

Rozpoznávání tváří II

Vojtěch Franc

Centrum strojového vnímání, ČVUT FEL Praha

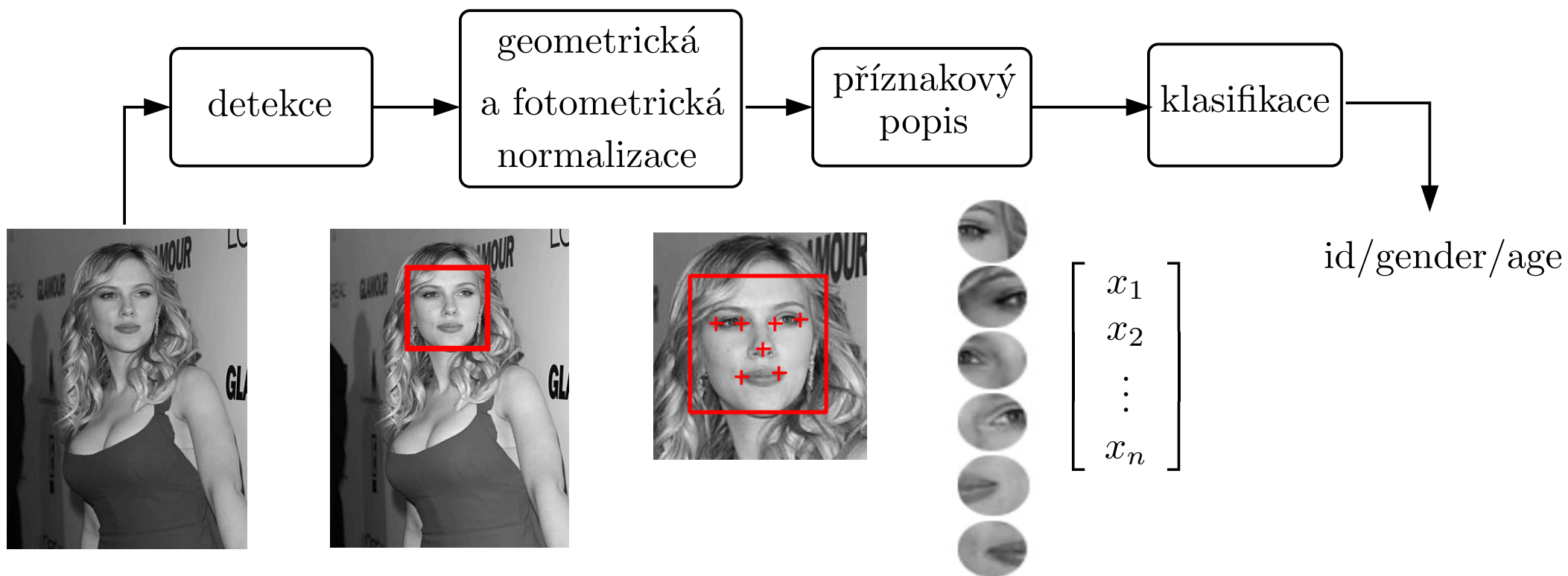


Biometrie ZS 2014

Poděkování Michalu Uříčarovi za obrázky popisující detekci významných bodů

Příklad rozpoznávacího systému, stavební bloky

- ◆ Detekce: lokalizace tváře v obraze (viz. minulá přednáška).
- ◆ Geometrická a fotometrická normalizace: cílem je získat reprezentaci tváře invariantní vůči geometrické transformaci či změně osvětlení
- ◆ Příznakový popis: tvář se popíše vektorem čísel
- ◆ Klasifikace: zařazení vektoru příznaků do třídy



Detekce významných bodů na tváři

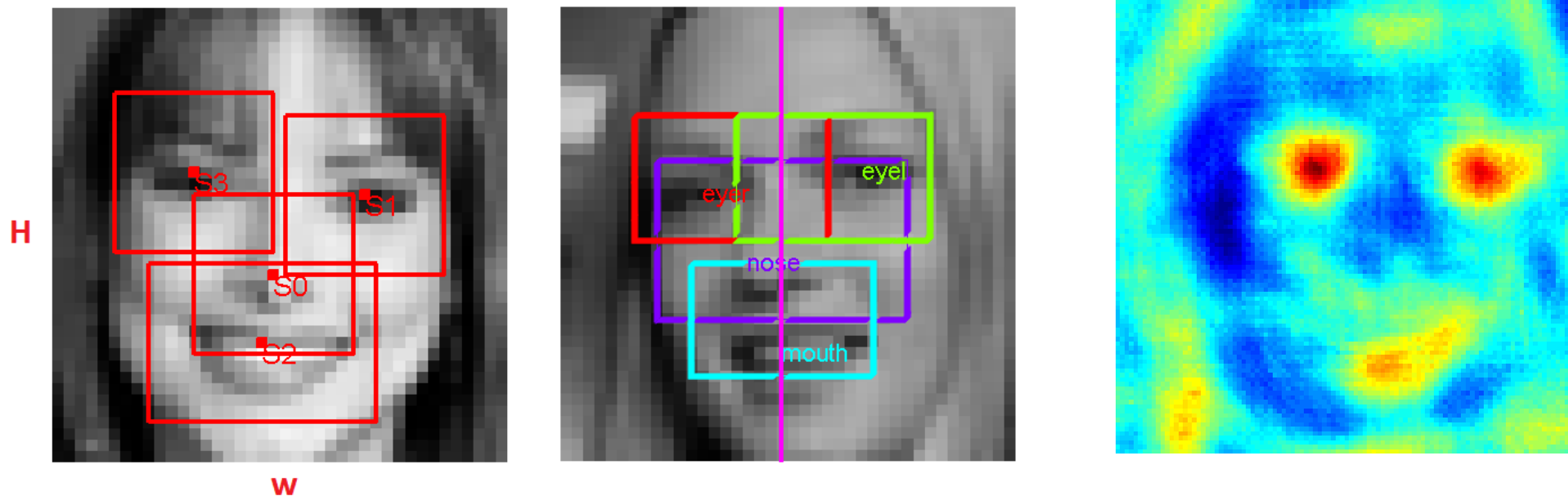
- ◆ Cíl: Lokalizovat předem definované významné body na lidské tváři
- ◆ Vstup: obrázek (nebo video)
- ◆ Výstup: pozice významných bodů



Nezávislý detektor pro každý významný bod

- ◆ $V = \{0, \dots, M - 1\}$ množina významných bodů
- ◆ $S_i \subset \{1, \dots, H\} \times \{1, \dots, W\}$ množina přípustných pozic i -tého významného bodu
- ◆ $q_i(I, s_i)$ “pravděpodobnost”, že i -tý významný bod je v obrázku I na souřadnici $s_i \in S_i$

odezva $q_i(I, s)$ pro levé oko



- ◆ Odhady pozic $\{\hat{s}_0, \dots, \hat{s}_{M-1}\}$ významných bodů v obrázku I lze nalézt nezávislými detektory (tj. jako “scanning window” detektor tváří z minulé přednášky)

$$\hat{s}_i = \operatorname{argmax}_{s \in S_i} q_i(I, s) \quad i \in V$$

- ◆ **Nevýhoda:** nevyužívá se toho, že pozice významných bodů nejsou nezávislé.

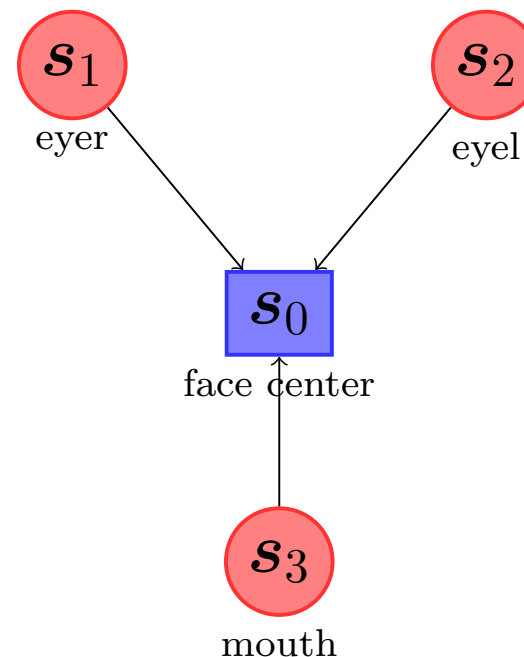
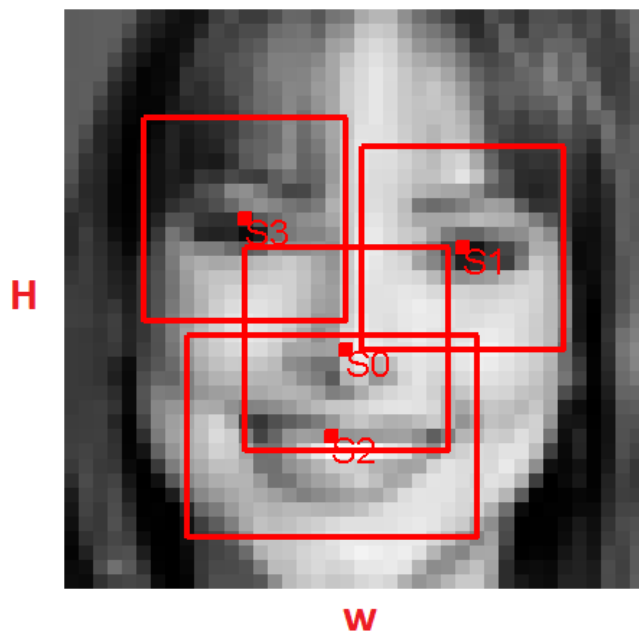
Deformable part model

- ◆ $G = (V, E)$ graf definující významné body a sousedství
- ◆ $g_{ij}(\mathbf{s}_i, \mathbf{s}_j)$ skóre konfigurace $(\mathbf{s}_i, \mathbf{s}_j)$ sousedících bodů $(i, j) \in E$
- ◆ $f(I, \mathbf{s})$ skóre pro konfiguraci pozic významných bodů $\mathbf{s} = (\mathbf{s}_0, \dots, \mathbf{s}_{M-1})$ v obrázku I

$$f(I, \mathbf{s}) = \sum_{i \in V} q_i(I, \mathbf{s}_i) + \sum_{(i, j) \in E} g_{ij}(\mathbf{s}_i, \mathbf{s}_j)$$

shoda
s obrázkem

deformační skóre
popisuje tvar tváře



Deformable part model

- ◆ Odhad pozice významných bodů v obrázku I vede na hledání konfigurace

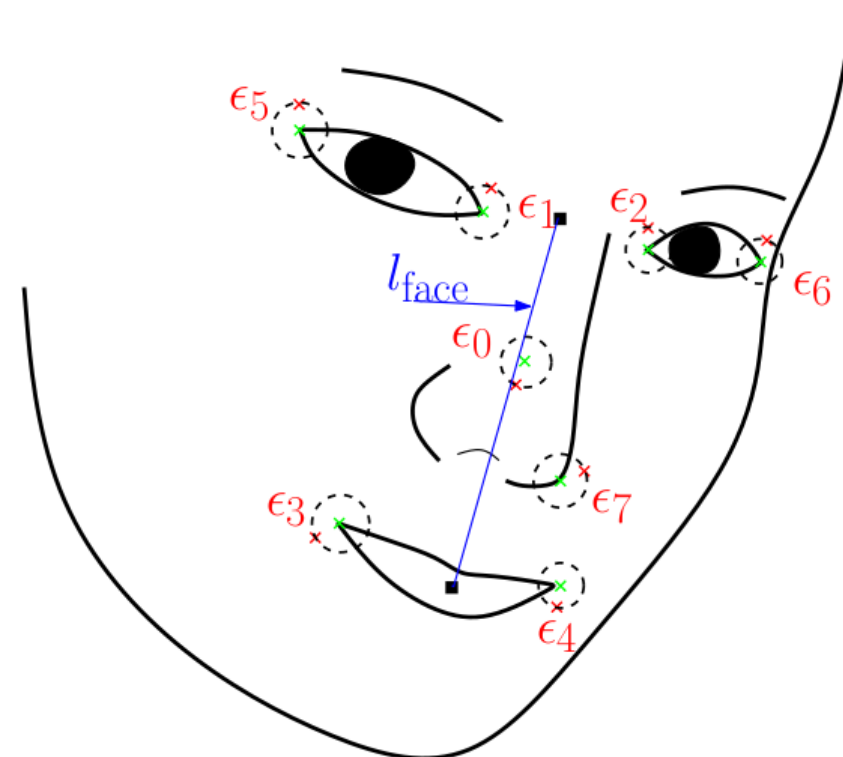
$$(\hat{\mathbf{s}}_0, \dots, \hat{\mathbf{s}}_{M-1}) = \underset{(\mathbf{s}_0, \dots, \mathbf{s}_{M-1}) \in S_0 \times \dots \times S_{M-1}}{\operatorname{argmax}} \left[\sum_{i \in V} q_i(I, \mathbf{s}_i) + \sum_{(i,j) \in E} g_{ij}(\mathbf{s}_i, \mathbf{s}_j) \right]$$

- ◆ Pro acyklický graf (V, E) lze nalézt maximální konfiguraci pomocí dynamického programování.

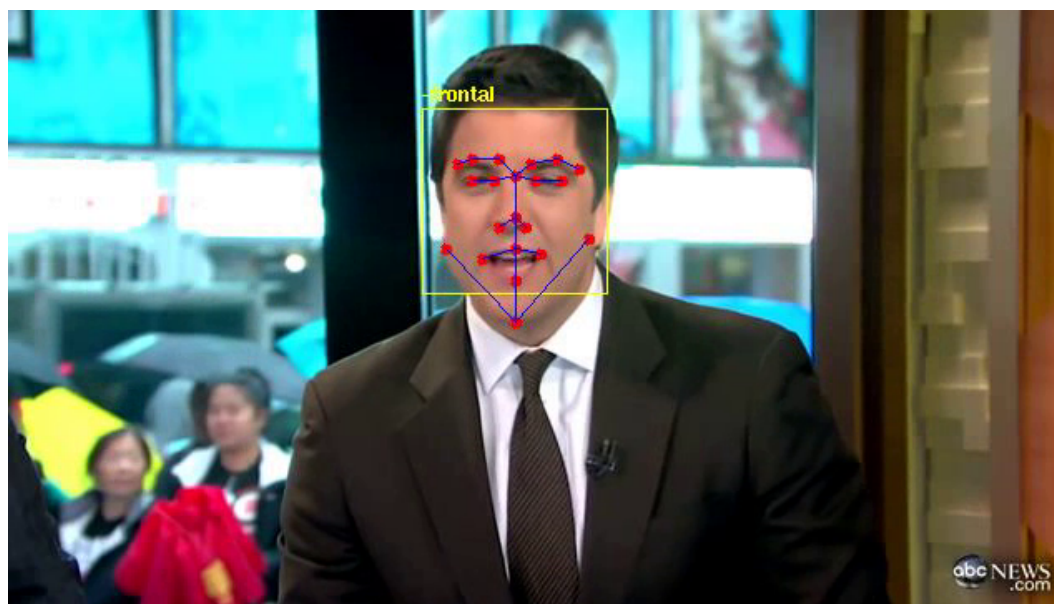
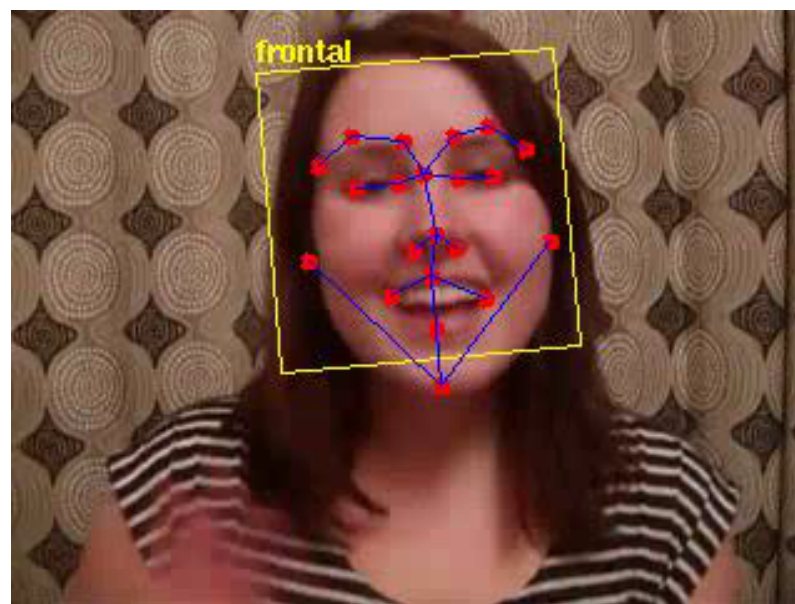
- ◆ Funkce q_i a g_{ij} lze učit z množiny $\{(I^1, \mathbf{s}^1), \dots, (I^m, \mathbf{s}^m)\}$, která obsahuje příklady obrázků I^j a jejich manuální anotaci \mathbf{s}^j .

- ◆ Cílem je naučit takové funkce q_i a g_{ij} , aby průměrná odchylka odhadu pozic byla minimální.

$$\text{odchylka} = \frac{\varepsilon_0 + \dots + \varepsilon_{M-1}}{M} \cdot \frac{1}{l_{face}}$$

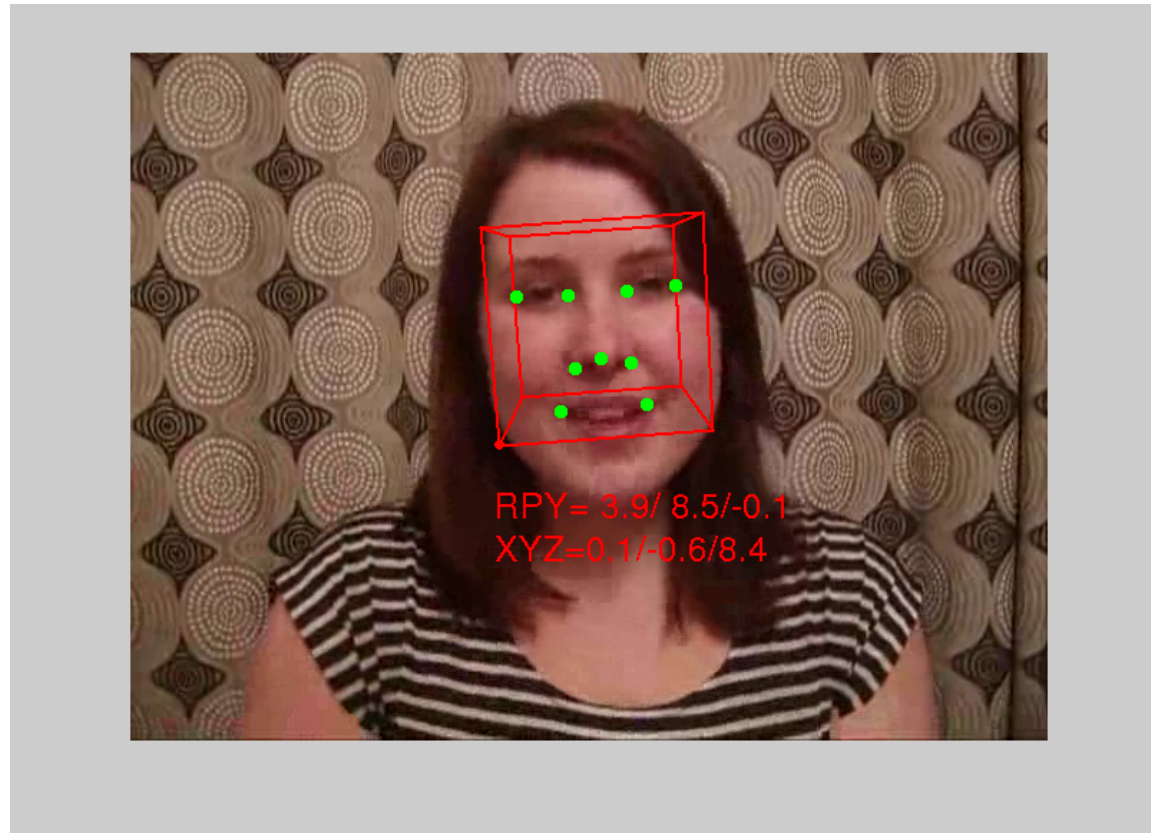


Deformable part model v akci



Videa převzata z <http://cmp.felk.cvut.cz/~uricamic/flandmark/>.

Detektor významných bodů ve 3D



Fitování rigidního 3D module tváře to 2D obázku vede na optimalizační úlohu

$$(\hat{\mathbf{t}}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \underset{(\mathbf{t}, \alpha, \beta, \gamma) \in \mathbb{R}^6}{\text{Argmax}} \text{score}(I, \mathbf{t}, \alpha, \beta, \gamma) := \sum_{i=1}^L f_i \left(I, K[R(\alpha, \beta, \gamma)|\mathbf{t}]\mathbf{X}_i \right),$$

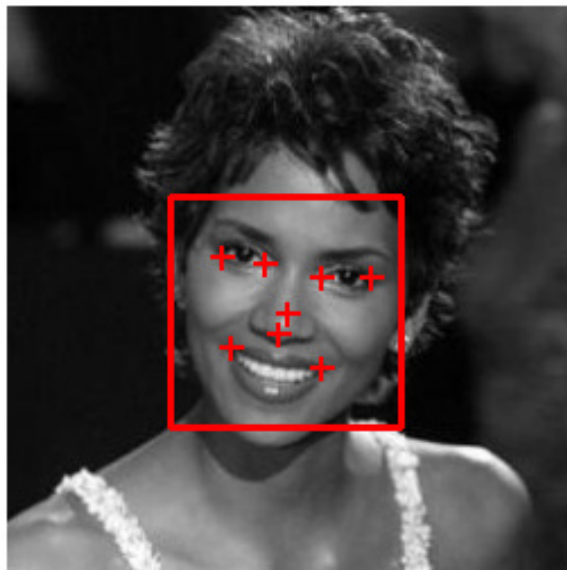
Videa převzata z *Cech et al. A 3D Approach to Facial Landmarks: Detection, Refinement, and Tracking. ICPR. 2014.*

Geometrická normalizace tváře - affinní vyrovnání

- ◆ Cíl: Převést vstupní tvář do polohy, která bude nezávislá na třídě geometrických transformací, vůči které chceme být invariantní (např. rotace, translace, změna měřítka).
- ◆ Metoda zajišťující invarianci vůči affinní transformaci: nalezneme parametry affinní transformace (např. metodou nejmenších čtverců)

$$s' = \mathbf{A}s + \mathbf{b} \quad \mathbf{A} \in \mathbb{R}^{2 \times 2}, \mathbf{b} \in \mathbb{R}^{2 \times 1}, \mathbf{s} \in \mathbb{R}^{2 \times 1}$$

která promítne nalezené významné body (s_0, \dots, s_{M-1}) na jejich kanonickou konfiguraci (s'_0, \dots, s'_{M-1}) .



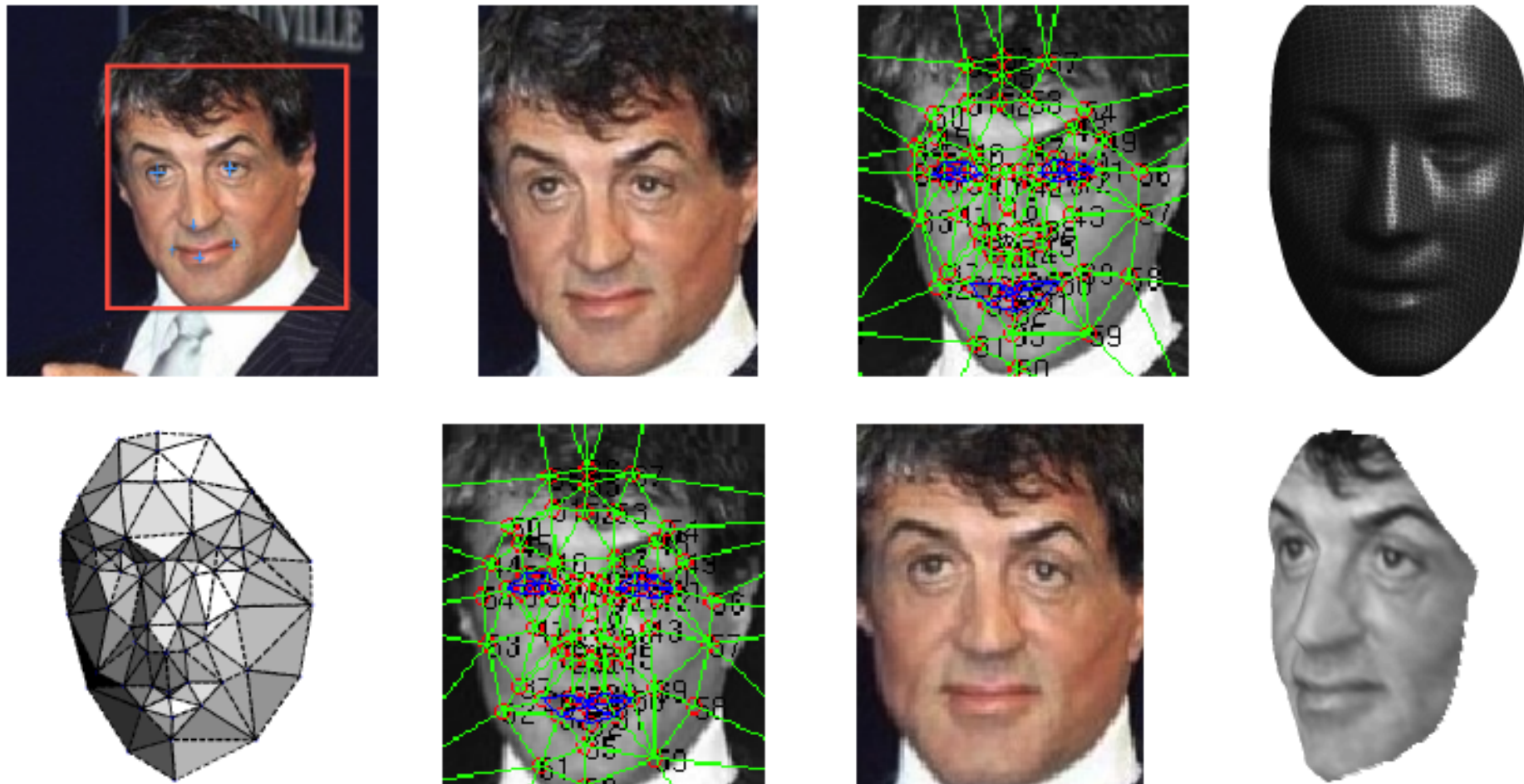
Vstupní obrázek



Normalizovaný obrázek

Geometrická normalizace tváře - frontalizace

- ◆ 3D model se použije k transformaci významných bodů detekovaných ve vstupním obrázku do konfigurace, kterou bychom viděli při frontálním pohledu na tvář.



Obrázek převzat z *Taigman et al: DeepFace: Closing the Gap to Human-Level Performance in Face Verification. CVPR 2014.*

Příznaková reprezentace obrázku

Cílem je získat numerický popis obrázku (příznakový vektor).

Éra I Příznaky extrahované pomocí Principal Component Analysis (PCA) a Linear Discriminant Analysis (LDA) přímo z jasových hodnot obrázku.

Éra II Manuálně navržené lokální deskriptory invariantní vůči změně jasu, měřítka, šumu v obrázku, atd.:

- ◆ Obecné deskriptory používané v počítačovém vidění jako např. Scale Invariant Feature Transform (SIFT) [Lowe 1999], Histogram Of Gradients (HOG) [Dalal & Triggs 2005], Local Binary Patterns (LBP) [Ojala et al 2004], atd.
- ◆ Deskriptory navržené specificky pro rozpoznávání tváří jako např. Biology Inspired Features (BIF) [Guo et al 2009] pro odhad věku.

Éra III Reprezentace obrázku učená automaticky z velkého množství příkladů pomocí “hlubokých” neuronových sítí.

Holistický popis vs. lokální popis

- ◆ **Holistický popis:** celý výřez geometricky normalizované tváře.

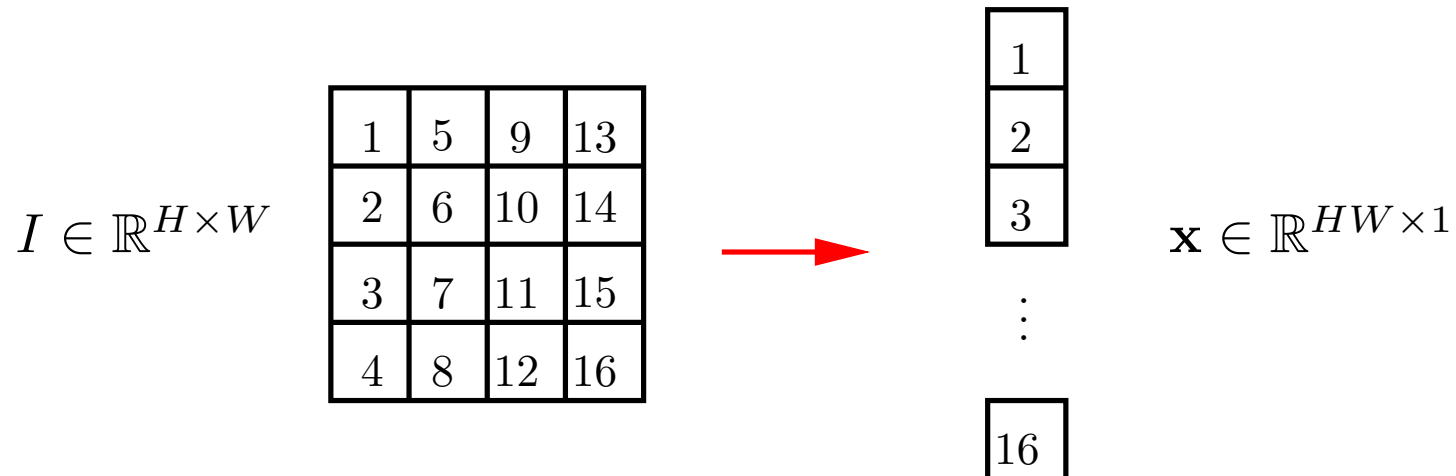


- ◆ **Lokální popis:** geometricky normalizované výřezy kolem významných bodů.



Příznaková reprezentace obrázku - jasové hodnoty

- ◆ Nejjednodušší reprezentace obrázku je použít přímo jasové hodnoty v jednotlivých pixelech jako příznaky



- ◆ Tato příznaková reprezentace je velmi citlivá změnu osvětlení.
- ◆ Pro zvýšení invariance vůči změně osvětlení se používají jasové normalizace. Např:
 - Normalizace na nulovou střední hodnotu a jednotkovou varianci:

$$I'_{ij} = \frac{I_{ij} - \mu}{\sigma}, \quad \mu = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H I_{ij}, \quad \sigma = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H (I_{ij} - \mu)^2$$

- Ekvalizace histogramu.
- ◆ Velká dimenze příznakového vektoru.

Metody extrakce příznaků

- ◆ Obrázek $H \times W$ je reprezentován v prostoru X .
- ◆ Rozhodovací funkce porovnává obrázky pomocí metriky $d: X \times X \rightarrow \mathbb{R}$. Např. klasifikátor tváří podle nejbližšího etalonu

$$h(x) = \operatorname{argmin}_{y=1,\dots,K} d(x, \mu_y)$$

kde μ_1, \dots, μ_K jsou etalony (reprezentanti) třídy 1 až K .

- ◆ Často se používá příznakový prostor $X = \mathbb{R}^n$ a Euklidovská metrika $d(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|$.
- ◆ **Problémy:**
 - Vysoká dimenze příznakového prostoru $X = \mathbb{R}^n$ ($n = HW$ pro obrázek $H \times W$ pixelů; např. $n = 10,000$ pro obrázek 100×100).
 - Některé příznaky nejsou dostatečně diskriminabilní, tj. vzdálenost $d(\mathbf{x}, \mathbf{x}')$ je malá pro odlišné tváře a velká pro podobné tváře.
- ◆ **Metody extrakce příznaků:** transformuj prostor $X = \mathbb{R}^n$ na prostor $Z = \mathbb{R}^p$, tak aby
 1. rekostrukční chyba byla nízká – **Principal Component Analysis**
 2. příznaky byly lépe diskriminabilní – **Linear Discriminant Analysis**

Principal Component Analysis: formulace

- ◆ Cíl: zadané body $\{\mathbf{x}_1, \dots, \mathbf{x}_m\} \in \mathbb{R}^{n \times m}$ chceme aproximovat v p -dimenzionálním affinním podprostoru

$$\tilde{X} = \left\{ \tilde{\mathbf{x}} \in \mathbb{R}^n \mid \tilde{\mathbf{x}} = \sum_{i=1}^p z_i \mathbf{a}_i + \mu \right\} \subseteq \mathbb{R}^n$$

kde $A = \{\mathbf{a}_1, \dots, \mathbf{a}_p\} \in \mathbb{R}^{n \times p}$, $\mu \in \mathbb{R}^n$ jsou parametry \tilde{X} .

- ◆ Projekce bodu \mathbf{x} na affinní podprostor \tilde{X} je bod

$$\tilde{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}' \in \tilde{X}} \|\mathbf{x} - \mathbf{x}'\|$$

K reprezentaci bodu $\tilde{\mathbf{x}} \in \mathbb{R}^n$ je potřeba pouze p souřadnic $\mathbf{z} = (z_1, \dots, z_p)^T \in \mathbb{R}^p$.

- ◆ Cílem je nalézt ortonormální bázi p -dimenzionálního affinního podprostoru \tilde{X} , pro něž je aproximační chyba

$$E(\tilde{X}) = \sum_{i=1}^m \|\mathbf{x}_i - \tilde{\mathbf{x}}_i\|^2$$

minimální.

Principal Component Analysis: řešení

- ◆ Řešením je affinní podprostor jehož bázové vektory $A = \{\mathbf{a}_1, \dots, \mathbf{a}_p\}$ tvoří p vlastních vektorů kovarianční matice

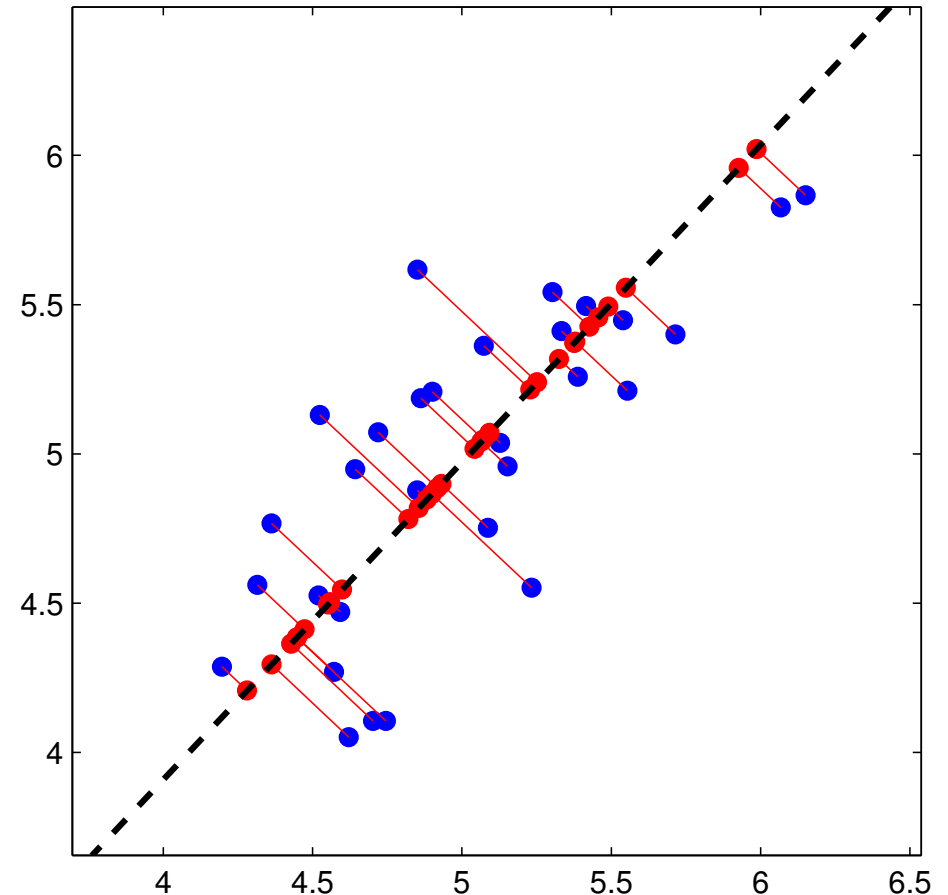
$$\mathbf{C} = \frac{1}{m} \sum_{i=1}^m (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T \quad \text{kde} \quad \mu = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i$$

- ◆ Projekce bodu \mathbf{x} na PCA prostor

$$\mathbf{z} = \mathbf{A}^T (\mathbf{x} - \mu)$$

- ◆ Zpětná projekce

$$\tilde{\mathbf{x}} = \mathbf{A}\mathbf{z} + \mu = \sum_{i=1}^p z_i \mathbf{a}_i + \mu$$

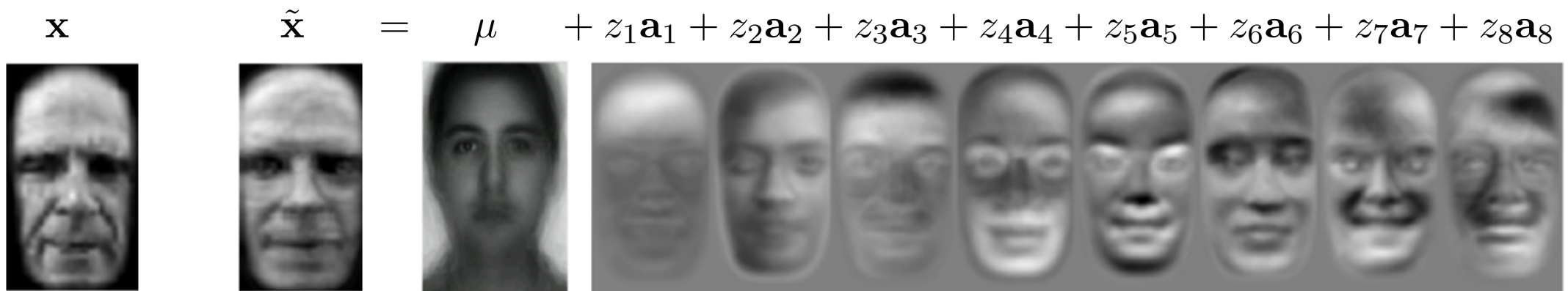


Principal Component Analysis: použití pro reprezentaci tváří

- ◆ Tvář \mathbf{x} se promítne do p -dimenzionálního PCA prostoru

$$\mathbf{z} = \mathbf{A}^T(\mathbf{x} - \mu) = (\underbrace{\mathbf{a}_1^T(\mathbf{x} - \mu)}_{z_1}, \dots, \underbrace{\mathbf{a}_p^T(\mathbf{x} - \mu)}_{z_p})^T$$

- ◆ Tvář \mathbf{x} lze aproximovat jako lineární kombinaci p bázových vektorů (“eigen tváří”)



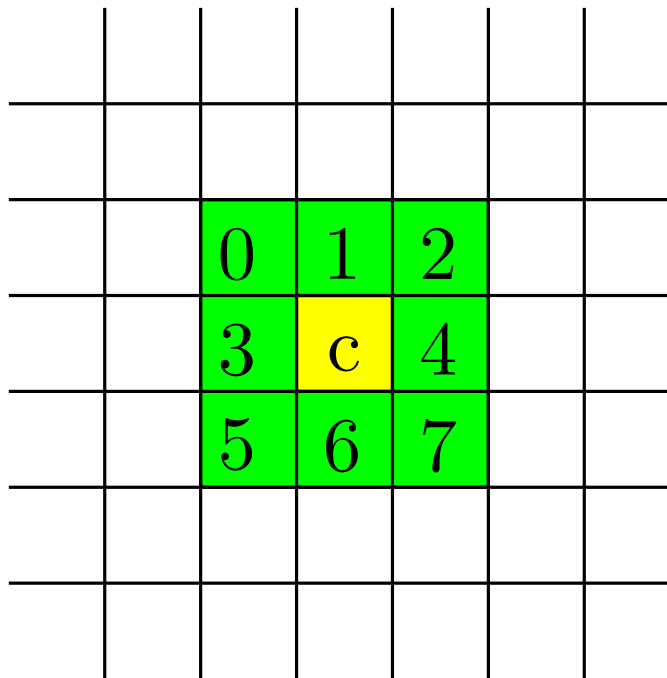
Každý obrázek $\mathbf{x} \in \mathbb{R}^{H \times W}$ lze reprezentovat pomocí $H \cdot W$ čísel.

Každý obrázek $\tilde{\mathbf{x}} \in \mathbb{R}^{H \times W}$ lze reprezentovat pomocí $p = 8$ čísel.

Příznaková reprezentace obrázku - Local Binary Patterns

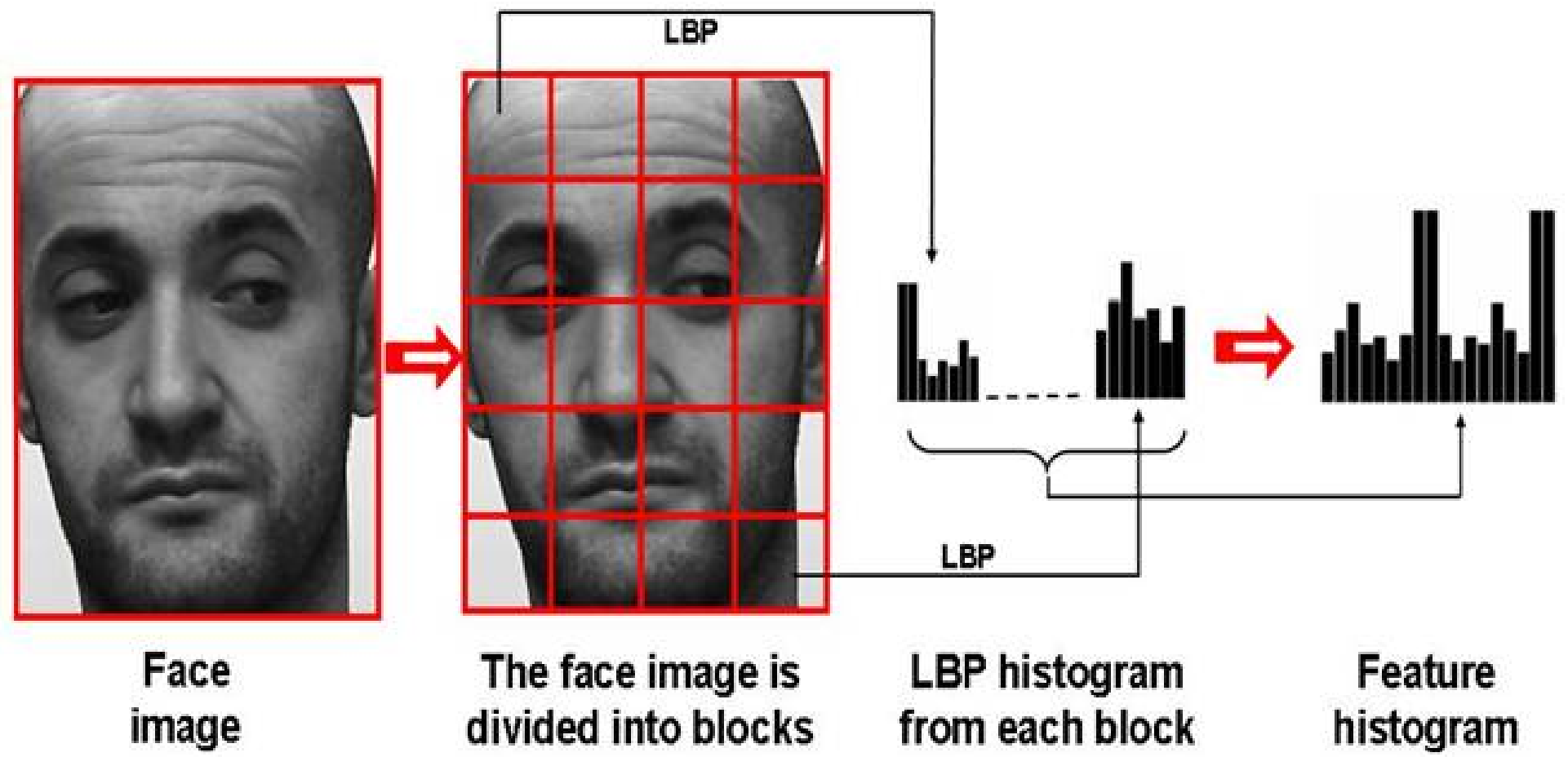
- ◆ LBP přiřadí oknu velikosti 3×3 pixelů 8-bitový kód

$$LBP = \sum_{p=0}^7 [I(c) \geq I(p)] 2^p$$



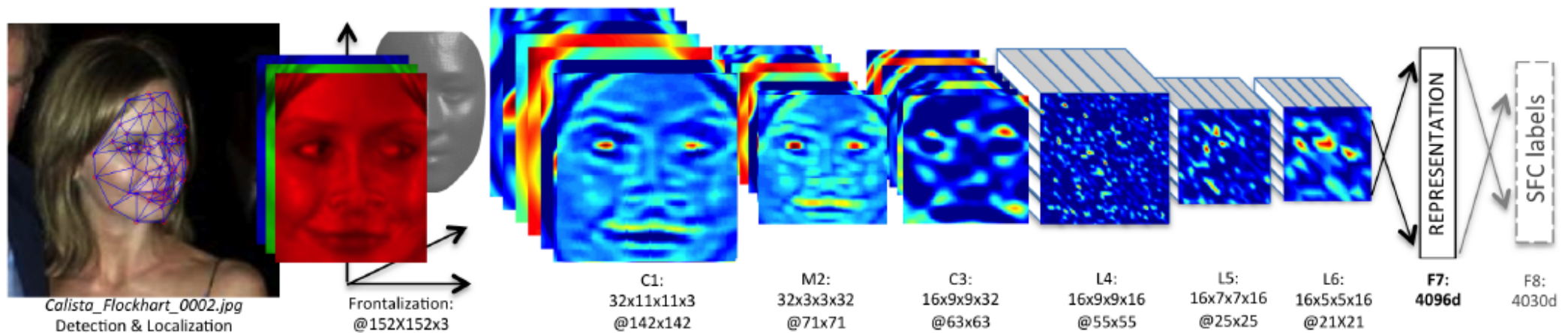
- ◆ LBP příznaky jsou invariantní vůči monotónní změně osvětlení obrázku.

Příznaková reprezentace obrázku - Local Binary Patterns



Příznaková reprezentace učená automaticky z příkladů pomocí hlubokých neuronových sítí

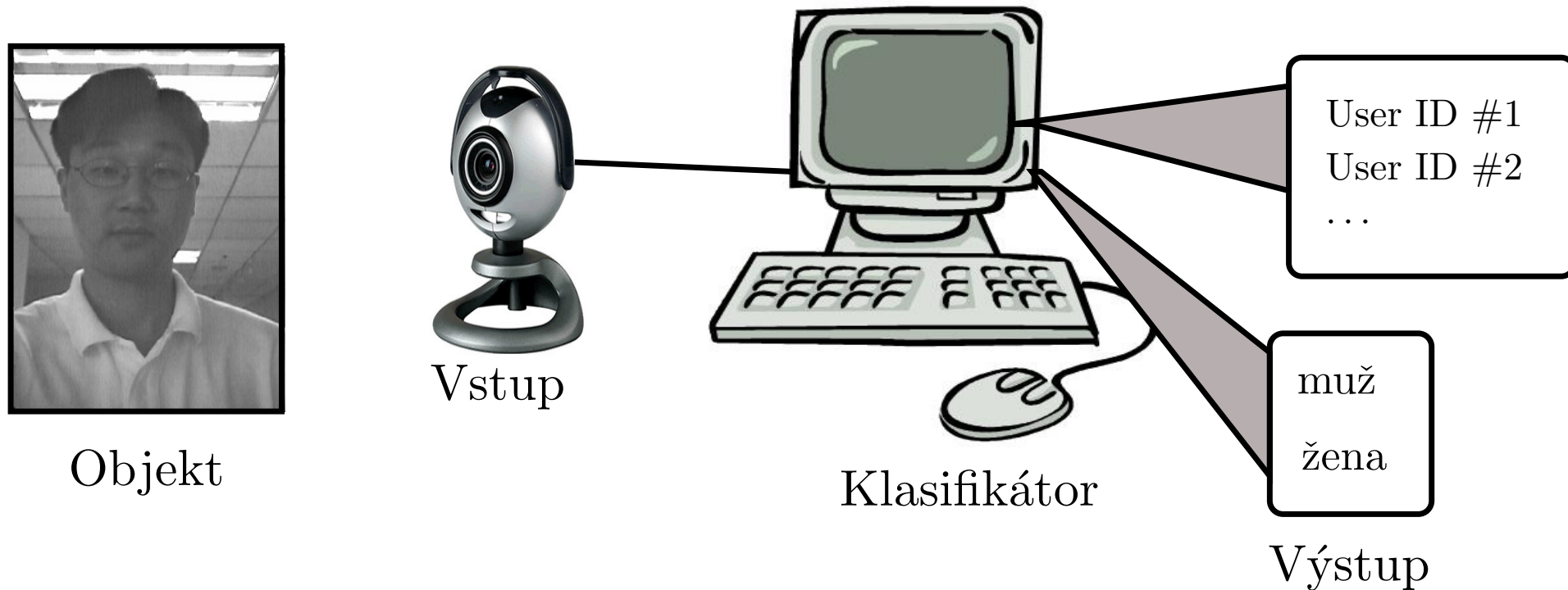
- ◆ Parametry konvoluční hluboké neuronové sítě (CNN) se učí z obrovské množiny příkladů (milióny anotovaných obrázků tváří) pomocí gradientní optimalizace (výpočet gradientu neuronové sítě = back propagation).
- ◆ CNN se naučí klasifikovat tváře to relativně malé množiny tříd; třídy odpovídají identitám lidí.
- ◆ Výstupy předposlední vrstvy se použijí jako příznakový popis vstupní tváře pro libovolnou rozhodovací úlohu, např. pro verifikaci tváří.



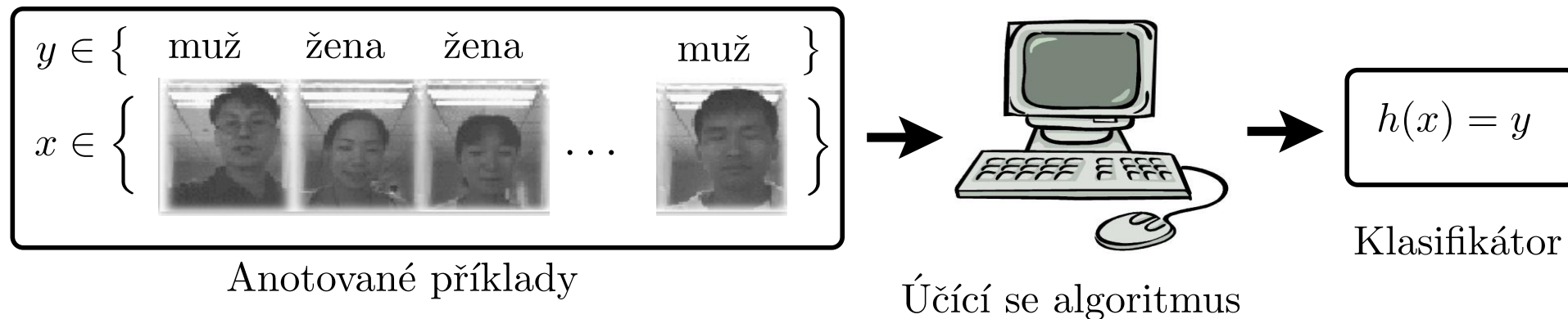
Obrázek převzat z *Taigman et al: DeepFace: Closing the Gap to Human-Level Performance in Face Verification. CVPR 2014.*

Klasifikace a učení

- ◆ Příklad klasifikace tváří do tříd:



- ◆ Metody učení klasifikátorů z příkladů:



Klasifikace a učení

Statistický model: objekt je popsán rozdělením pravděpodobnosti $p(x, y)$ kde

- ◆ $x \in X$... příznakový popis klasifikovaného obrázků např. lokální příznaky nebo výstup hluboké NN sítě.
- ◆ $y \in Y$... skrytých stav klasifikovaného objektu.
Např. pohlaví osoby na obrázku, věk, identita.

Cílem je nalézt klasifikátor $h: X \rightarrow Y$, který pro danou ztrátovou funkci $\ell: Y \times Y \rightarrow \mathbb{R}$ minimalizuje očekávané riziko

$$R_{\text{exp}}(h) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \ell(y, h(x))$$

Problém je, že místo rozdělení $p(x, y)$ máme k dispozici jen trénovací příklady z něj náhodně a nezávisle generované

$$\{(x_1, y_1), \dots, (x_m, y_m)\} \in (X \times Y)^m$$

Učení: Trénovací příklady se použijí k nalezení:

1. Odhadu statistického modelu $p(x, y)$ a následně ke konstrukci klasifikátoru $h(x)$, tzv. generativní učení.
2. Klasifikátoru $h(x)$, tzv. diskriminativní učení.

Příklad klasifikačního problému:

Odhad pohlaví z obrázku zaregistrované tváře

- ◆ $x \in X$... např. LBP příznaky spočítané na geometricky normalizované lidské tváři
- ◆ $y \in Y = \{\text{muž, žena}\}$
- ◆ Pokud chceme minimální pravděpodobnost chybné klasifikace zvolíme $\ell(y, h(x)) = \mathbb{1}[y \neq h(x)]$ a pak

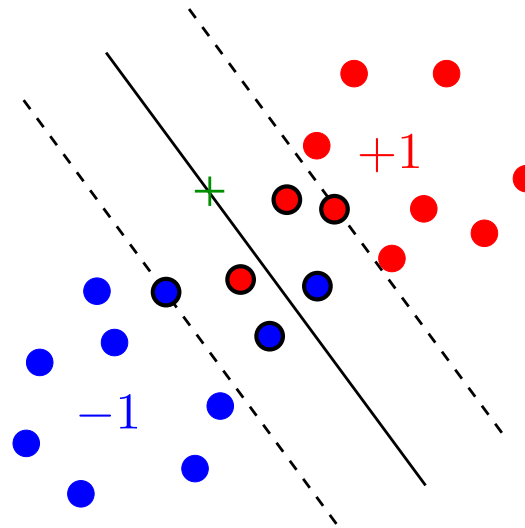
$$\begin{aligned}
 R_{\text{exp}}(h) &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \ell(y, h(x)) \\
 &= \sum_{x \in X} p(x, y = \text{muž}) \mathbb{1}[h(x) = \text{žena}] + \sum_{x \in X} p(x, y = \text{žena}) \mathbb{1}[h(x) = \text{muž}]
 \end{aligned}$$



Linear Support Vector Machines

- ◆ Klasifikátor $h: \mathbb{R}^n \rightarrow \{-1, +1\}$

$$h(\mathbf{x}; \mathbf{w}) = \text{sign}(\langle \mathbf{x}, \mathbf{w} \rangle)$$



- ◆ Učení parametrů \mathbf{w} z příkladů $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\} \in (\mathbb{R}^n \times \{+1, -1\})$ je formulováno jako konvexní problém

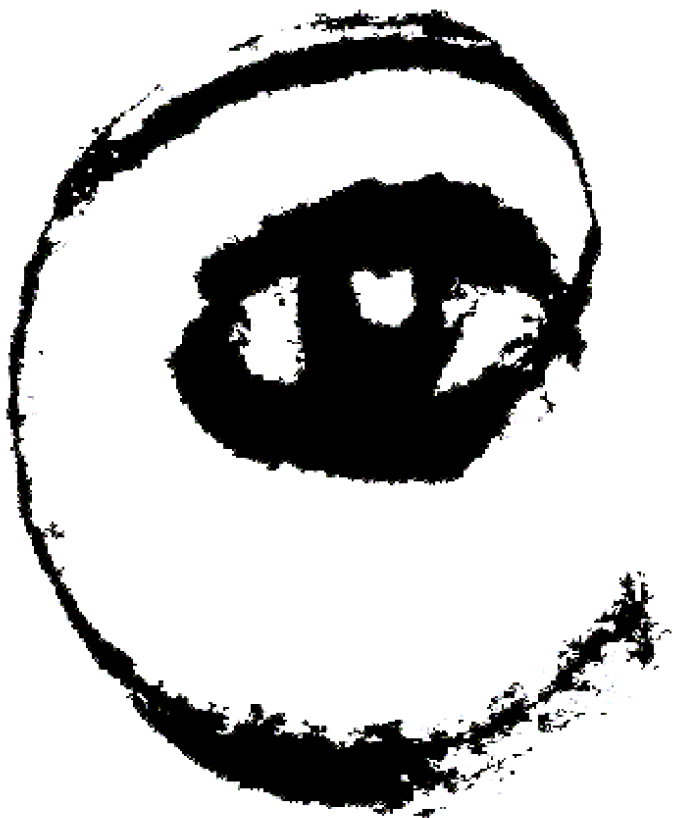
$$\mathbf{w}^* = \underset{\mathbf{w} \in \mathbb{R}^n}{\text{argmin}} \left[\frac{\lambda}{2} \|\mathbf{w}\|^2 + \sum_{i=1}^m \max\{0, 1 - y_i \langle \mathbf{w}, \mathbf{x}_i \rangle\} \right]$$

kde $\Omega(\mathbf{w})$ je kvadratický regularizátor a $R(\mathbf{w})$ je konvexní horní mez pro

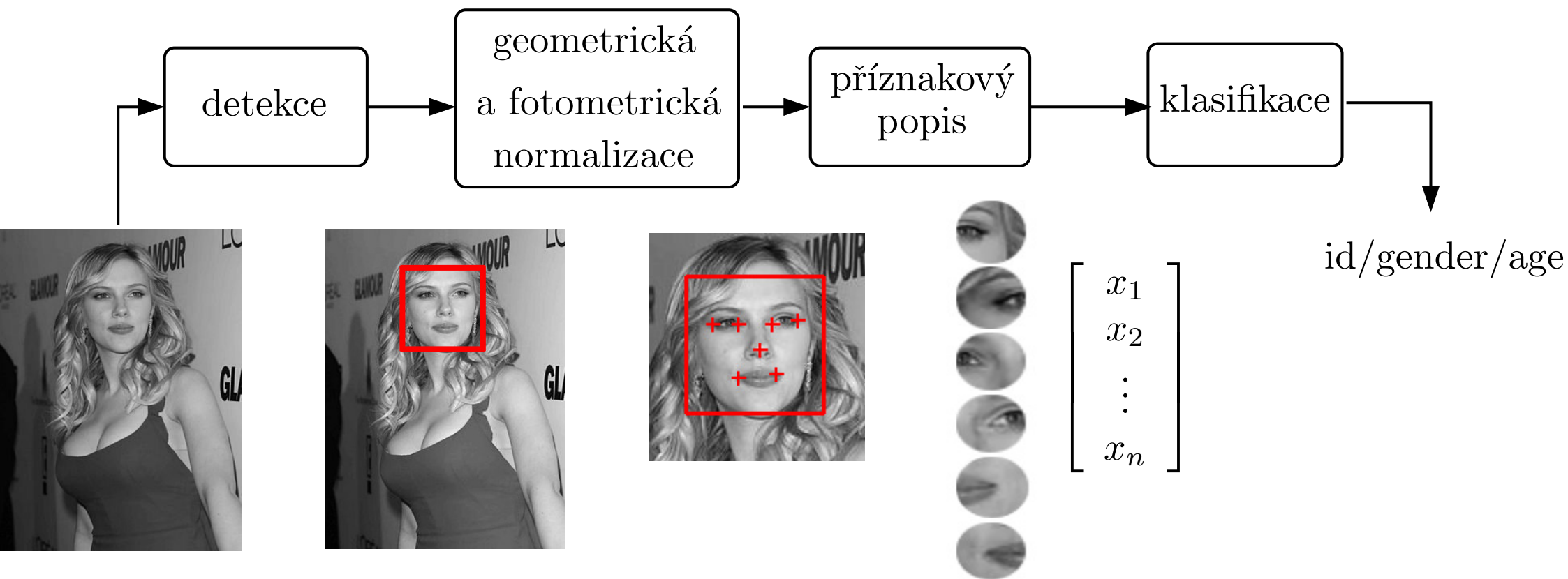
$$\sum_{i=1}^m [y_i \neq h(\mathbf{x}_i; \mathbf{w})]$$

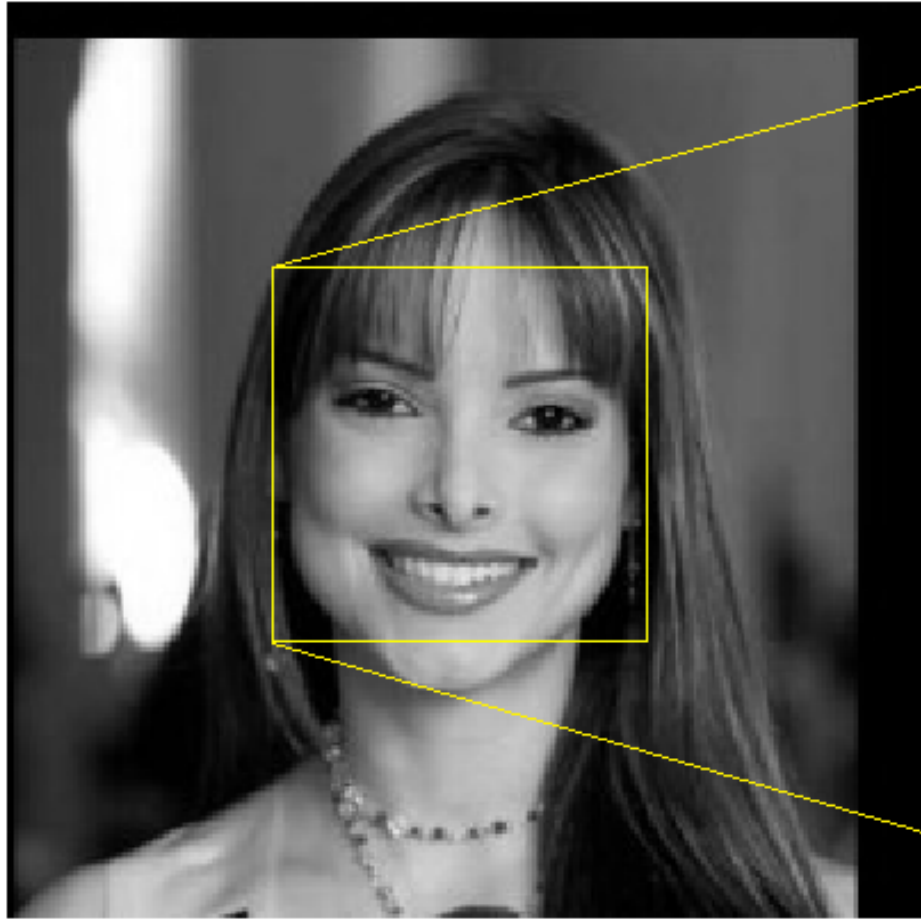
tj. počet chyb, kterých se klasifikátor $h(\mathbf{x}; \mathbf{w})$ dopustí na trénovací příkladech.

Konec

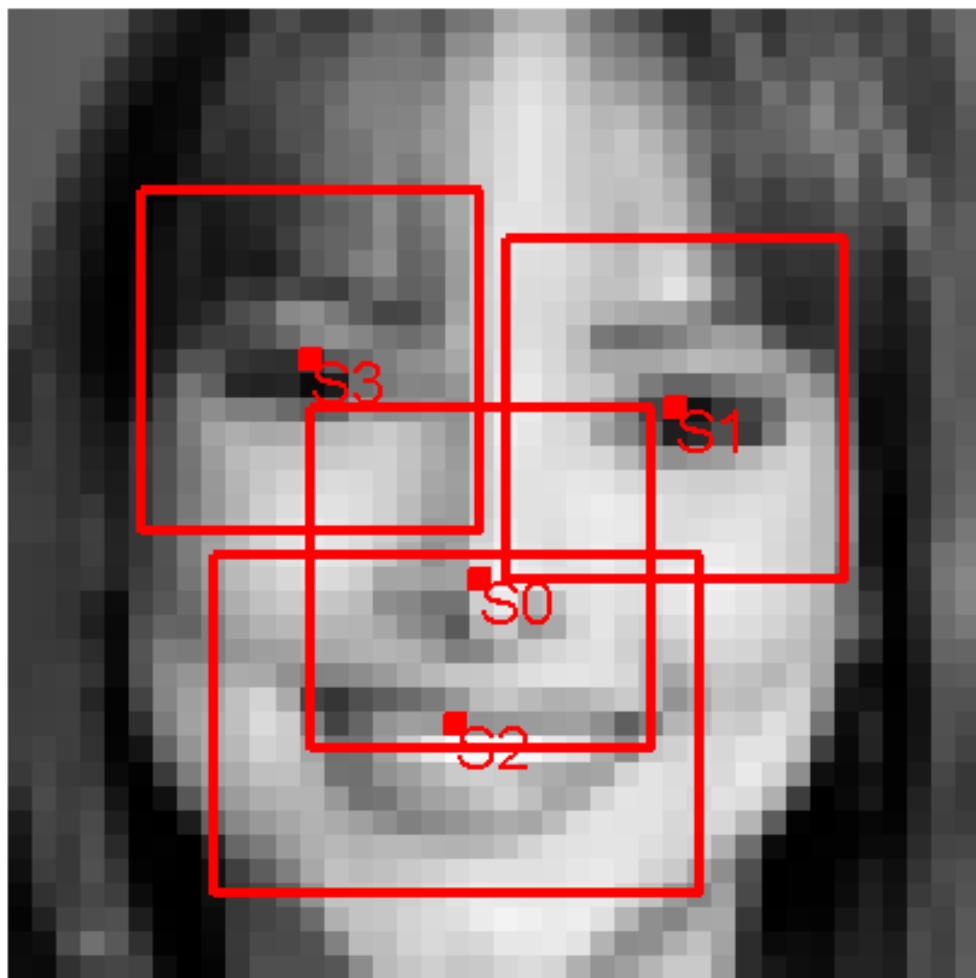


m p



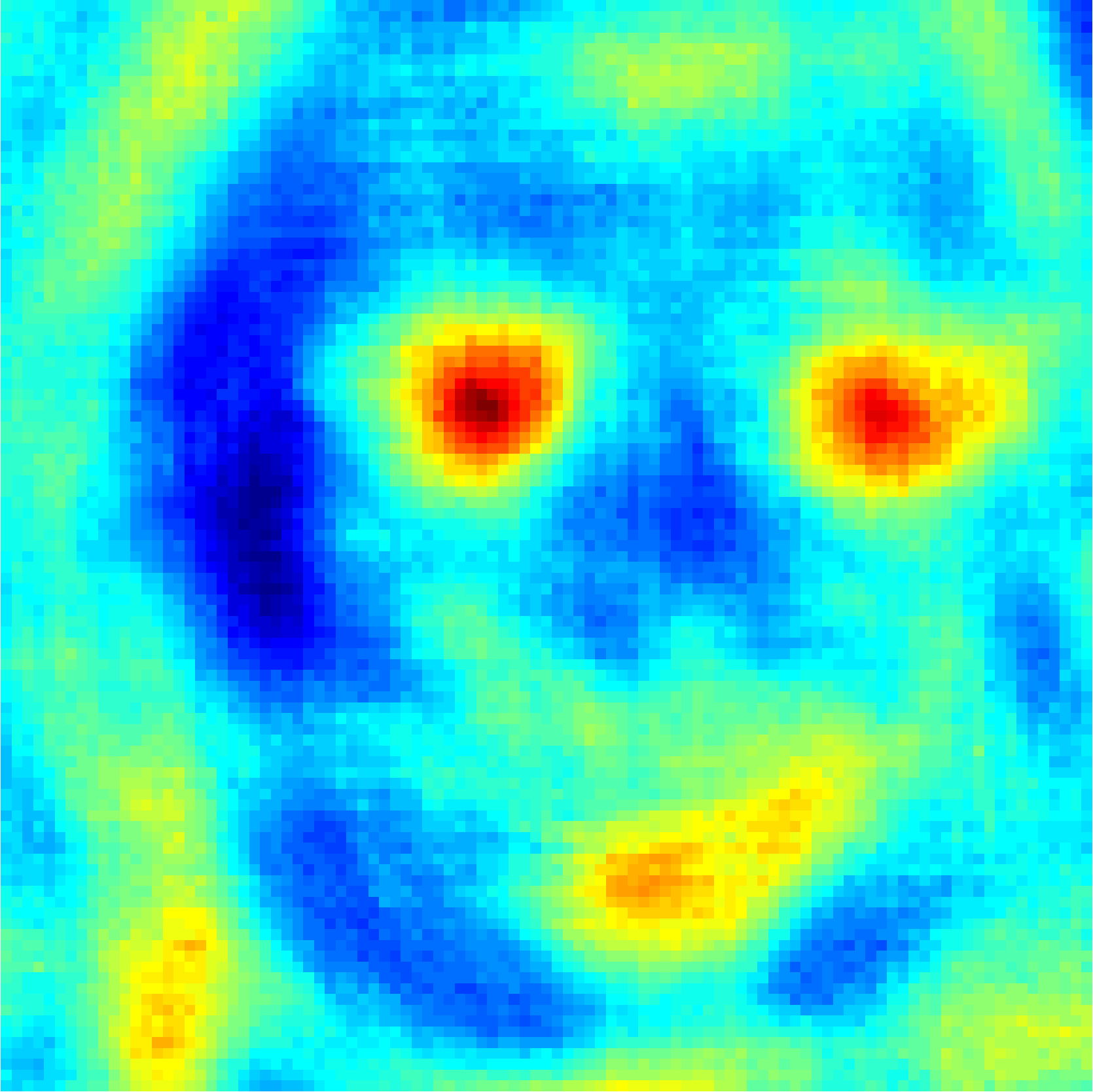


H

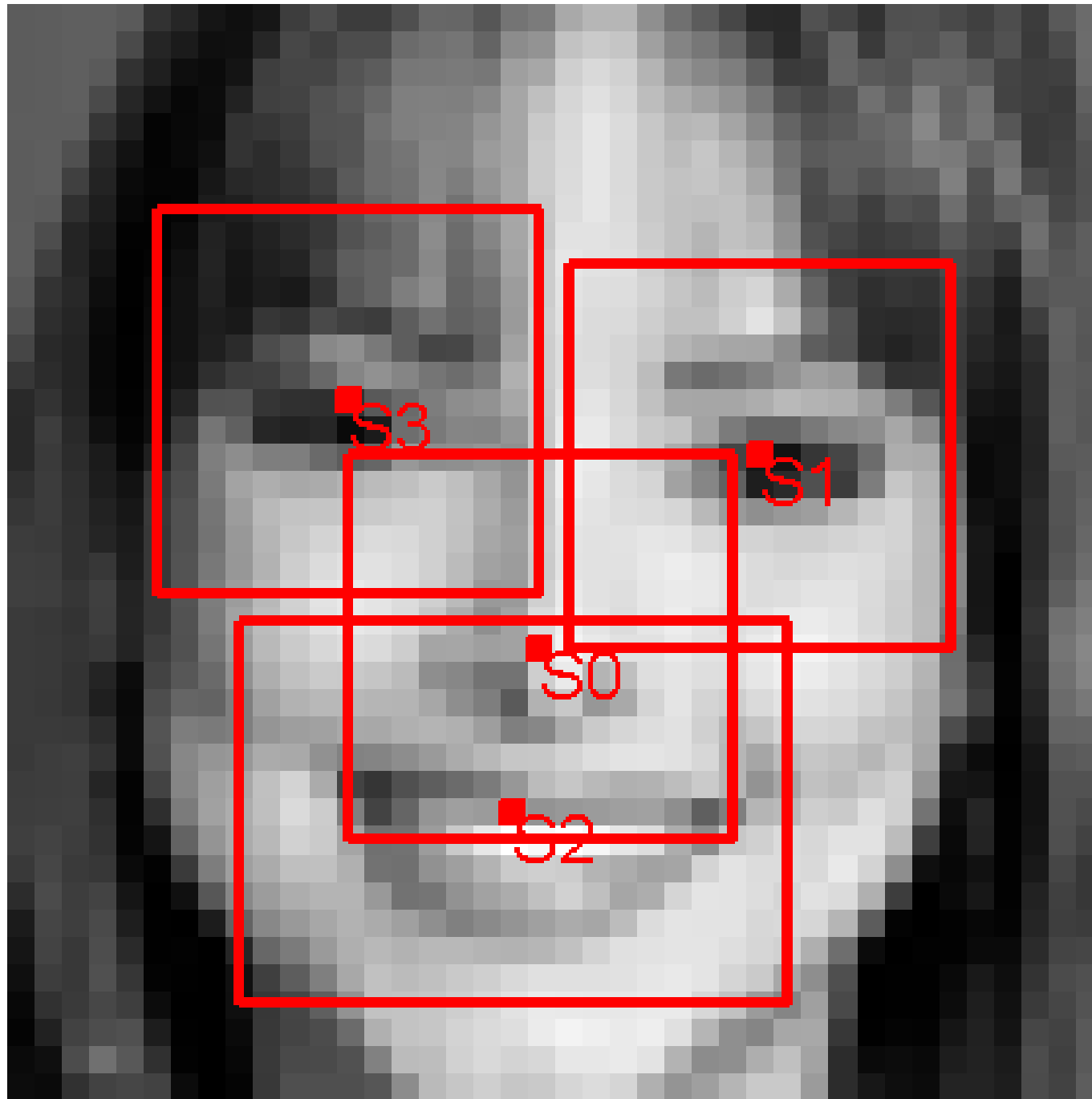


W

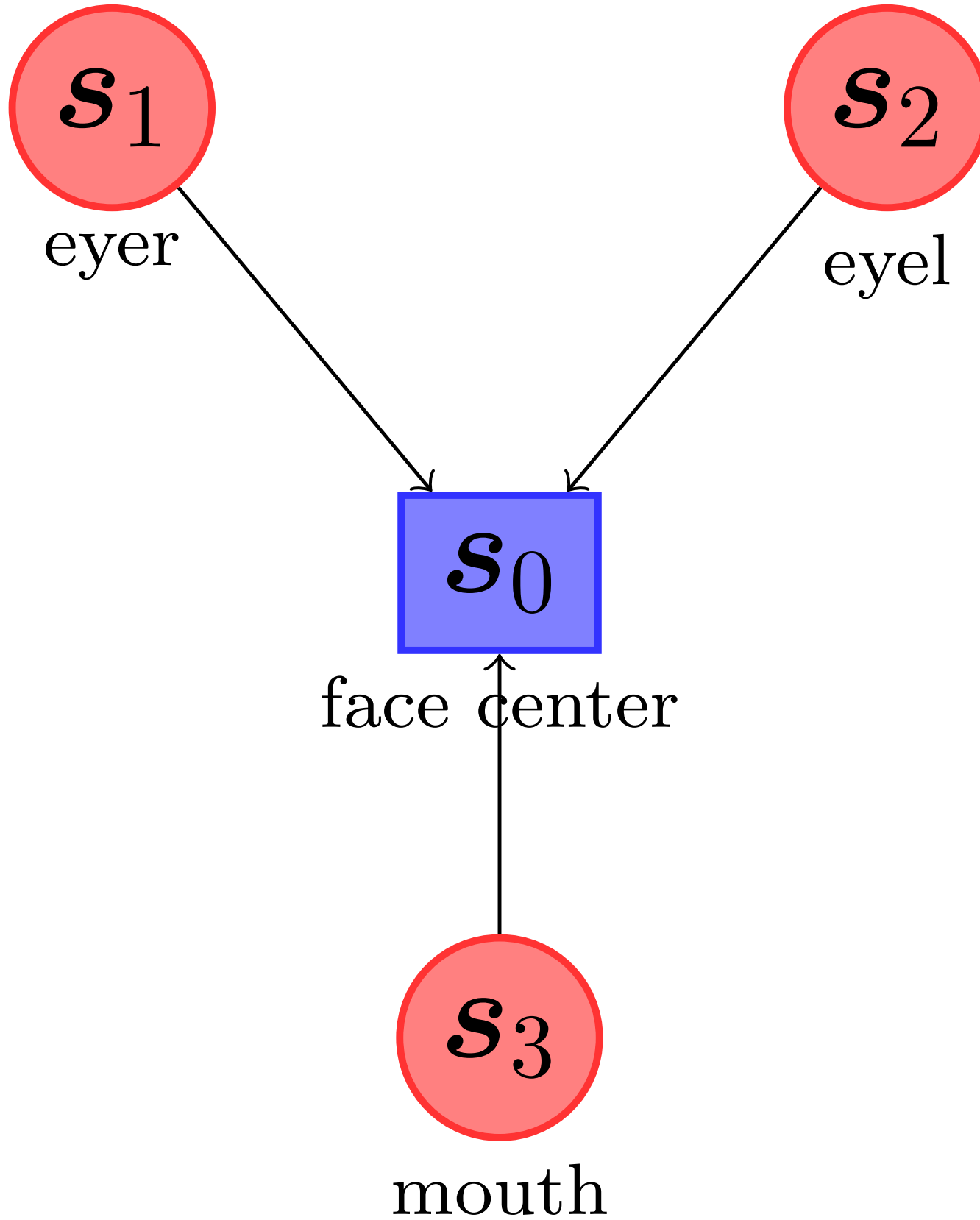


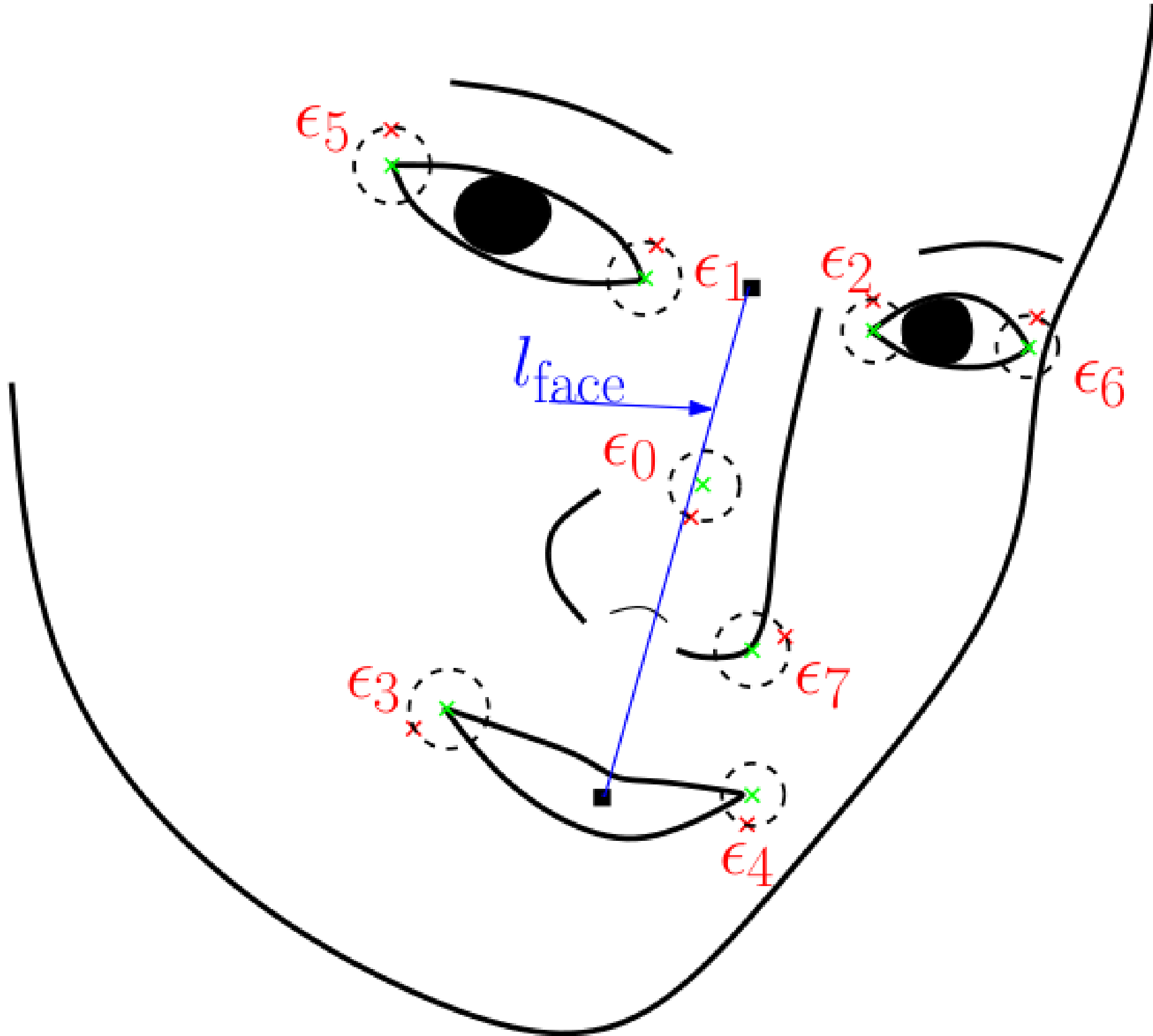


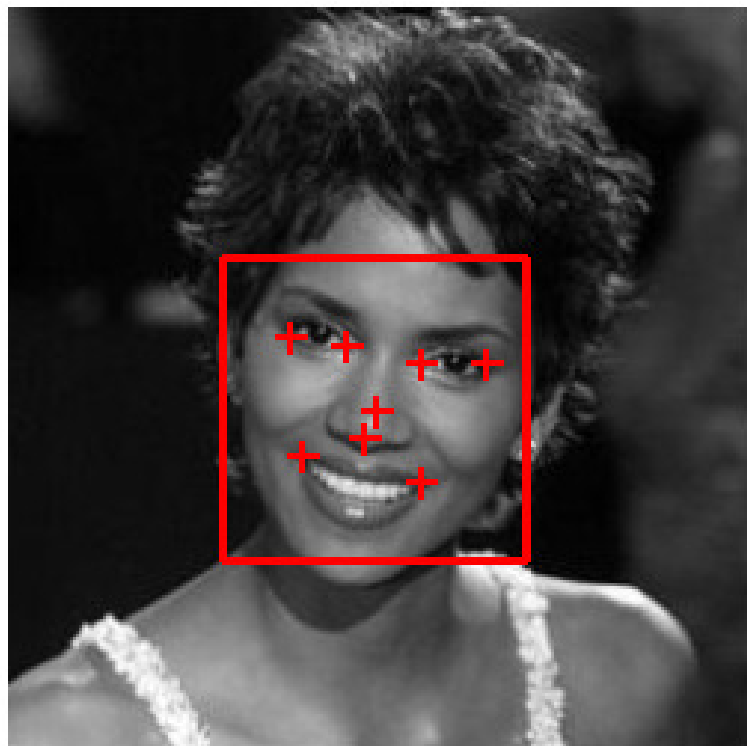
H

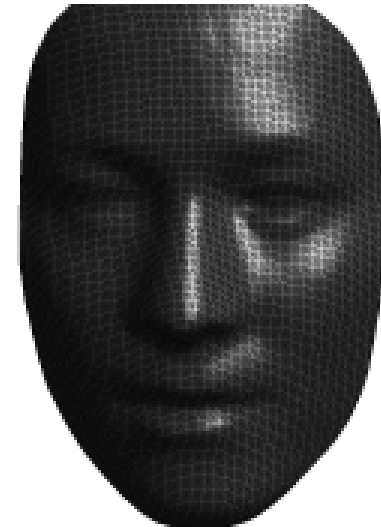
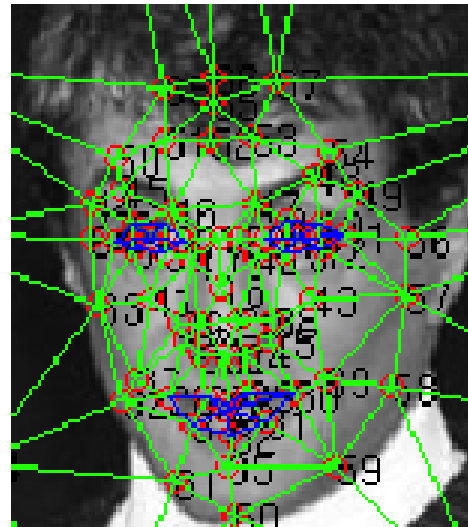


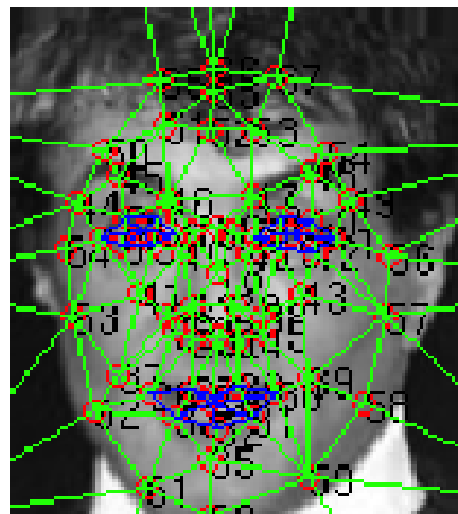
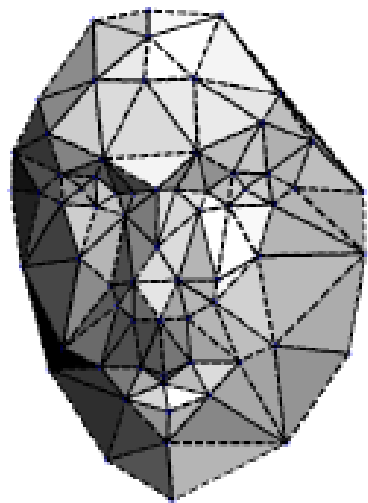
W

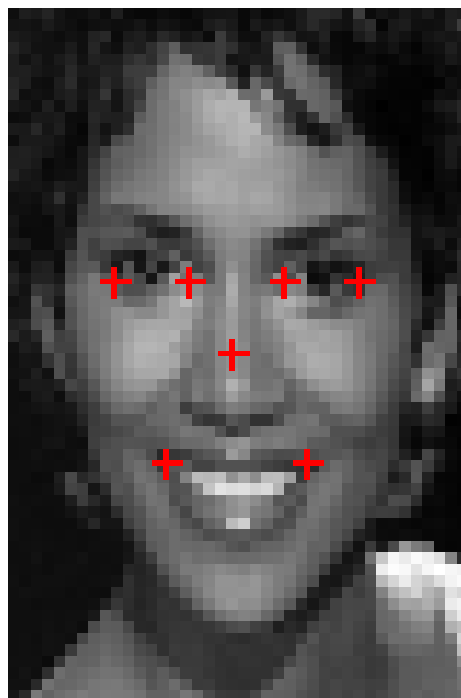
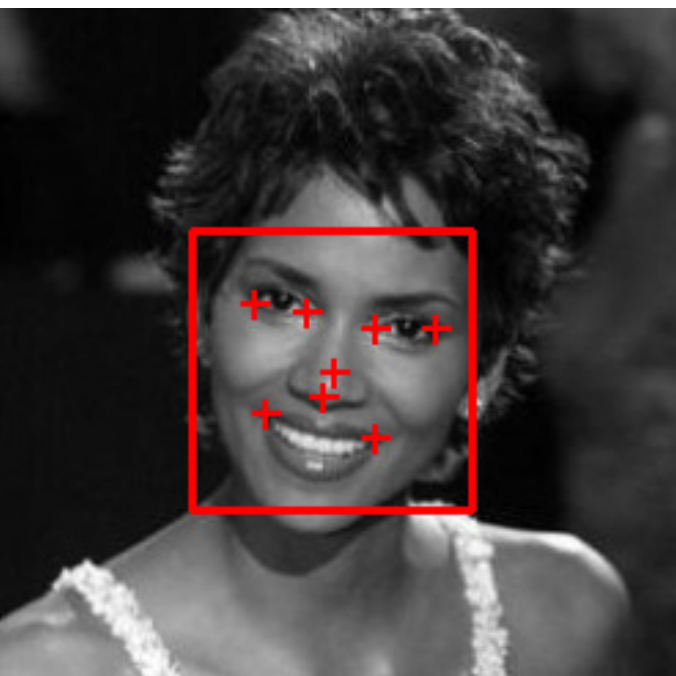


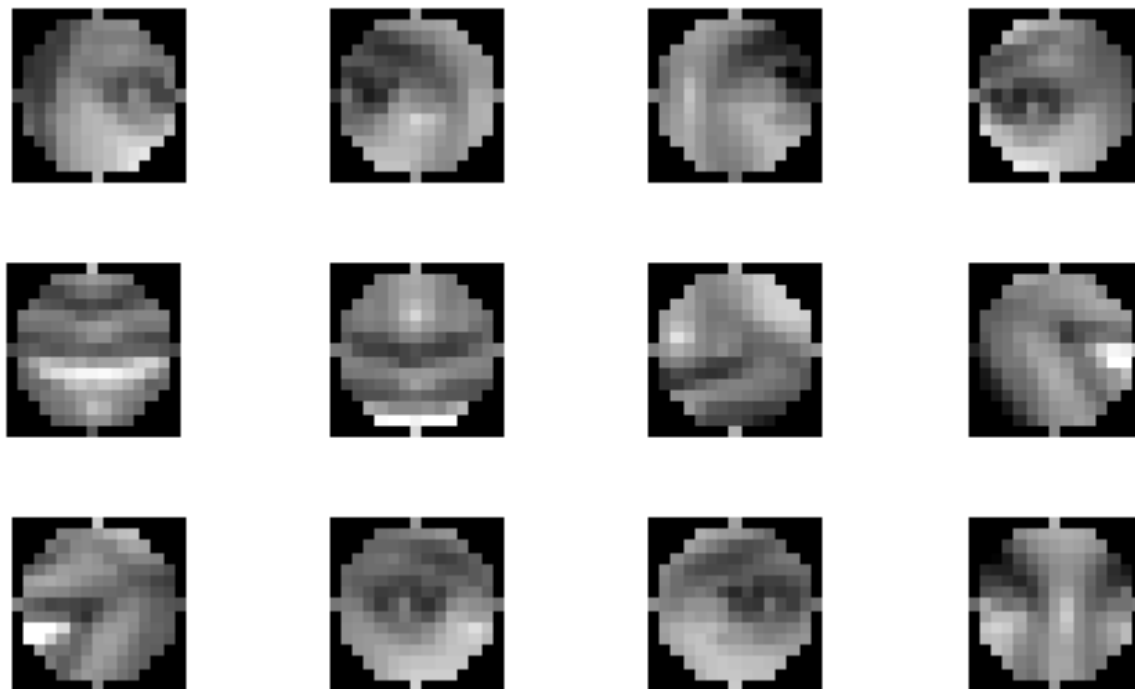
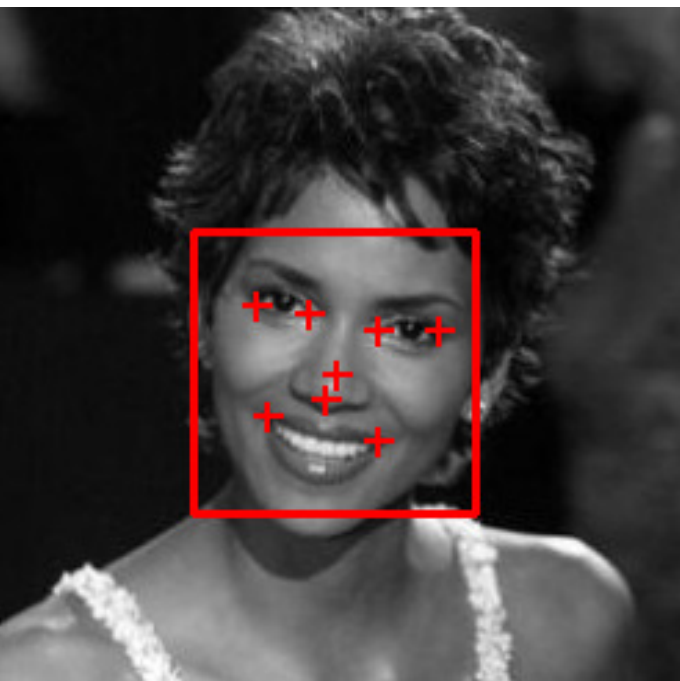








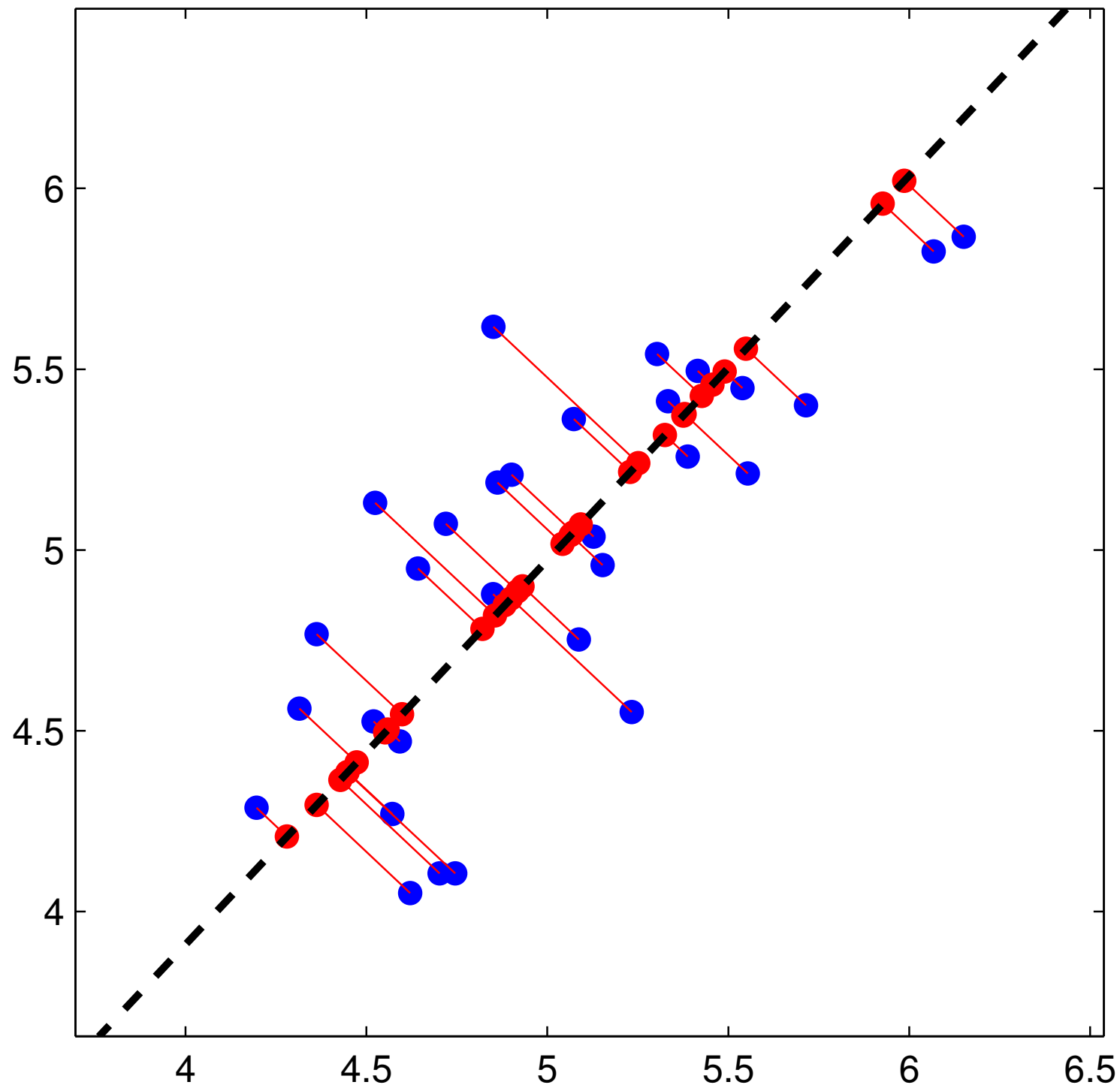




1	5	9	13
2	6	10	14
3	7	11	15
4	8	12	16



1
2
3
⋮
16

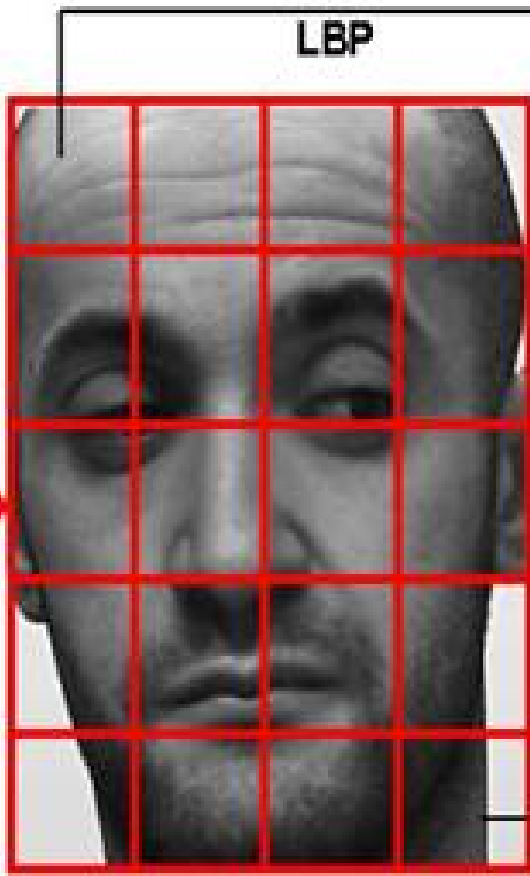




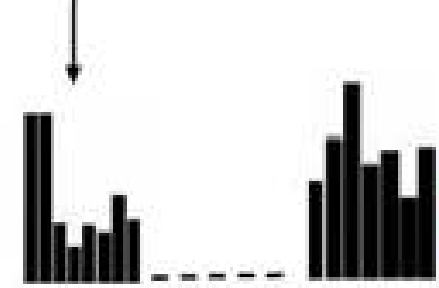
		0	1	2	
		3	c	4	
		5	6	7	



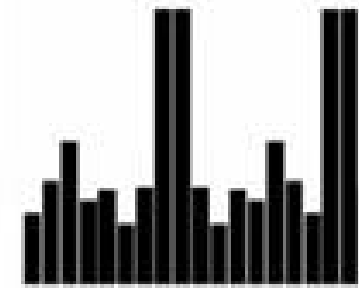
Face image



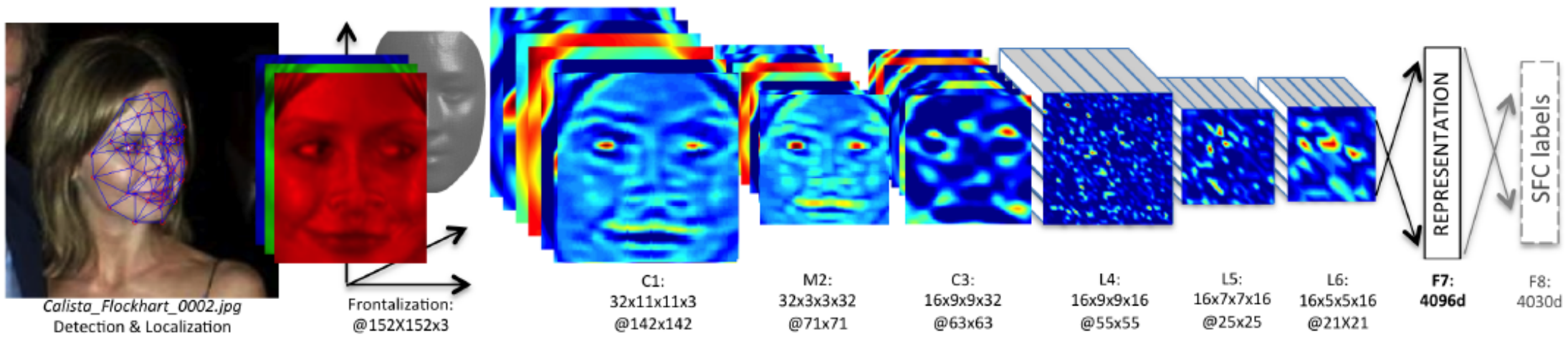
The face image is divided into blocks



LBP histogram from each block



Feature histogram





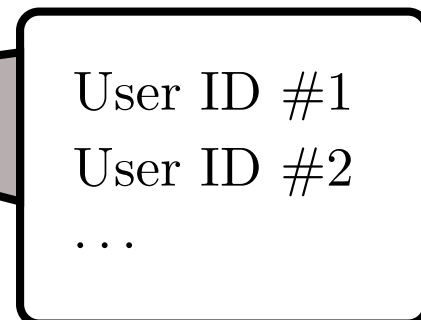
Objekt



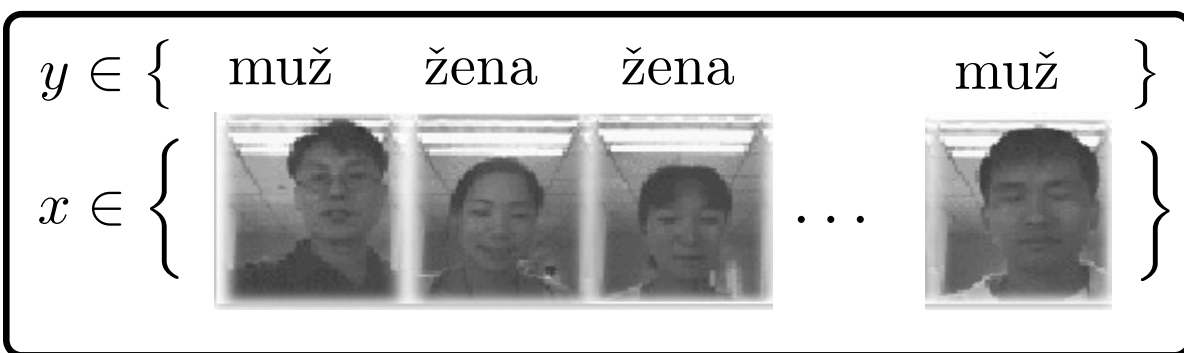
Vstup



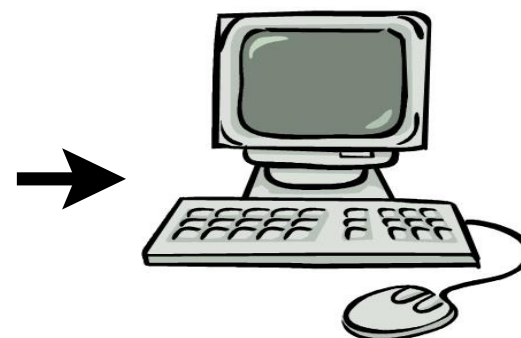
Klasifikátor



Výstup



Anotované příklady



Účící se algoritmus

$h(x) = y$

Klasifikátor





