoznávací skóre, ale i větší výpočetní zátěž. Příznakovými vektory se obsáhleji zabývá kapitola 3.

2.7.3 Trénování modelů

K trénování modelů se používá různé množství fotografií, čím méně jich je třeba, při zachování funkčnosti modelu, tím lépe. Pokud je k dispozici více fotografií než je nezbytně třeba, je vhodné z nich vybrat takové, které navzájem nesou co největší množství informací. Tím je myšleno, že fotografie by měly zachycovat co největší variaci výrazů, osvětlení, účesů a dalších podmínek, které se mohou během fotografování jedné osoby měnit. Čím více takových informací je dodáno při trénování, tím lépe poté model funguje i při výskytu nových, dosud nespátrých podmínek.

Nalézt vztah mezi počtem fotografií a mírou nové informace, která se dá z každé další přidané fotografie extrahovat, je prakticky nemožné. Vše závisí na podmínkách při fotografování, vhodný počet snímků se dá nejlépe určit testováním. Při jistém počtu již rozpoznávací skóre nestoupá, nebo jen velmi málo. Počet snímků pro trénování je většinou výsledkem kompromisu mezi výpočetní náročností a úspěšností identifikace.

3 Příznakové vektory

3.1 Charakteristické příznaky v obrazech

Vzhledem k velikosti a dvourozměrné reprezentaci rozpoznávaných obrazů je velmi nevhodné využívat při procesu rozpoznávání přímo tyto. Mnohem výhodnější je nejprve v nich vyhledat charakteristické znaky, a ty posléze uložit vhodným způsobem do vektoru, který lze dále zpracovat pomocí metody HMM a vytvořit z nich rozpoznávací schéma.

Charakteristické příznaky by měly splňovat tato kritéria:

- Reprezentativnost – co nejlépe odlišovat jednotlivé osoby
- Nekorelovanost – co nejmenší vazba mezi příznaky
- Praktičnost – dostupnost a použitelnost

V některých případech platí, že čím více příznaků je k dispozici, tím kvalitnější jsou výsledky rozpoznávání. Větší počet příznaků však s sebou přináší vyšší počet výpočetních operací a tím i delší výpočetní časy.

Neexistuje žádný obecný postup, který by určil vhodné znaky. Ty jsou voleny intuicí a následovány sérií testů, které na reprezentativním vzorku objektů (tedy osob) určí, zda splňují požadovaná kritéria. V rámci této práce bylo několik metod extrakce příznaků teoreticky nastudováno a po jejich zhodnocení byly některé otestovány vzhledem ke konečným výsledkům rozpoznávání.

3.2 Tvorba příznakových vektorů

Existuje celá řada sofistikovaných i zcela prostých metod (viz kapitola 2), používaných k extrakci příznakových vektorů ze vstupního obrazu. Jejich aplikace je závislá na povaze řešené úlohy, výpočetních možnostech použitého systému a povaze rozpoznávaných obrazů. Existují metody, které používají jen velmi málo výpočtů, získávání těchto příznaků je rychlé, ale jejich velikost je neúměrná vzhledem k dalšímu zpracovávání i v případě jejich archivace. Metody výpočetně náročné zabírají sice více času, ale příznaky jimi získané lépe odlišují jednotlivé osoby a jejich malá velikost zaručuje rychlé výpočty v následných částech zpracování, čímž se zcela kompenzuje časová ztráta na začátku procesu.

Několik těchto metod je popsáno dále, v teoretické analýze byl brán zřetel zejména na použití v kombinaci s HMM a obrazy obličejů.

3.2.1 Dvourozměrná diskretní kosinová transformace 2D-DCT

Dvourozměrná diskretní kosinová transformace [10,11] bývá nejčastěji používanou metodou extrakce příznaků. Na rozdíl od ostatních zmíněných metod se vytváří malé množství parametrů, se kterými se dá rychle a efektivně pracovat. Tento postup se provádí na šedotónových obrázcích. Prokázal se jako dostatečně efektivní při použití neměnného osvětlení, avšak při proměnlivém osvětlení (intenzitě i úhlu) dochází často k chybné detekci. Přesto byla tato metoda vzhledem k výsledkům jiných výzkumů použita při testech viz kapitola 4.

3.2.2 CTAG

Další možností získávání vektorů je metoda využívající CTAG (Color Texture Adjacent Graf) koeficientů. Analýzou textury v dané oblasti vzniká graf hodnot, které se dají přímo použít jako příznakové vektory. Tyto grafy se nesestavují pro celý obraz, ale pro jednotlivé části zvlášť, čímž se částečně zachovává dvojrozměrnost vektoru. Také nezávislost na podmínkách osvětlení je celkově vyhovující. Tato metoda byla z testování předem vyřazena z důvodu přílišné shody textur jednotlivých obličejů.

3.2.3 Jasová intenzita

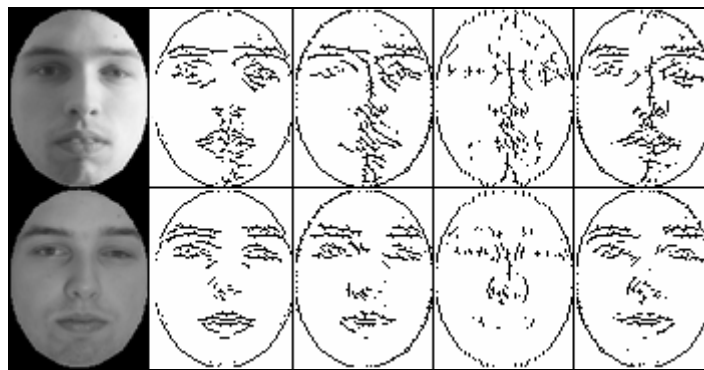
Nejjednodušší metodou je zřejmě využití jasové intenzity jednotlivých pixelů. Po převedení na šedotónový obraz vzniká matice hodnot, která se již pouze transformuje na vektor. Takto získané parametry mají tu nevýhodu, že se jejich počet zvyšuje úměrně s velikostí zkoumaného obrázku. Další nevýhodou, jak je patrné na první pohled, je naprostá nepřizpůsobivost vůči variabilním světelným podmínkám. Při každé změně osvětlení se musí znovu provádět trénování databáze a pokud se osvětlení mění často, je tento přístup naprosto nepoužitelný. Tato metoda nebyla do testování zařazena z důvodů výše uvedených.

3.2.4 Detekce hran

Tato metoda byla podrobně prověřena v předchozím projektu [12]. Pro vytváření příznakových vektorů jsou používány šedotónové obrázky. Pro zajištění bezvadné

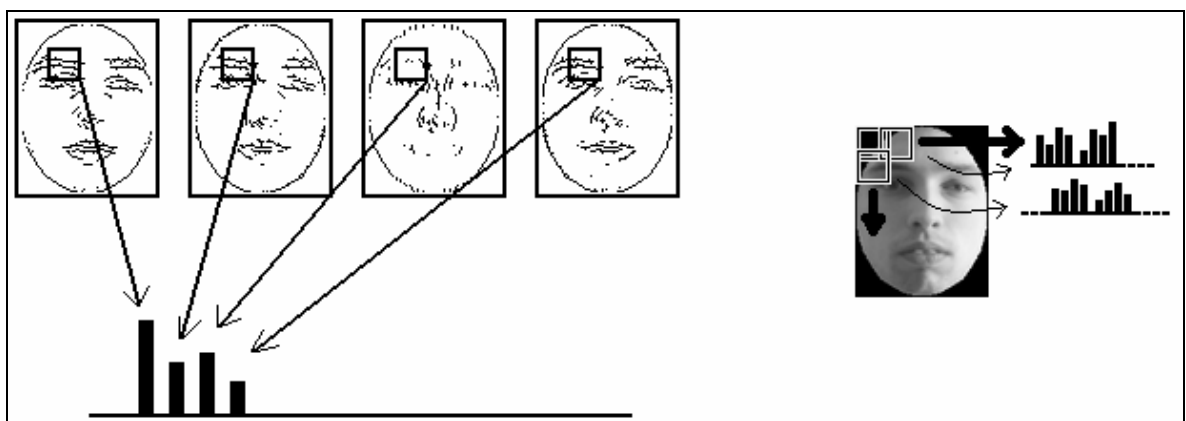
funkčnosti jsou obrázky ručně vymaskovány tak, že v obraze zůstane pouze výřez obličeje bez vlasů obsahující všechny potřebné části (čelo, oči, nos, ústa, brada) viz obr. 12, ROI 2. Obrázky jsou poté zmenšeny na rozměry 90x70 pixelů.

Dále následuje hledání hran, a to pomocí operátoru Prewittové. Jedná se o matici, kterou se vynásobí obrazová matice a vznikne tak černobílý binární obraz hledaných hran. Pro potlačení vlivu směru osvětlení se hrany vyhledávají postupně ve čtyřech směrech. Toho se docílí tak, že operátor se před každým vyhledáváním otočí o 45°. Vzniknou tak postupně 4 matice s hranami pro různé směry, tedy horizontální, +45°, vertikální a -45°, viz obr. 9.



Obr. 9: 4-směrová hranová detekce

Tyto 4 matice s hranami jsou pak najednou procházeny sčítacím okénkem 5x5 pixelů, které sečte všechny pixely označené jako hrany a pro každou matici uloží tento součet do příznakového vektoru viz obr. 10. Okénko se obrazem posunuje horizontálně s přesahem 1 pixelu přes předcházející (horizontálně i vertikálně) a s každým posunutím přidá do vektoru 4 součty. Tento průběh je také patrný na obr. 10. Při použití obrázků o rozměrech 90x70 pixelů obsahuje výsledný vektor 1496 parametrů.



Obr. 10: Tvorba příznakových vektorů

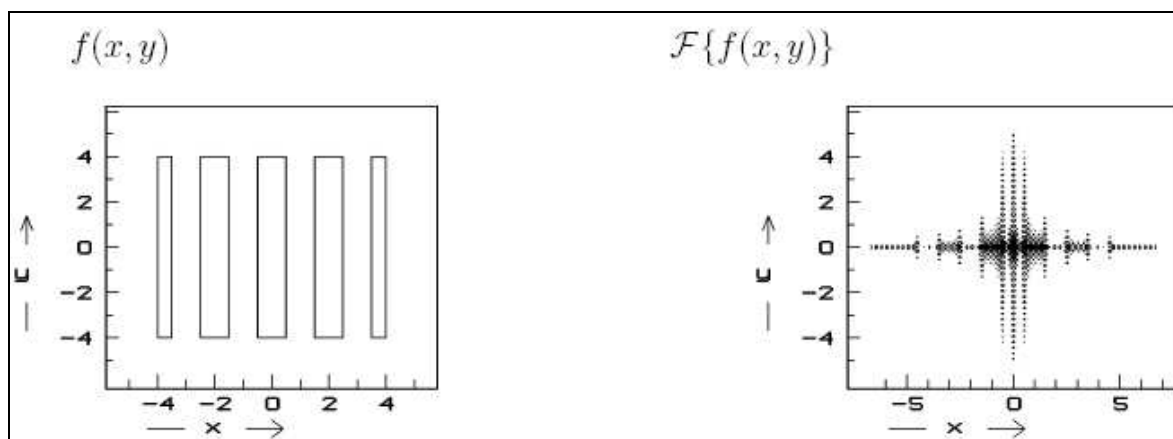
Velikost vytvořeného vektoru je závislá pouze na rozměrech parametrizovaných obrazů, nikoliv na zjištěných parametrech a v tomto případě odpovídá 5,85 kB. Díky volbě malé velikosti vzorků se dá i velké množství snímků (250) zpracovat v relativně krátkém čase (25 s) při použití běžného PC (Celeron 2,66 GHz, 448 MB RAM).

Rozpoznávací skóre dosažené za použití této metody se pohybovalo kolem 70 % a pro účely této práce bylo tedy nedostatečné. Také výpočetní čas byl vzhledem k použití mnohem větší databáze neúměrně vysoký a proto tato metoda nebyla v rámci této práce znovu otestována.

3.2.5 Dvourozměrná Fourierova transformace (2D-FFT)

Dvourozměrnou Fourierovou transformací (Fourierovým spektrem) komplexní funkce $g(x, y)$ nazveme funkci:

$$F\{g\} = \int_{-\infty-\infty}^{+\infty+\infty} G(f_x, f_y) \exp(-i2\pi(xf_x + yf_y)) dx dy \quad (3.1)$$



Obr. 11: Svislé čáry a jejich dvourozměrná Fourierova transformace

Jak bylo zjištěno během rešerše, Fourierovu transformaci lze s obstojnými výsledky aplikovat při úloze identifikace osob. Množství koeficientů je sice značné, dá se ale redukovat výběrem těch nejvýraznějších.

3.2.6 Vlnková transformace (Wavelet Transform, WT)

Aplikaci algoritmu WT [2] si lze představit následujícím způsobem. Mějme signál obsahující složky do 50Hz. Tento signál je nejdříve rozdělen na dva s užšími frekvenčními

pásmy (např. 0-25Hz a 25-50Hz). K tomu lze použít filtry typu dolní a horní propust. Potom je stejný postup aplikován buď na oba nebo pouze na jeden z nově vzniklých signálů. Obvykle je dále rozdělen signál dolnofrekvenční části (tj. v tomto případě 0-25Hz). Postup se opakuje, dokud signál nedosáhne předem určené krajní hladiny. Těmito operacemi jsou získány shluky signálů, které reprezentují ten samý signál v různých frekvenčních pásmech.

U většiny signálů je důležitější obsah nižších frekvencí, neboť právě tyto frekvence reprezentují podstatu nebo charakter signálu. Vysoké frekvence naopak ovlivňují odstín signálu. Projevují se na nich třeba i nechtěné jevy (např. měřený signál za přítomnosti šumu o vyšší frekvenci). *Diskrétní WT* (discrete wavelet transform; DWT) užívá pro tyto dva rozdílné aspekty signálu názvy *aproximační koeficienty* (approximation coefficients) a *detailní koeficienty* (detail coefficients).

Algoritmus DWT je následující:

Analyzovaný signál je přefiltrován dvěma komplementárními propustmi. Dolní propust oddělí aproximaci prvního řádu a horní propust detail prvního řádu. Je tak však získán dvojnásobný počet dat. To je dvakrát více dat, než je třeba pro reprezentaci signálu. Počet dat lze zredukovat převzorkováním signálu. To znamená, že je například vynechán každý druhý prvek signálu. Tento proces se nazývá *decimace* (oversampling). Je však třeba vzít v úvahu, že při tomto postupu může dojít k jevu zvanému *překrývání spekter* (aliasing). Zabránit mu lze použitím vhodných filtrů. Aplikací filtrace jsou získány *koeficienty aproximace 1. řádu* $kA1$ a *koeficienty detailu 1. řádu* $kD1$. Bude-li se pokračovat dále, lze se dostat ke koeficientům aproximací a detailů vyšších řádů. Při každém dalším řádu této operace se zdvojnásobuje měřítko (při prvním řádu máme měřítko 2_1 , při n -tém řádu 2_n).

Tato transformace byla použita při testování, a to na základě výsledků získaných z rešerše [6]. Bylo také zřejmé, že vlivem transformací bude počet koeficientů získaných touto metodou velmi nízký, což bylo pro účely této práce zásadní.

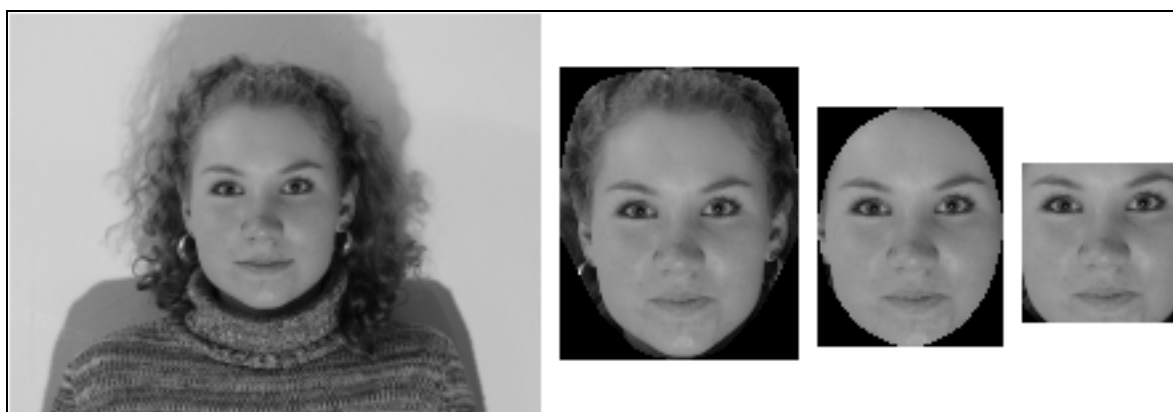
4 Testování

Rozpoznávací skóre může ovlivňovat mnoho různorodých faktorů. Některé z nich jsou rozvedeny dále v této kapitole. Může to být například jasová variace snímků, mimika obličeje, velikost snímku a celkově kvalita vstupní fotografie. V neposlední řadě pak také metoda extrakce příznaků z obrazu. V testování nebyly zahrnuty všechny myslitelné vlivy, a to zejména z důvodu omezenosti fotografické databáze. Bylo použito několik klíčových faktorů, které byly vzájemně kombinovány s cílem získat nejlepší možné výsledky identifikace.

4.1 ROI

Jedna z možných složek, ovlivňujících výsledné rozpoznávací skóre, je velikost a poloha ROI – oblasti zájmu (round of interest). Byly testovány 3 možné varianty, které by mohly být výstupem z face detektoru, viz obr. 12. Požadovaný face detektor bohužel nebyl k dispozici, snímky v databázi jsou tedy pouze „laboratorní simulací“ ideálního výstupu a byly vytvářeny ručně.

Celkem byly zpracovány snímky pro 100 osob, pro každou 35 vzorků (20 na trénování a 15 na testování), to vše třikrát pro 3 oblasti ROI, celkově tedy **10500** snímků zpracovaných ručně. Tento proces zabral značný objem času určeného pro ostatní části této práce.



Obr. 12: Původní obraz, ROI 1, ROI 2, ROI 3

První výřez o velikosti 90x110 pixelů zobrazuje celý obličej, včetně podstatné části vlasové pokrývky hlavy. Okolní prostor je černě vymaskován, aby různorodost pozadí

nenarušovala výsledky rozpoznávání. Druhý výřez, 70x90 pixelů, zahrnuje pouze čelo, oči, nos, ústa a bradu. Vlasy byly tentokrát odstraněny. Třetí výřez je čtvercový, 60 pixelů, a obsahuje pouze minimální oblast obličeje, pro člověka však stále ještě dostatečný pro rozpoznání dané osoby.

Teoreticky větší počet informací předurčuje lepší identifikaci, což bylo nakonec dokázáno při testování v kombinaci s dalšími faktory, ale překvapivě tomu tak nebylo ve všech případech. Rychlost zpracování byla samozřejmě úměrná počtu pixelů jednotlivých oblastí.

4.2 Počet vzorků

Počet vzorků použitých k trénování samozřejmě hraje také roli v konečném rozpoznávacím skóre. Ideální by bylo, aby se k trénování dal využít jediný obraz dané osoby. Testováním bylo zjištěno, jakým způsobem je ovlivněno rozpoznávání a kolik vzorků je třeba, aby se jeho úroveň pohybovala v požadovaných mezích.

Testování bylo prováděno za použití metody 2D-FFT, výsledky jsou uvedeny v tabulce 1. Dá se očekávat, že s využitím ještě většího množství snímků by se dosáhlo lepších výsledků, ovšem tím by se dále zvyšovala velikost databáze těchto snímků, zpracování by trvalo déle a celý proces by se neúměrně komplikoval. Bylo tedy stanoveno, že 20 snímků bude ideální jak z hlediska úrovně rozpoznávání, tak získávání a zpracování těchto fotografií.

Počet vzorků	10	15	20
Rozpoznávání [%]	80	84	94

Tab. 1: Závislost rozpoznávacího skóre na počtu vzorků

4.3 Metody extrakce příznaků

Nejdůležitějším faktorem, ovlivňujícím kvalitu rozpoznávání, je metoda použitá k extrakci příznaků z obrazu. V kapitole 3 je teoretický popis vlastností jednotlivých metod a vhodnosti jejich použití pro účely této práce. Na základě teoretického rozboru byly posléze vybrány 3 metody, které byly testovány v kombinaci s ostatními faktory. Pro všechna testování bylo použito 20 trénovacích a 15 testovacích snímků. K testování byl použit program **MATLAB**, který nabízí širokou paletu nástrojů pro práci s obrazem a aplikace jednotlivých transformací byla z hlediska programování nejjednodušší.

4.3.1 Vlnková transformace

Stavy	Mixtury	ROI 1 [%]	ROI 2 [%]	ROI 3 [%]
1	5	-	52	-
1	10	-	62	-
1	15	-	58	-
1	20	-	58	-
10	1	-	60	-
15	1	-	78	-
20	1	-	88	-
25	1	55	94	58
29	1	-	94	-

Tab. 2: Testování vlnkové transformace

Jak je vidět z tabulky 2, nejlépe se metoda osvědčila při použití ROI 2 a 25-ti stavového Markovského modelu. V závislosti na těchto výsledcích byla metoda otestována se stejným nastavením na ROI 1 a ROI 3. Zde se úspěšnost této metody nepotvrdila. Bohužel, při použití této metody na odlišných vzorcích ROI 2 rozpoznávací skóre prudce kleslo na úroveň 60 %. Původní výsledky byly tedy pouze statistickou odchylkou a celá metoda byla pro finální použití zamítnuta jako nevhodná.

4.3.2 2D-FFT

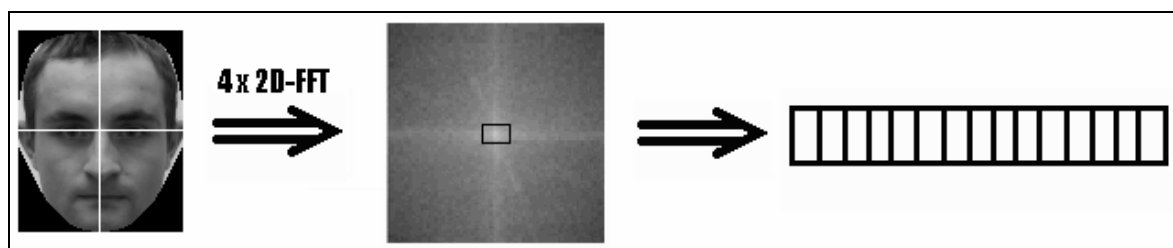
Počet stavů	Rozměry okénka	Rozpoznávání [%]
10	5x5	56
15	5x5	66
20	5x5	72
25	5x5	74
20	10x10	58
5	3x5	70
10	3x5	88
14	3x5	90

Tab. 3: Testování 2D-FFT

Jako nejlepší se ukázalo použití výřezu 3x5 FFT koeficientů kolem nejnižších frekvencí, které obsahuje většinu energie obrazu. Tím je získáno 15 příznaků, nejvyšší hodnota je určena jako referenční 1 a podle ní jsou ostatní normalizovány v rozsahu 0 - 1. Tato nejvyšší hodnota se poté do vektoru neukládá, protože je vždy 1 a nenesou žádnou užitečnou informaci. Tím je také dáno maximální možné množství stavů HMM – 14. Tento

počet stavů se také ukázal jako nejlepší možný pro rozpoznávání. S využitím tohoto nastavení bylo dosaženo rozpoznávacího skóre 90 %.

Dodatečně, pro další navýšení rozpoznávacího skóre, byl obraz rozdělen do čtyř kvadrantů, částečně se překrývajících. Z každého kvadrantu je poté vypočítána 2D-FFT a výřezy frekvencí jsou uloženy do společného vektoru. Tímto způsobem byla úroveň identifikace navýšena až na 99 %. Proto byl tento postup označen jako nejideálnější a získávání příznakových vektorů je ve vytvořeném programu založeno právě na něm. Postup je zobrazen na obr. 13.



Obr. 13: Tvorba příznakového vektoru pomocí 2D-FFT

4.3.3 2D-DCT

V závislosti na výsledcích předchozího testu 2D-FFT byla metoda 2D-DCT testována s použitím ROI 1 a se získáváním koeficientů ze 4 kvadrantů. Pro nejlépe fungující čtrnáctistavový HMM byl otestován i ROI 2 a 3, výsledky byly ale horší, takže kompletní testy nebyly již zpětně prováděny.

Počet stavů	Počet mixtur	Rozpoznávání [%]
1	1	50
1	2	64,7
1	3	68,7
1	5	64
2	1	20
5	1	74,7
10	1	87,3
14	1	94

Tab. 4: Testování 2D-DCT

Dle výsledků testů, viz tab. 4, by bylo možné použít k extrakci příznaků i 2D-DCT, nicméně ani při nejlepším nastavení nepřekonal rozpoznávací skóre 94% metodu 2D-FFT s jejími 99 procenty úspěšnosti. Jako metoda použitá ve vytvářeném programu byla tedy ponechána dvourozměrná Fourierova transformace.

4.4 Jasová závislost

Účelem tohoto testu bylo zjistit, jak ovlivní změna jasu identifikaci osob. Jasové změny byly prováděny uměle pomocí gamma korekce, za použití 10 snímků, které byly dříve bez problémů identifikovány. K extrakci příznakových vektorů byla použita 2D-FFT.

Gamma korekce	Rozpoznávání [%]
+ 0,1	60
+ 0,2	50
- 0,1	70
- 0,2	40

Tab. 5: Vliv jasové korekce na úroveň identifikace

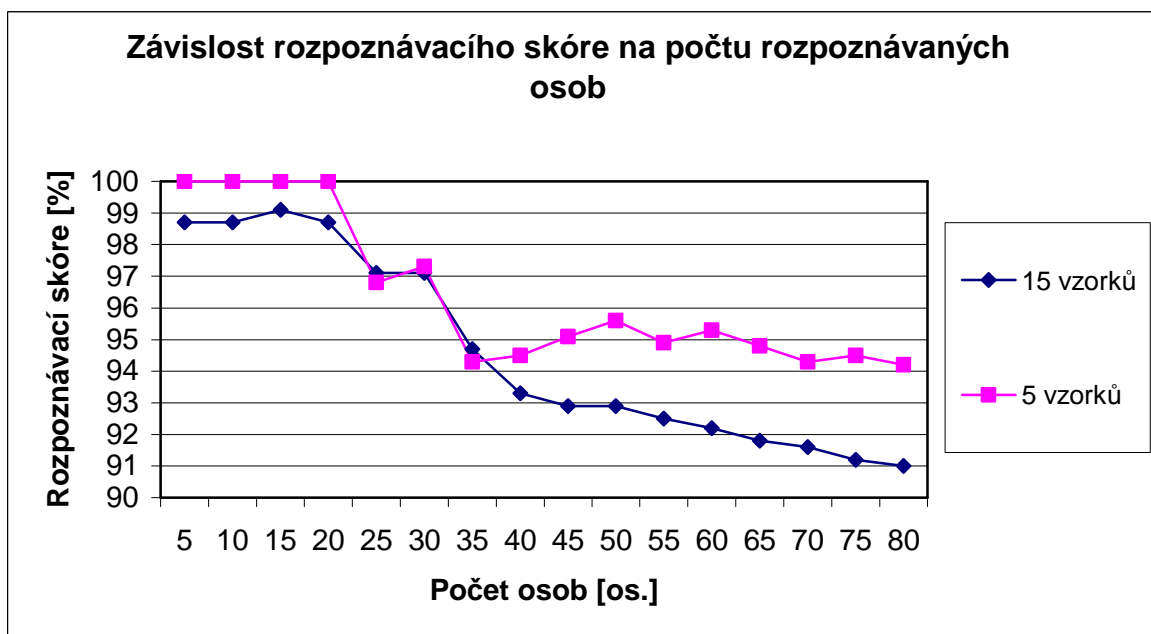
Jak je patrné z tab. 5, snímky, které byly dříve rozpoznávány se 100% úspěšností, nebyly HMM schopny správně přiřadit odpovídajícím osobám. To je dáno zejména tím, že při testování byly použity snímky s jednotnou úrovní jasu, natrénované modely byly tedy málo robustní vůči změnám podmínek. Tento problém by se dal částečně kompenzovat přidáním trénovacích obrazů s uměle sníženým a zvýšeným jasem.

4.5 Shrnutí testů

Z rozsahu celé práce byla většina času věnována právě testování. Bylo třeba vyzkoušet mnoho různých kombinací vlivů, působících na výslednou identifikaci osob.

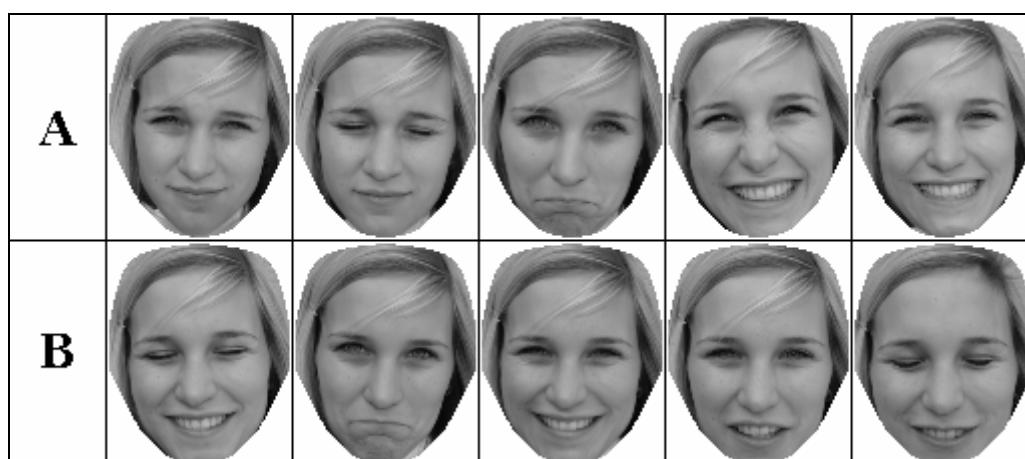
Nejlepším řešením tvorby příznakových vektorů je metoda dvourozměrné Fourierovy transformace, použitá na každý analyzovaný obraz právě 4x, a to po jednotlivých kvadrantech. Jako vstupní obraz se používá výřez obličeje o velikosti 110x90 pixelů, a to včetně vlasů. Pro vytvoření čtrnáctistavových HMM je použito dvacet výřezů na každého člověka.

Na závěr byl proveden test, který měl vyjádřit závislost kvality identifikace na počtu osob, mezi kterými je třeba provést rozpoznávání. Výsledky jsou uvedeny v grafu 1. Pro ilustraci bylo identifikováno 5 obrazů pro každou osobu a poté 15 obrazů na osobu. Jak je vidět, rozpoznávací skóre klesá přibližně o 10 % na každých 100 osob, i když nebylo možné otestovat větší množství osob, vzhledem k omezené velikosti databáze.



Tab. 1: Závislost rozpoznávacího skóre na počtu osob v databázi.

Závěrem ještě několik poznámek k používané databázi obličejů. Podmínky, za kterých bylo testování prováděno, se dají nazvat jako laboratorní. K dispozici byly ideálně pořízené fotografie, s vhodným osvětlením a bez natočení obličeje. Všechny fotografie byly pořízeny v krátkém časovém intervalu, na tom samém místě a za stejných podmínek. To platí jak pro trénovací, tak testovací snímky. Odchyly mezi nimi jsou minimální, viz obr.14.



Obr. 14: Vzorek databáze: A) Trénovací sada B) Testovací sada

V reálném světě se s něčím takovým nedá absolutně počítat. Trénovací fotografie budou pořízeny najednou, s vhodnými prostředky a osvětlením, například v ateliéru.

Fotografie určené pro identifikaci jsou pak získávány z jakýchkoliv zdrojů, ať už se jedná o průmyslové bezpečnostní kamery, fotobuňky nebo jiná záznamová zařízení. U takovýchto snímků bude kvalita velmi proměnlivá, rozhodně však horší než u natrénovaného setu. Úroveň identifikace bude tedy v reálných podmínkách pravděpodobně snížena, v závislosti na různorodosti prostředí. Míru této degradace nebylo možné zjistit z důvodu nedostatku reálných dat. Výsledek 90% úspěšnosti na 100 osob tedy musí být označen jako *za ideálních podmínek*.

5 Program Identifikátor

Výsledný program byl vytvořen v programovacím jazyce Pascal. Je velmi jednoduchý na obsluhu. Po kliknutí na nápis *Vložte foto* a vybrání fotografie je stiskem jediného tlačítka *Spustit rozpoznávání* identifikována osoba na obraze a je zobrazeno její jméno nebo alespoň číslo v databázi. Vytvořit novou databázi osob je poněkud složitější, ale průměrně zkušený uživatel nebude mít problém ani s touto částí programu. Stačí vybrat složku, ve které se nachází sada fotografií nové osoby, případně několika osob najednou. Program zároveň nabízí i funkce určené pro zkušené uživatele a testery. Využitím těchto funkcí může uživatel vstupovat do vnitřních procedur a ovlivňovat jednotlivé fáze identifikace, případně nastavováním proměnných dále vylepšovat program. Hlavní okno programu je zobrazeno na obr. 15.



Obr. 15: Okno programu

Program v sobě kombinuje několik funkčních celků. Na vstupu programu měl být zabudován face detektor, který by provedl výřez obličeje z vložené fotografie a tento by byl posléze zpracován v dalších částech programu. Tento detektor bohužel nebyl k dispozici,

program tedy využívá pouze referenční databázi výřezů obličejů, která byla vytvořena v rámci této práce. Nicméně program je napsán takovým způsobem, aby mohl být Face detektor jednoduše integrován bez zásadních zásahů do kódu.

První částí programu je tedy algoritmus, který provede extrakci příznakových vektorů z vybraného obrázku či setu obrázků. Používá k tomu metodu upravené dvourozměrné Fourierovy transformace, která byla vyvinuta v rámci této práce. Více o této metodě je uvedeno v kapitole 4.

Další částí je balíček programů HTK pro realizaci funkcí skrytých Markovských modelů, tedy jejich tvorbu a rozpoznávací algoritmy. Software HTK byl dodán jako samostatně fungující celek, který však pracoval pouze pod příkazovou řádkou. V rámci vytvářeného programu bylo třeba realizovat komunikaci s tímto softwarem. Identifikátor vytváří dávkové soubory *.bat pro HTK, které poté spouští v příkazové řádce. Tímto způsobem se provádí inicializace, trénování a rozpoznávání.

Program dále generuje seznamy souborů s parametry, gramatiku pro HMM, seznam objektů k identifikaci a další skripty potřebné k řádné funkci HTK, to vše v závislosti na vstupních obrazech. Všechny generované skripty se dají vytvářet jak jednotlivě, zejména pro potřeby doladování, testování a budoucího vylepšování programu, tak i hromadně, v rámci vytváření nového Markovského modelu.

Finální částí programu je algoritmus pro interpretaci výsledků identifikace. Výstup z HMM je sám o sobě příliš složitý, jsou z něj tedy vybrány jen relevantní informace a uživateli je zobrazeno pouze číslo, případně jméno identifikované osoby, pokud bylo uloženo.

Závěr

System, vytvořený v rámci této práce, umožňuje identifikaci osob s přesností 99%, klesající o 10% na každých 100 osob v databázi. Program dále umožňuje vytvářet nové databáze osob, přidávat další osoby do již vytvořených databází a z nich pak generovat skryté Markovské modely používané při identifikaci. Z vložených fotografií dokáže extrahovat příznakové vektory.

Nalezení vhodných příznaků pro identifikaci osob bylo důležitou částí celé práce. Pro vyřešení této úlohy byla provedena celá řada testů, ze kterých nakonec vyšla jako nejlepší metoda extrakce založená na dvourozměrné Fourierově transformaci. Ta je prováděna ve 4 kvadrantech obrazu, pro zvýšení úrovně rozpoznávání. Jako příznaky jsou pak vybrány pouze koeficienty nejnižších frekvencí, které obsahují většinu energie obrazu.

Pro trénování HMM bylo použito 20 fotografií od každé osoby. Markovské modely jsou složeny ze 14 stavů. Fotografie použité v testu jsou pořízeny na jednom místě, v jediném čase a jsou velmi kvalitní, lze je tedy považovat za ideální vzorky. Skutečná data budou velmi pravděpodobně kvalitativně na nižší úrovni, rozpoznávací skóre bude klesat úměrně s kvalitou fotografií.

Závěrem této práce lze říci, že skryté Markovské modely se dají využít v úloze identifikace osob v obraze, ale spíše za použití malých a středních databází osob, maximálně do několika stovek lidí. Kladem je, že identifikace může probíhat v reálném čase. Nevýhodou je, že modely je třeba natrénovat z relativně velkého počtu snímků (20). Vzhledem k těmto podmínkám se dá usuzovat, že vytvořený program by se lépe uplatnil pro sledování pohybu osob v malé až střední komunitě lidí, například ve firmě, než pro identifikaci osob ve městech nebo dokonce celosvětově.

Do budoucna by se dala dále zlepšovat metoda extrakce příznaků z obrazu, která je klíčová pro úroveň identifikace.

Seznam použité literatury

- [1] Gender Classification with Support Vector Machines, Baback Moghaddam, Ming-Hsuan Yang, MERL - Mitsubishi Electric Research Laboratory, 2000
- [2] *Wavelet Toolbox, User's Guide. Revised for version 2(Release 12)*, Misiti M., Misiti Y., Oppenheim G., Poggi J., The Mathworks Inc., 2000
- [3] http://en.wikipedia.org/wiki/Hidden_Markov_model
- [4] An Embedded HMM-based Approach for Face Detection and Recognition, Ara V. Nefian, Monson H. Hayes III., Georgia Institute of Technology, Atlanta, 1999
- [5] Face Identification System using Single Hidden Markov Model and Single Sample Image per Person, Hung-Son Le and Haibo Li, Digital Media Lab, Dept. of Applied Physics and Electronics, Umeå University, 2004
- [6] Using Hidden Markov Models and Wavelets for face recognition, M. Bicego, U. Castellani, V. Murino, Dipartimento di Informatica, Università di Verona, 2003
- [7] 1D-HMM for Face Verification: Model Optimization Using Improved Algorithm and Intelligent Selection of Training Images, S. Naderi, M. S. Min, N. Moghaddam Charkari, 2004
- [8] Parameterisation of a Stochastic Model for Human Face Identification, F. S. Samaria, A. C. Hartert, 1994
- [9] Combination of Face Classifiers for Person Identification, Bernard Achermann, Horst Bunke, Institut für Informatik und angewandte Mathematik, Universität Bern, 1996
- [10] Face detection and recognition using Hidden Markov Models, Ara V. Nefian, Monson H. Hayes III., Georgia Institute of Technology, Atlanta, 1998
- [11] Face Recognition Using A DCT-HMM Approach, Vinayadatt V. Kohir, U. B. Desai, Indian Institute of Technology, Mumbai 1998
- [12] Identifikování osob v obraze, Zdeněk Šulc, TUL Liberec, 2008