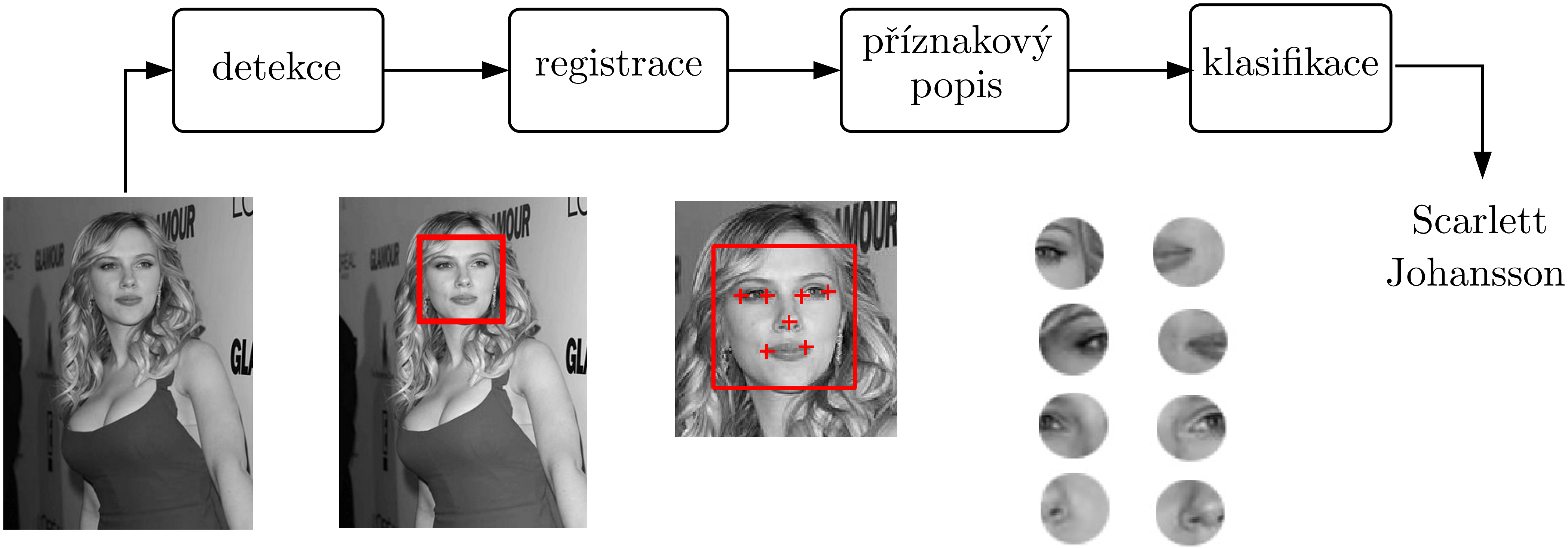
Rozpoznávání tváří II

Vojtěch Franc
Centrum strojového vnímání, ČVUT FEL Praha

Biometrie ZS 2012

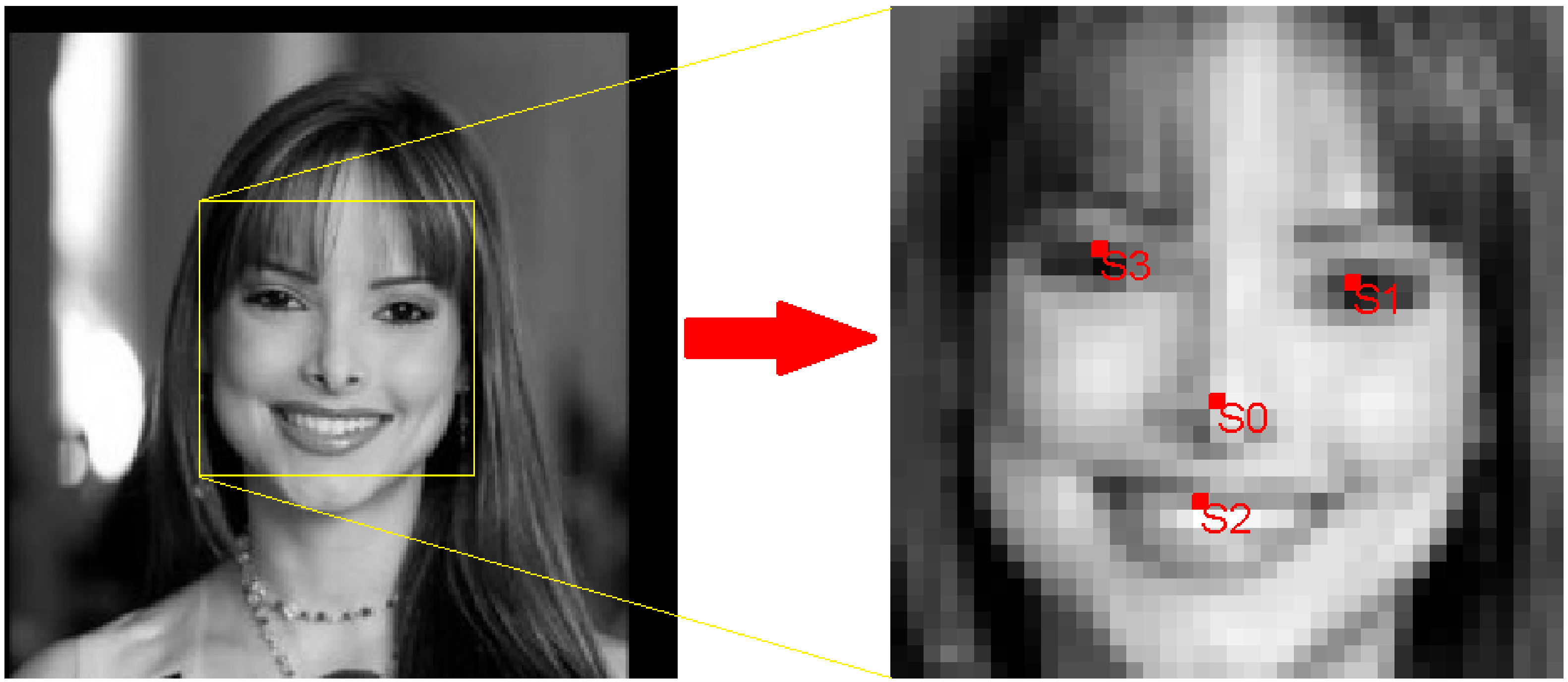
Poděkování Michalu Uřičařovi za obrázky popisující detekci významných bodů

Stavební bloky typického rozpoznávacho systému



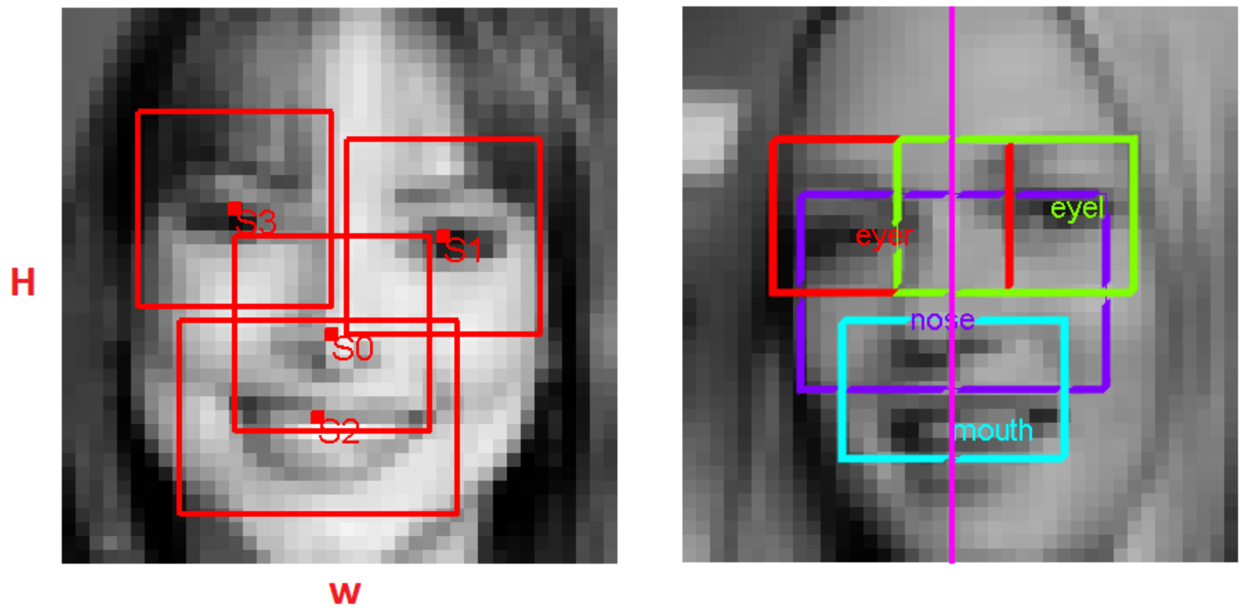
Detekce významných bodů na tváři

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



Nezávislý detektor pro každý významý 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
- $q_i(I, \mathbf{s}_i)$ "pravděpodnost", že i-tý významný bod je v obrázku I na souřadnici $\mathbf{s}_i \in S_i$



- Odhady pozic $\{\hat{\mathbf{s}}_0, \dots, \hat{\mathbf{s}}_{M-1}\}$ významných bodů v obrázku I lze nálézt nézávislými detektory

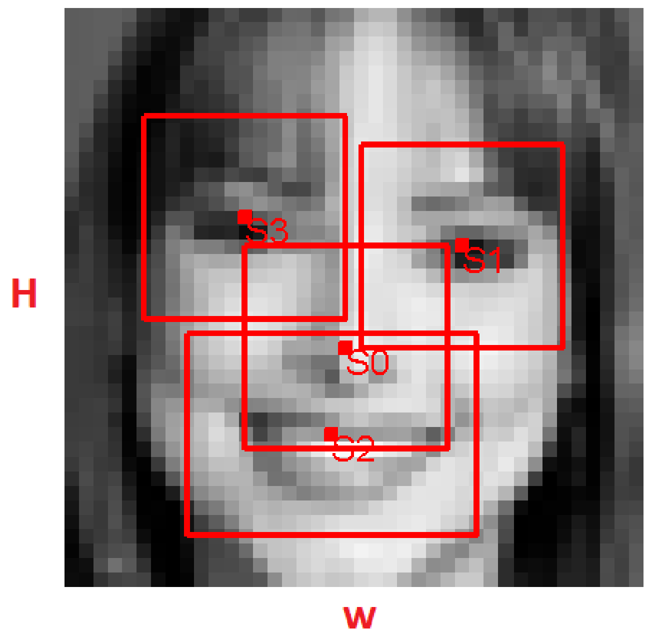
$$\hat{\mathbf{s}}_i = \operatorname*{argmax}_{\mathbf{s} \in S_i} q_i(I, \mathbf{s}) \qquad i \in V$$

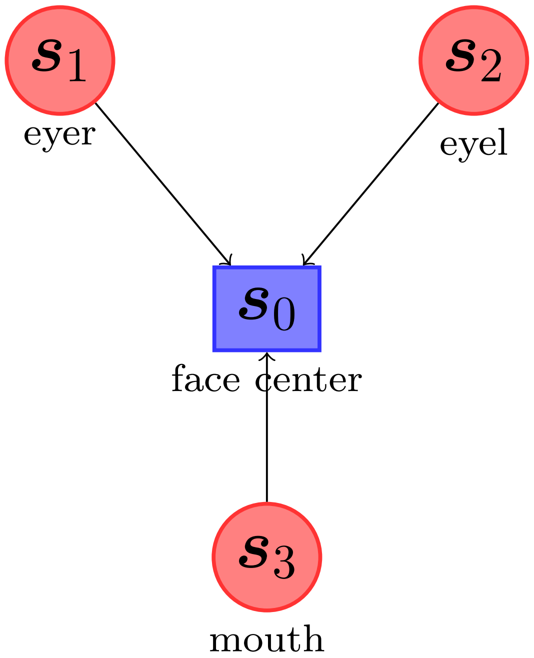
- Nevýhoda: nevyužívá se toho, že pozice významných bodů nejsou nezávislé.

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 $\mathbf{s} = (\mathbf{s}_0, \ldots, \mathbf{s}_{M-1})$ v obrázku I

$$f(I, \mathbf{s}) = \underbrace{\sum_{i \in V} q_i(I, \mathbf{s}_i)}_{\text{shoda s obrázkem}} + \underbrace{\sum_{(i,j) \in E} g_{ij}(\mathbf{s}_i, \mathbf{s}_j)}_{\text{deformační skóre}}$$





Deformable part model

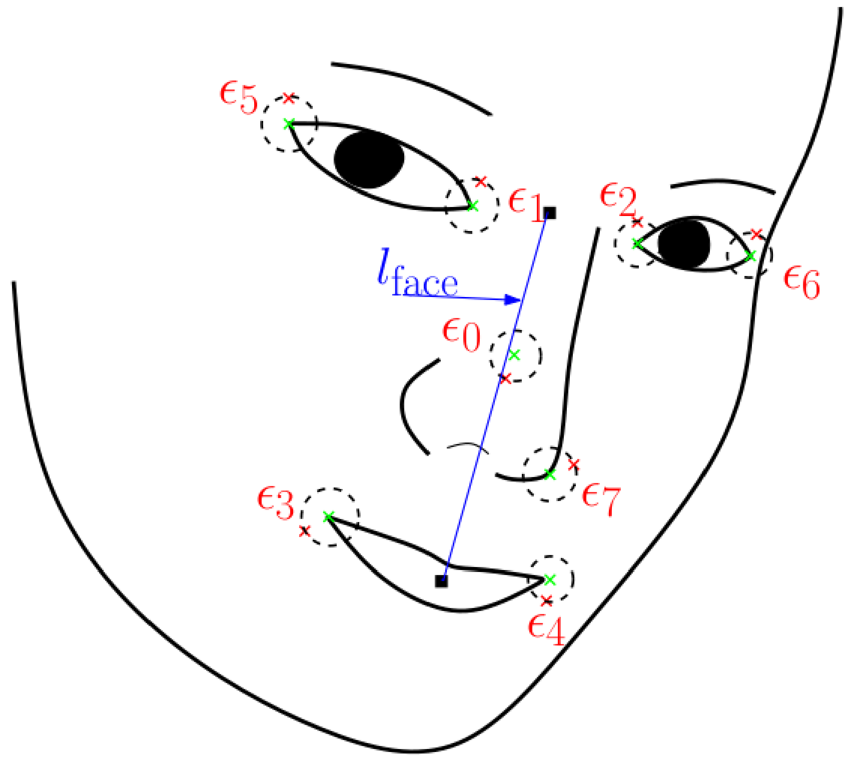
- Odhad pozice významných bodů v obrázku I vede na hledání konfigurace

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

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

- Funkce q_i a g_{ij} lze učit z množiny $\{(I^1,\mathbf{s}^1),\ldots,(I^m,\mathbf{s}^m)\}$, 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.



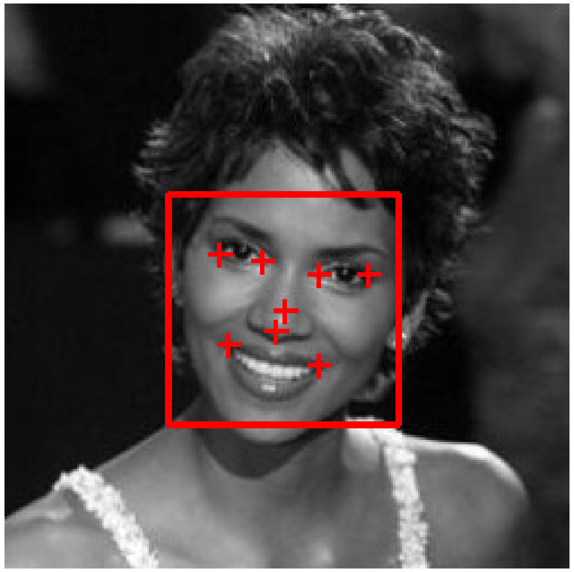
$$\text{odchylka} = \frac{\varepsilon_0+\cdots+\varepsilon_{M-1}}{M}\cdot\frac{1}{l_{face}}$$

Geometrická normalizace tváře

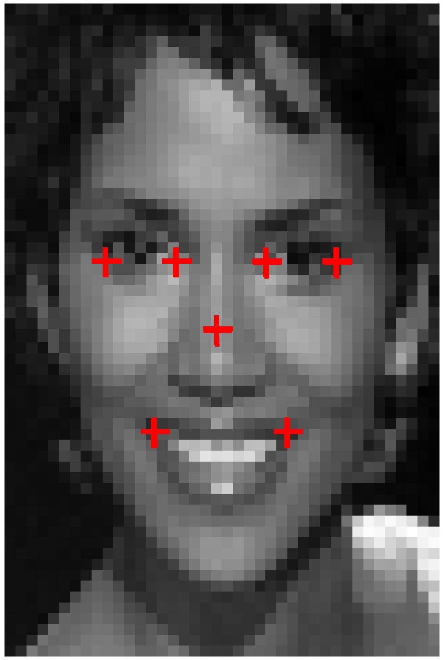
- Cíl: převést tvář do normalizované polohy
- Jeden z možných způsobů: nalezneme affinní transformaci (např. metodou nejmenších čtverců)

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

  která promítne nalezené významné body $(\mathbf{s}_0, \ldots, \mathbf{s}_{M-1})$ na jejich normalizovanou konfiguraci $(\mathbf{s}'_0, \ldots, \mathbf{s}'_{M-1})$.



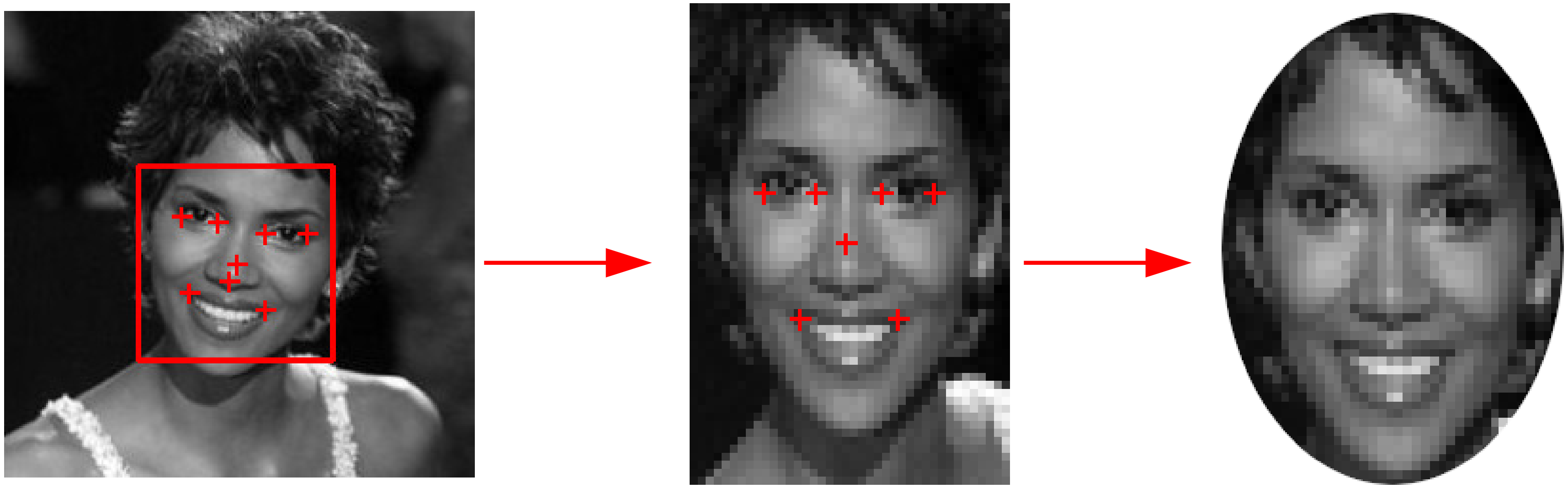
Vstupní obrázek



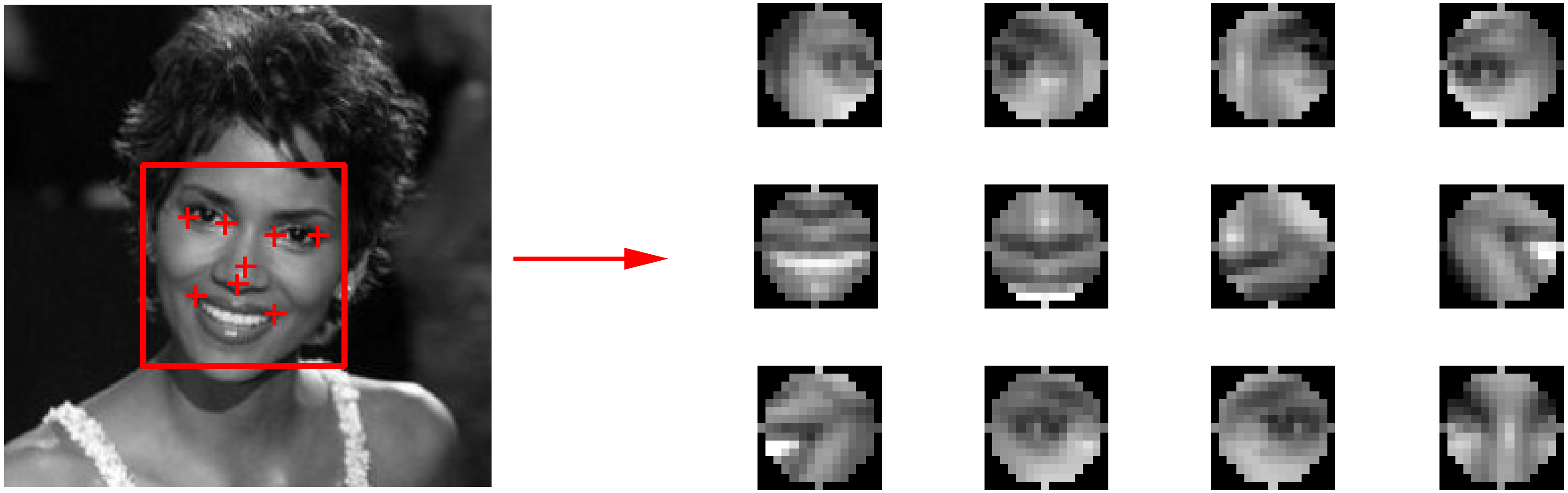
Normalizovaný obrázek

Holistický popis vs. lokalní 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

$$I \in \mathbb{R}^{H\times W} \quad \begin{array}{|c|c|c|c|} \hline 1 & 5 & 9 & 13 \\ \hline 2 & 6 & 10 & 14 \\ \hline 3 & 7 & 11 & 15 \\ \hline 4 & 8 & 12 & 16 \\ \hline \end{array} \quad \longrightarrow \quad \begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline 3 \\ \hline \end{array} \; \vdots \; \begin{array}{|c|} \hline 16 \\ \hline \end{array} \quad \mathbf{x} \in \mathbb{R}^{HW\times 1}$$

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

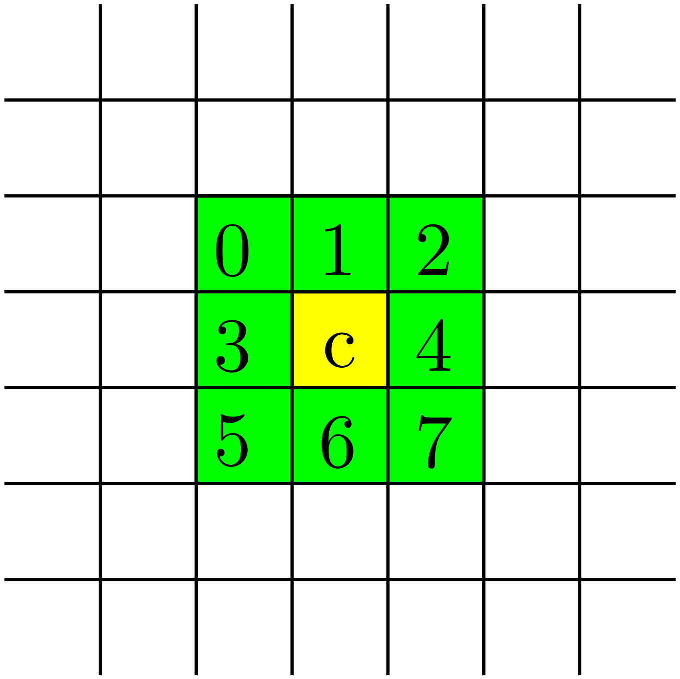
$$I'_{ij} = \frac{I_{ij} - \mu}{\sigma}, \qquad \mu = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H} I_{ij}\,, \qquad \sigma = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}(I_{ij} - \mu)^2$$

  - Ekvalizace histogramu.

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

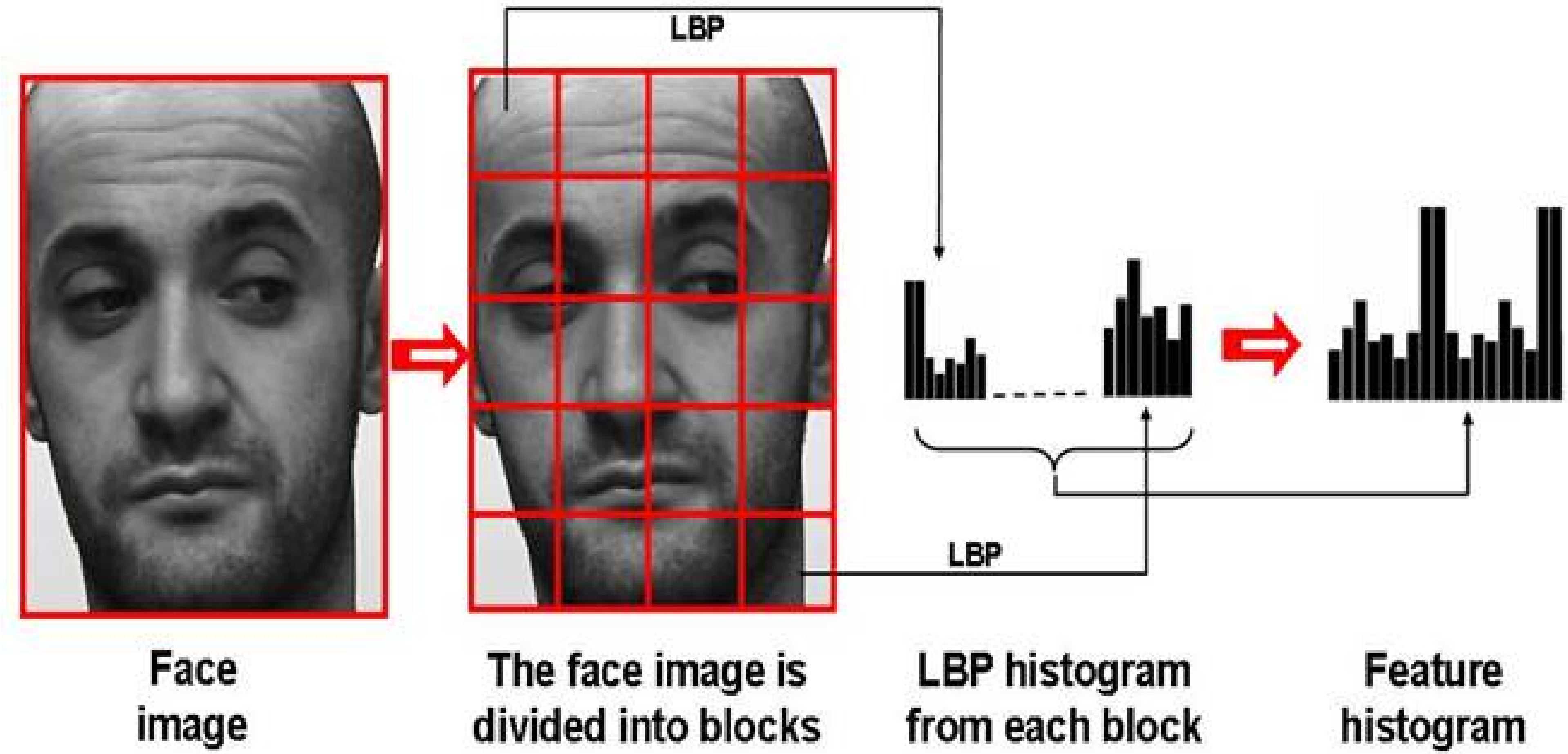


Image downloaded from http://www.scholarpedia.org/article/Local_Binary_Patterns

Metody extrakce příznaků

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

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

kde $\mu_1, \ldots, \mu_K$ 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).
  - 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 (PCA)

- 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 nějž je aproximační chyba

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

minimální.

Principal Component Analysis (PCA)

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

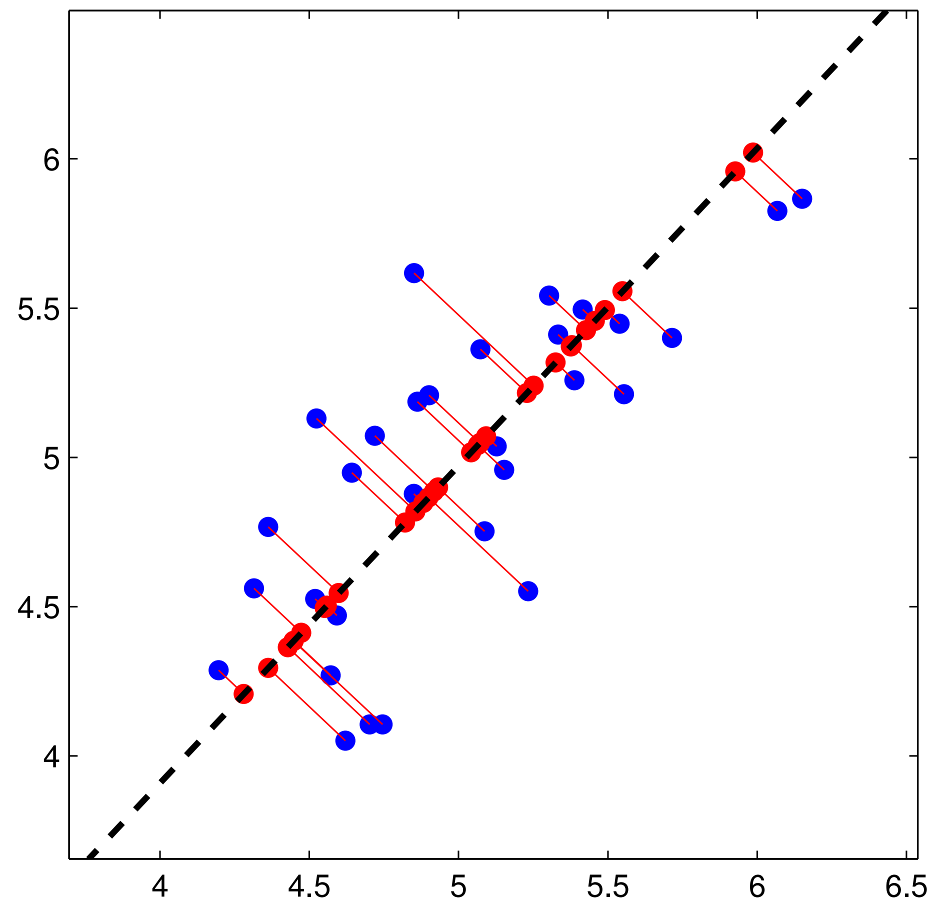
$$\mathbf{C} = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T \qquad \text{kde} \qquad \mu = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i$$

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

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

- Zpětná projekce

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



Eigen faces - Aplikace PCA na tváře

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

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

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

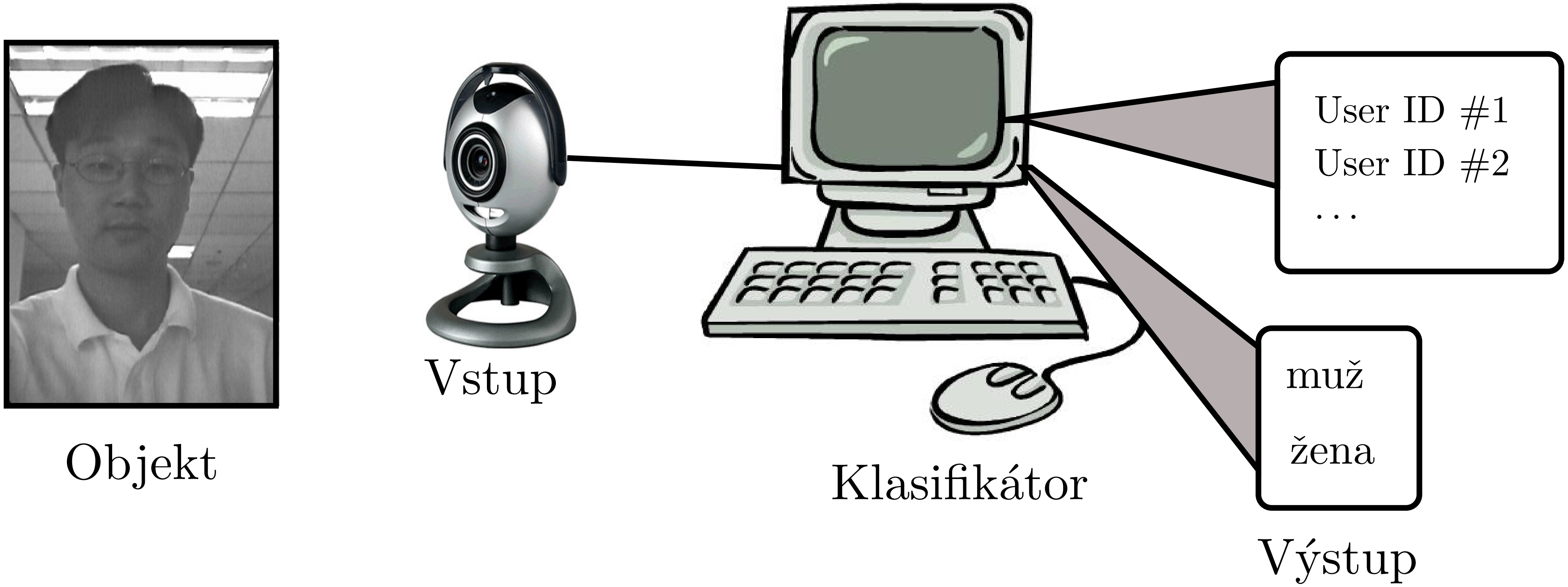
$$\mathbf{x} \qquad \tilde{\mathbf{x}} = \mu + z_1\mathbf{a}_1 + z_2\mathbf{a}_2 + z_3\mathbf{a}_3 + z_4\mathbf{a}_4 + z_5\mathbf{a}_5 + z_6\mathbf{a}_6 + z_7\mathbf{a}_7 + z_8\mathbf{a}_8$$

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.

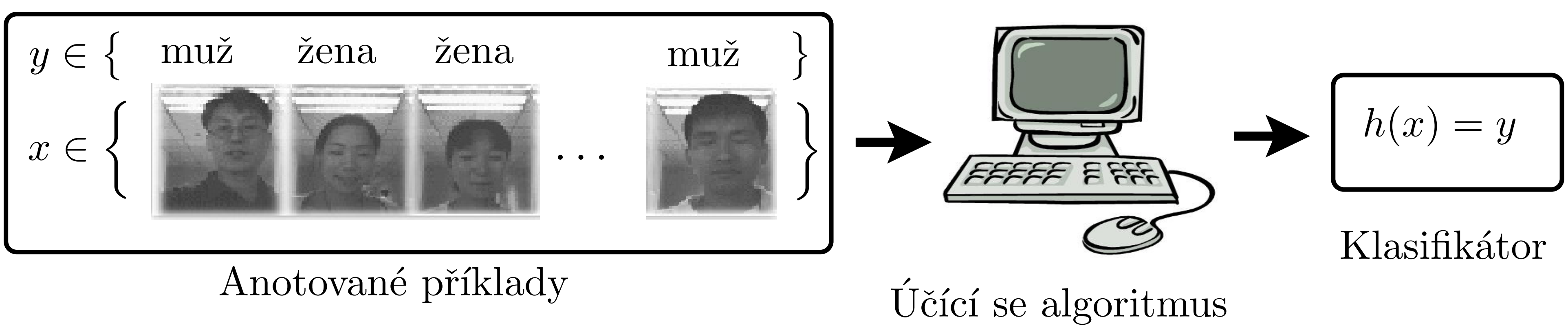
Image downloaded from http://vision.stanford.edu/teaching/cs231a/lecture/lecture2_face_recognition_cs231a_marked.pdf

Klasifikace a učení - neformální popis problému

- Příklad klasifikace tváří do tříd:



- Metody učení klasifikátorů z příkladů:



Klasifikace - formální (statistický) popis

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

- $x \in X$ ... popis klasifikovaného obrázku.
  Např. příznaky spočítané na geometricky normalizované tváři.
- $y \in Y$ ... skrytých stav klasifikovaného objektu.
  Např. pohlaví osoby na obrázku, věk, identita.

Cílem je nalézt klasifikátor $h\colon X \to Y$, který pro danou ztrátovou funkci $\ell\colon Y \times Y \to \mathbb{R}$ 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 příklady z něj náhodně a nezávisle generované

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

Příklad klasifikačního problému: Odhad pohlaví z obrázku zaregistrované tváře

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

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

Metody učení klasifikátorů z příkladů

Generativní učení

- Na základě příkladů $\{(x_1, y_1), \ldots, (x_m, y_m)\}$ najdeme odhad $\hat{p}(x, y)$ zkutečného rozdělení $p(x, y)$. Např. použijeme princip maxima věrohodnosti.
- Do vzorce pro očekávané riziko $R_{\text{exp}}(h)$ dosadíme $\hat{p}(x, y)$ a hledáme klasifikátor, který tuto aproximaci skutečného rizika minimalizuje

$$h^* \in \operatorname*{argmin}_h \hat{R}(h) \qquad \text{kde} \qquad \hat{R}(h) = \sum_{x \in X} \sum_{y \in Y} \hat{p}(x, y)\, \ell(y, h(x))$$

Diskriminativní učení

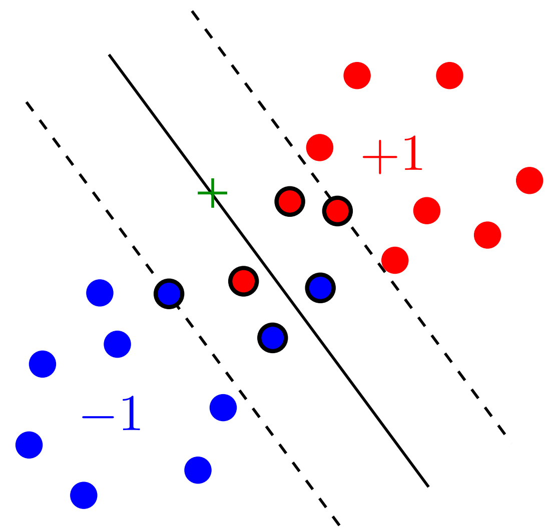
- Předpokládáme, že známe tvar rozumného klasifikátoru. Tj. klasifikátor nemůže být libovolný, ale patří do množiny $\mathcal{H}$. Např. $\mathcal{H}$ je množina všech lineárních klasifikátorů.
- V množině $\mathcal{H}$ hledáme (učíme) klasifikátor, který dobře funguje na trénovacích příkladech, tj. minimalizujeme emprické riziko

$$h^* \in \operatorname*{argmin}_{h \in \mathcal{H}} R_{\text{emp}}(h) \qquad \text{kde} \qquad R_{\text{emp}}(h) = \sum_{i=1}^{m} [\![\ell(y_i, h(x_i))]\!]$$

Linear Support Vector Machines

- Klasifikátor $h \colon \mathbb{R}^n \to \{-1,+1\}$

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



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

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

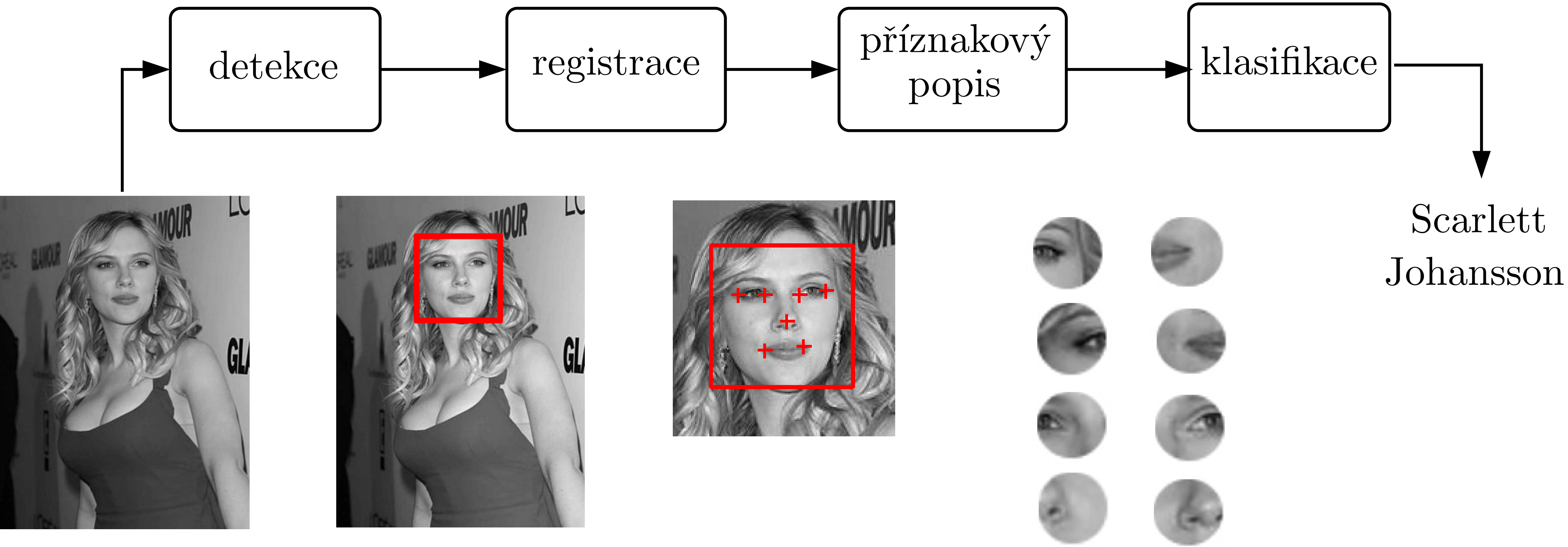
kde $\Omega(\mathbf{w})$ je kvadratický regularizátor a $R(\mathbf{w})$ je konvexní horní mez pro

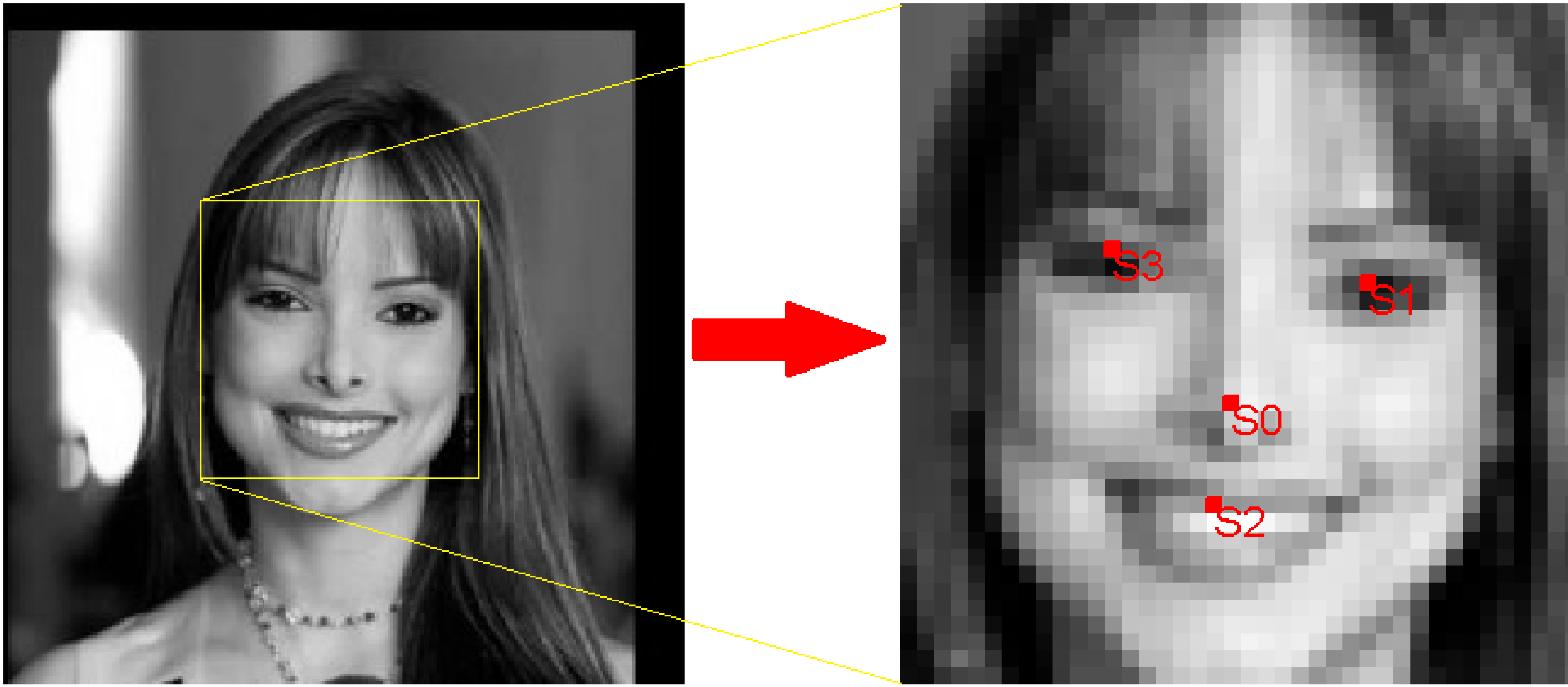
$$\sum_{i=1}^{m} [\![y_i \neq h(\mathbf{x}_i;\mathbf{w})]\!]$$

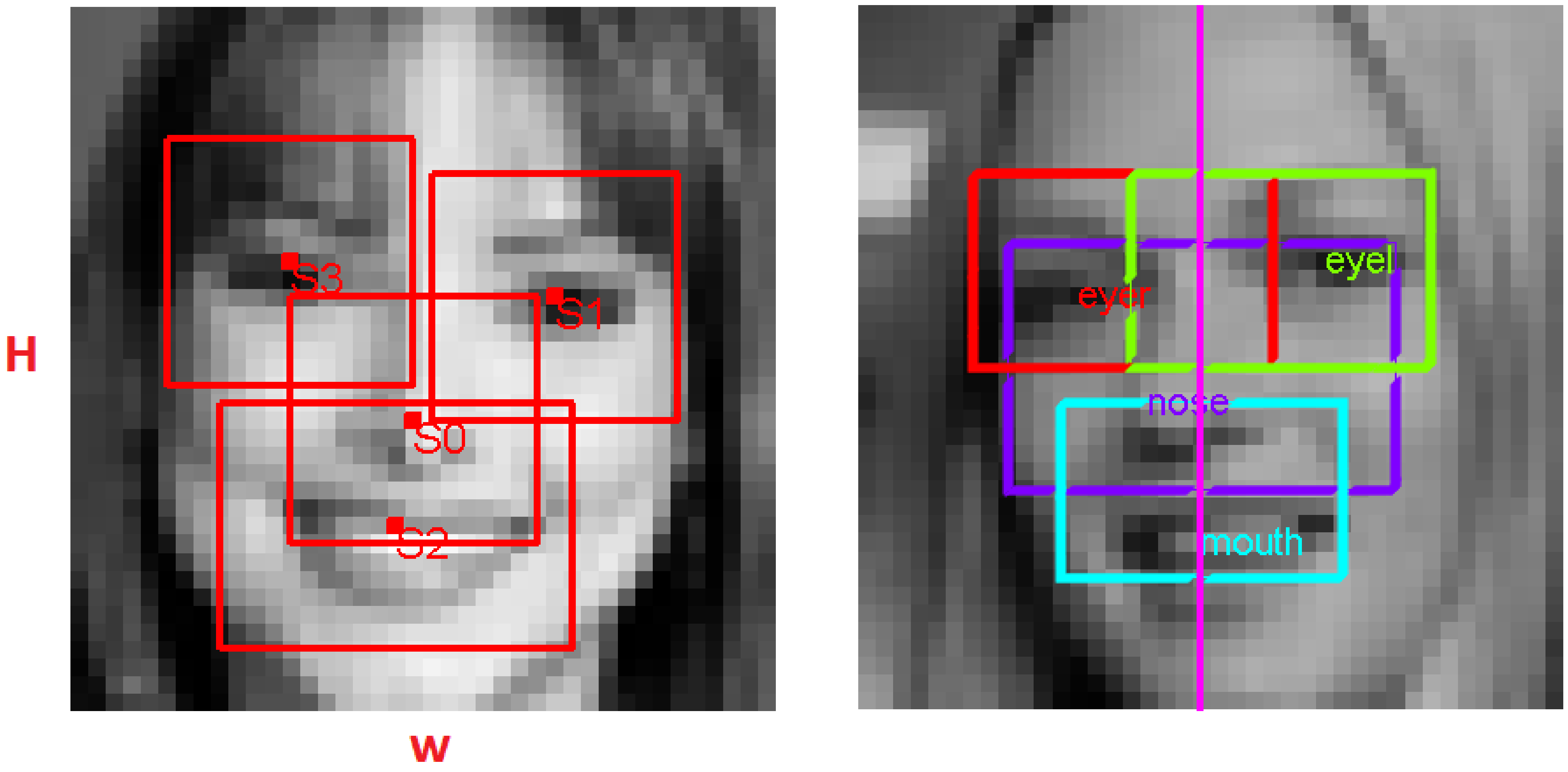
tj. počet chyb, kterých se klasifikátor $h(\mathbf{x};\mathbf{w})$ dopustí na trénovací příkladech.

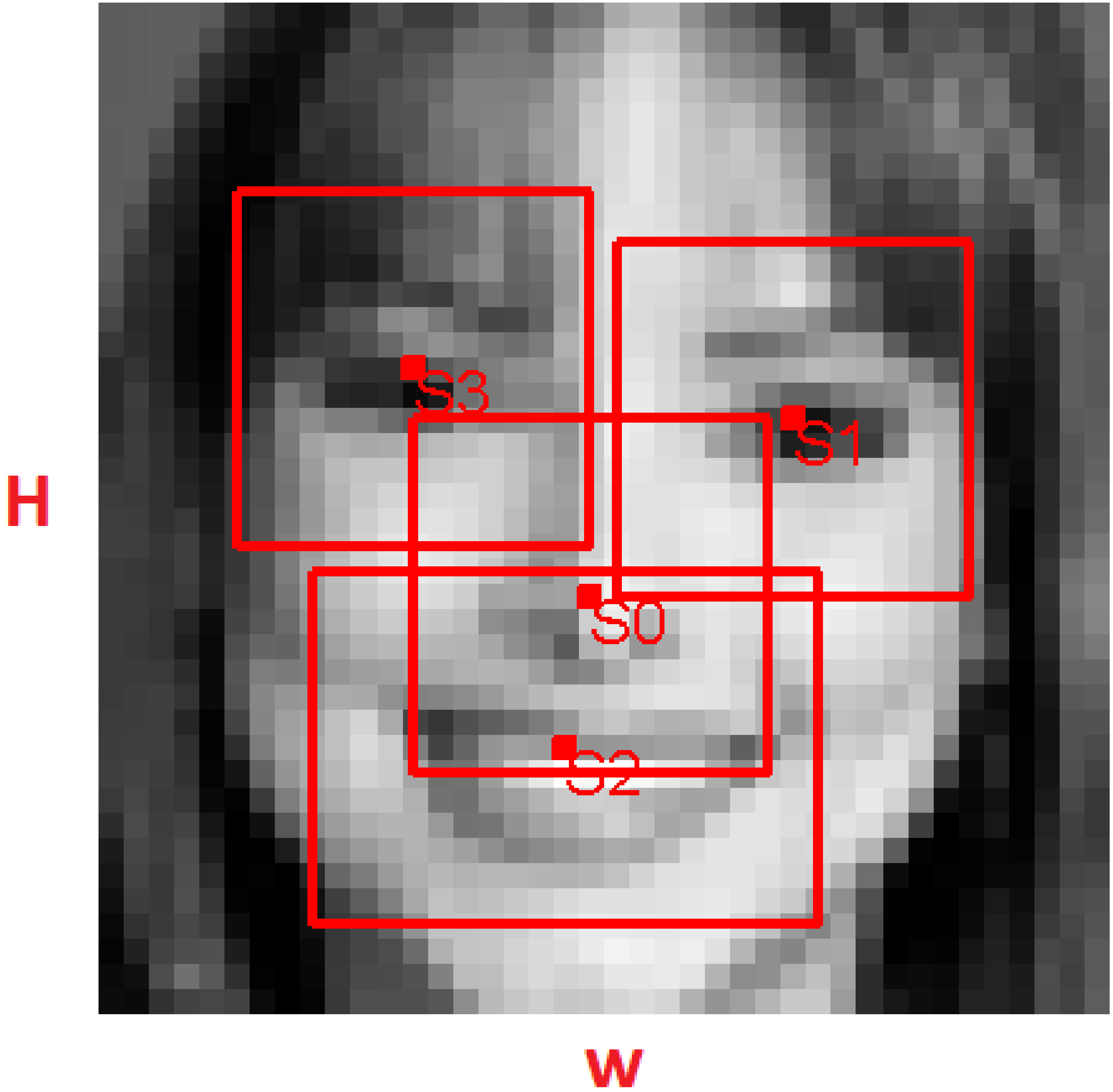
Konec

m p

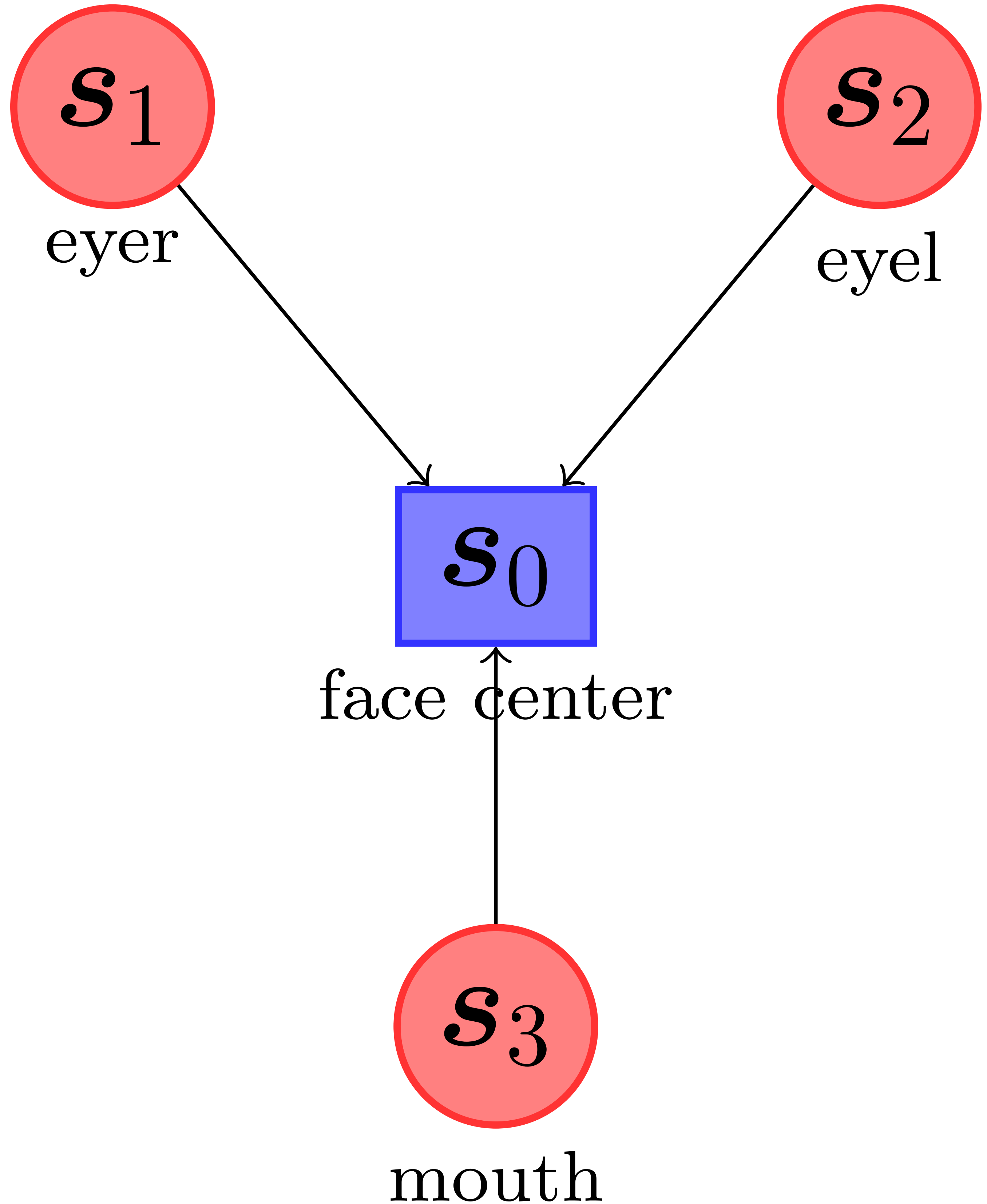

detekce
registrace
příznakový popis
klasifikace
Scarlett Johansson


S3
S1
S0
S2

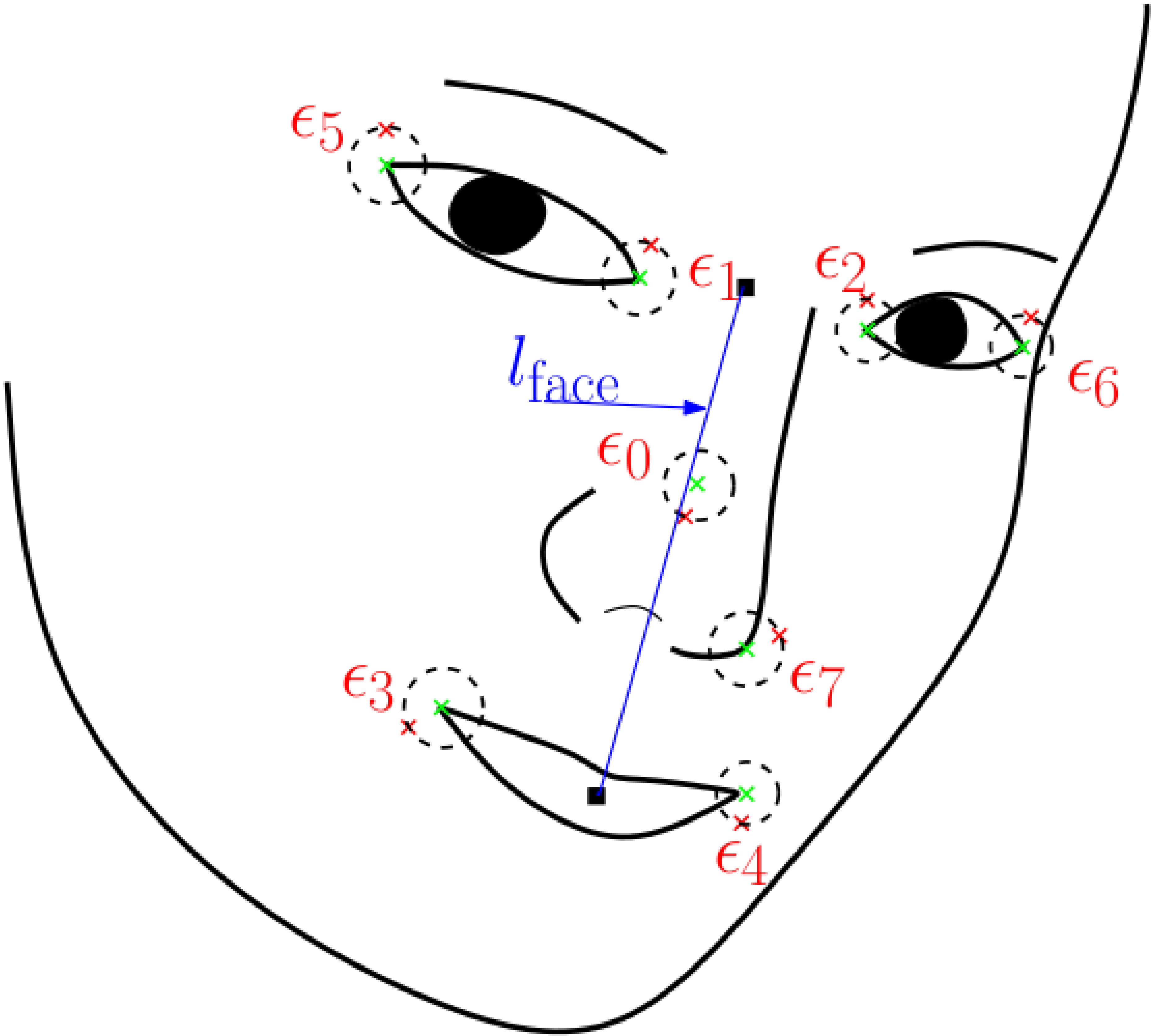

H
S3
S1
S0
S2
W
eyer
eyel
nose
mouth



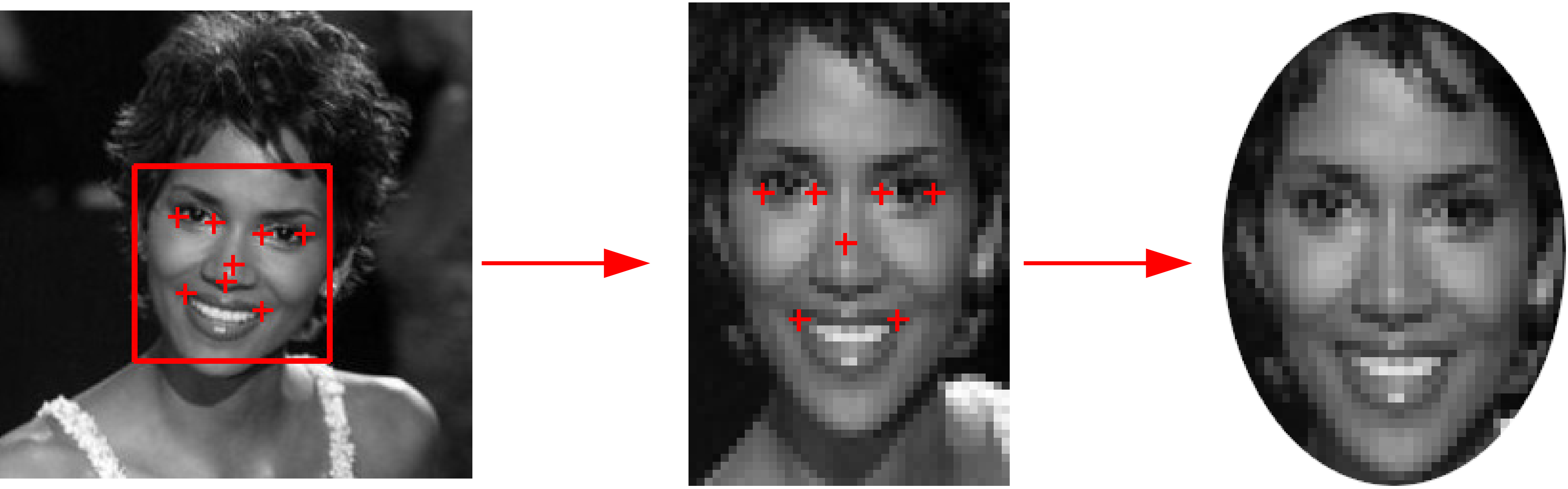
H
S3
S1
S0
S2
W

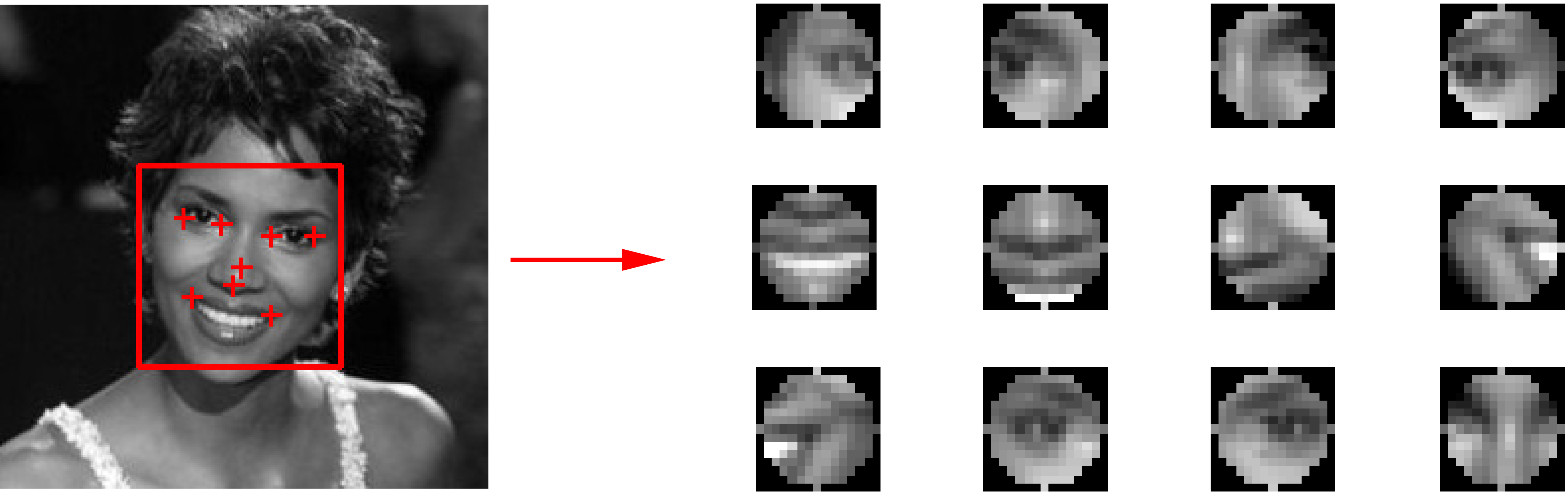


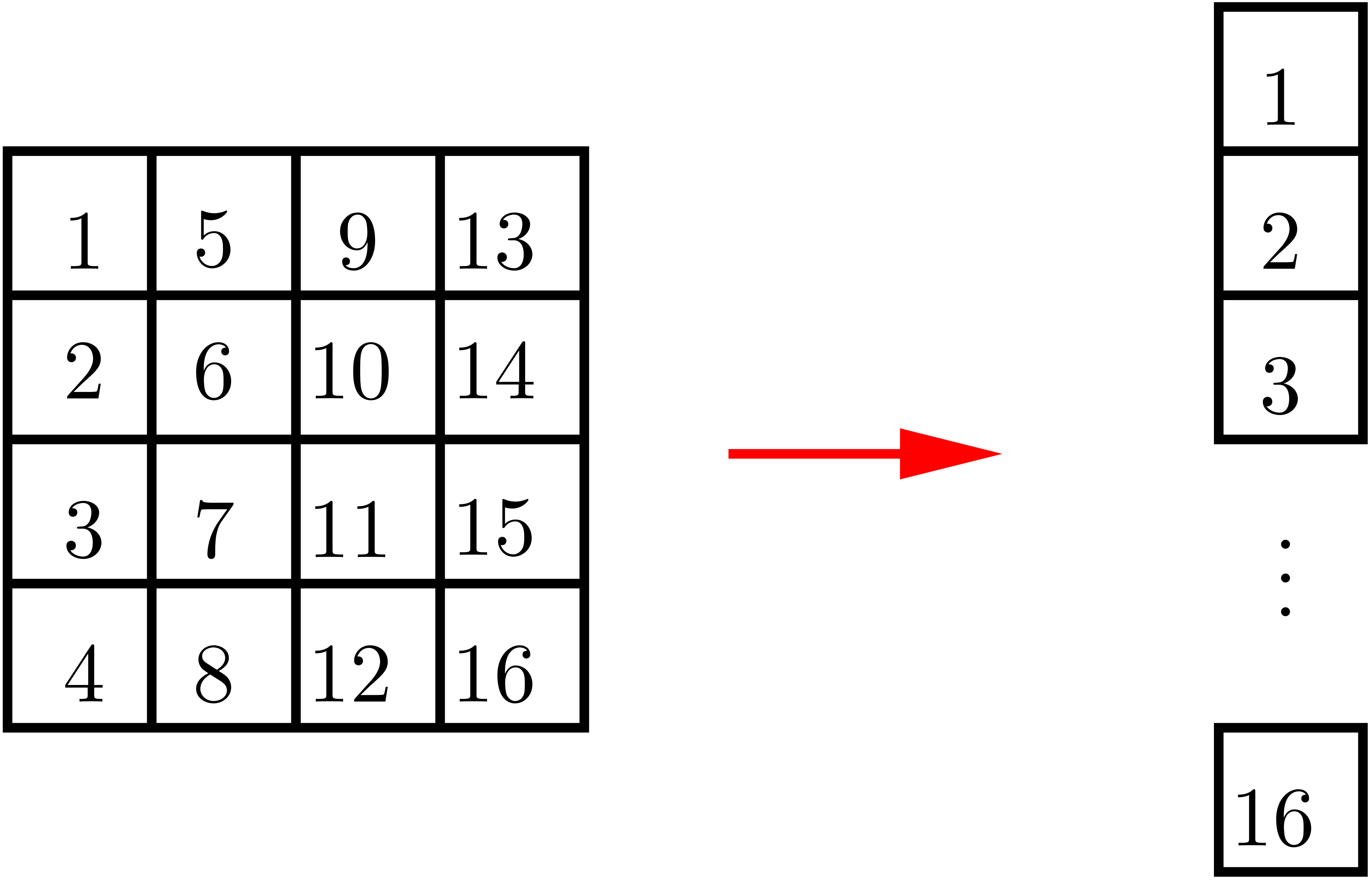
s_1
eyer
s_2
eyel
s_0
face center
s_3
mouth

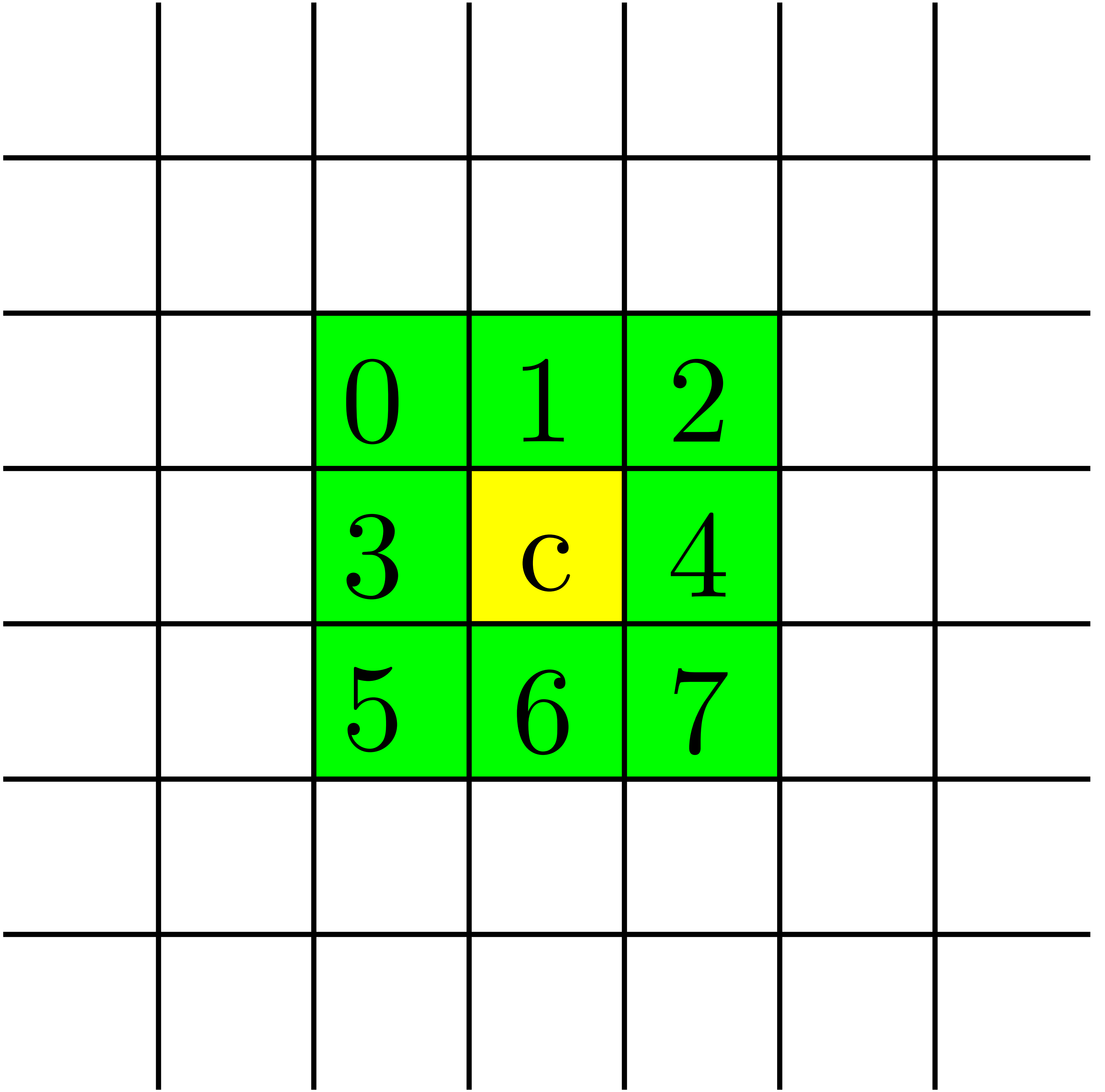


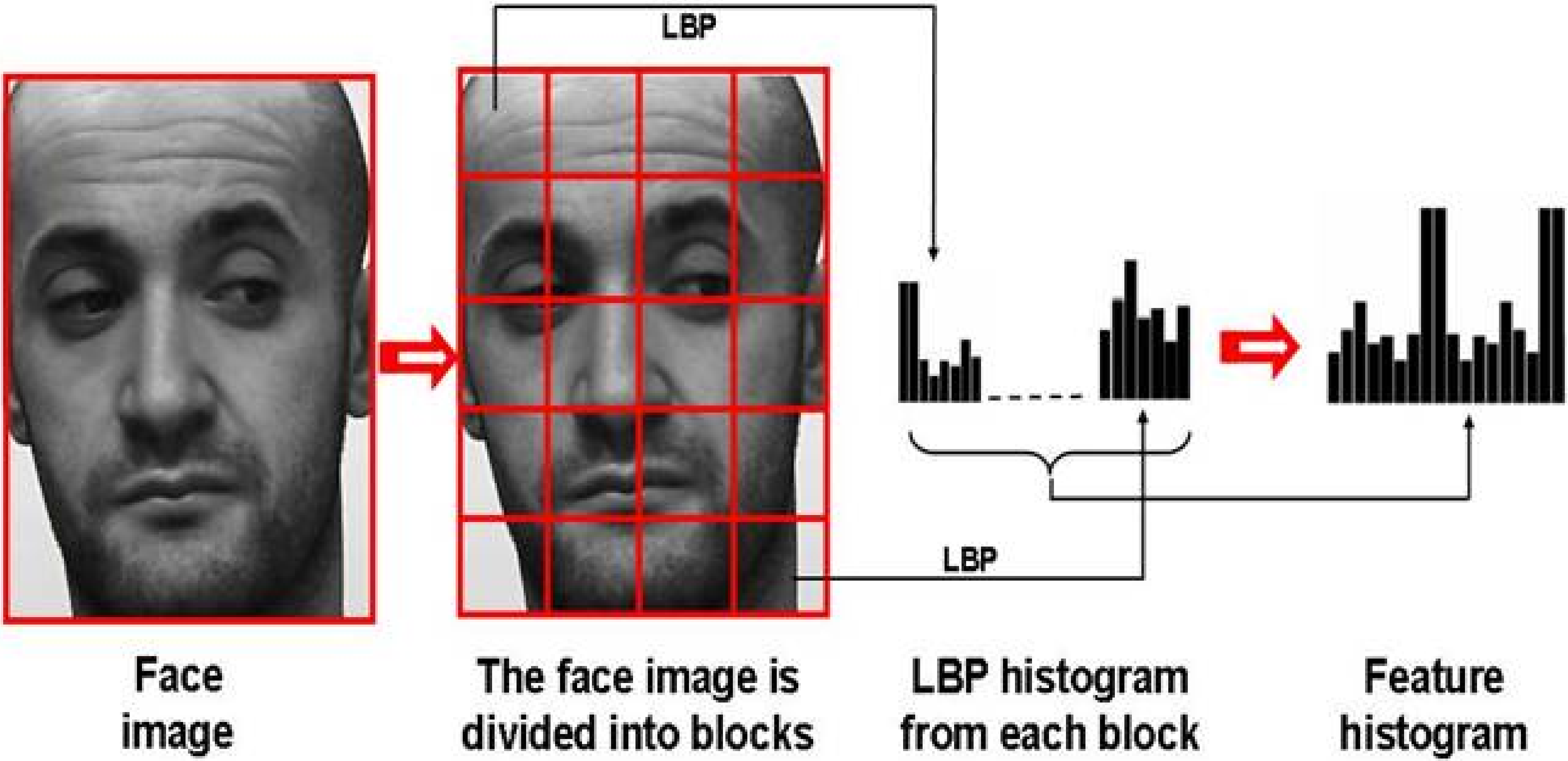
ϵ_5
ϵ_1
ϵ_2
l_{face}
ϵ_6
ϵ_0
ϵ_7
ϵ_3
ϵ_4

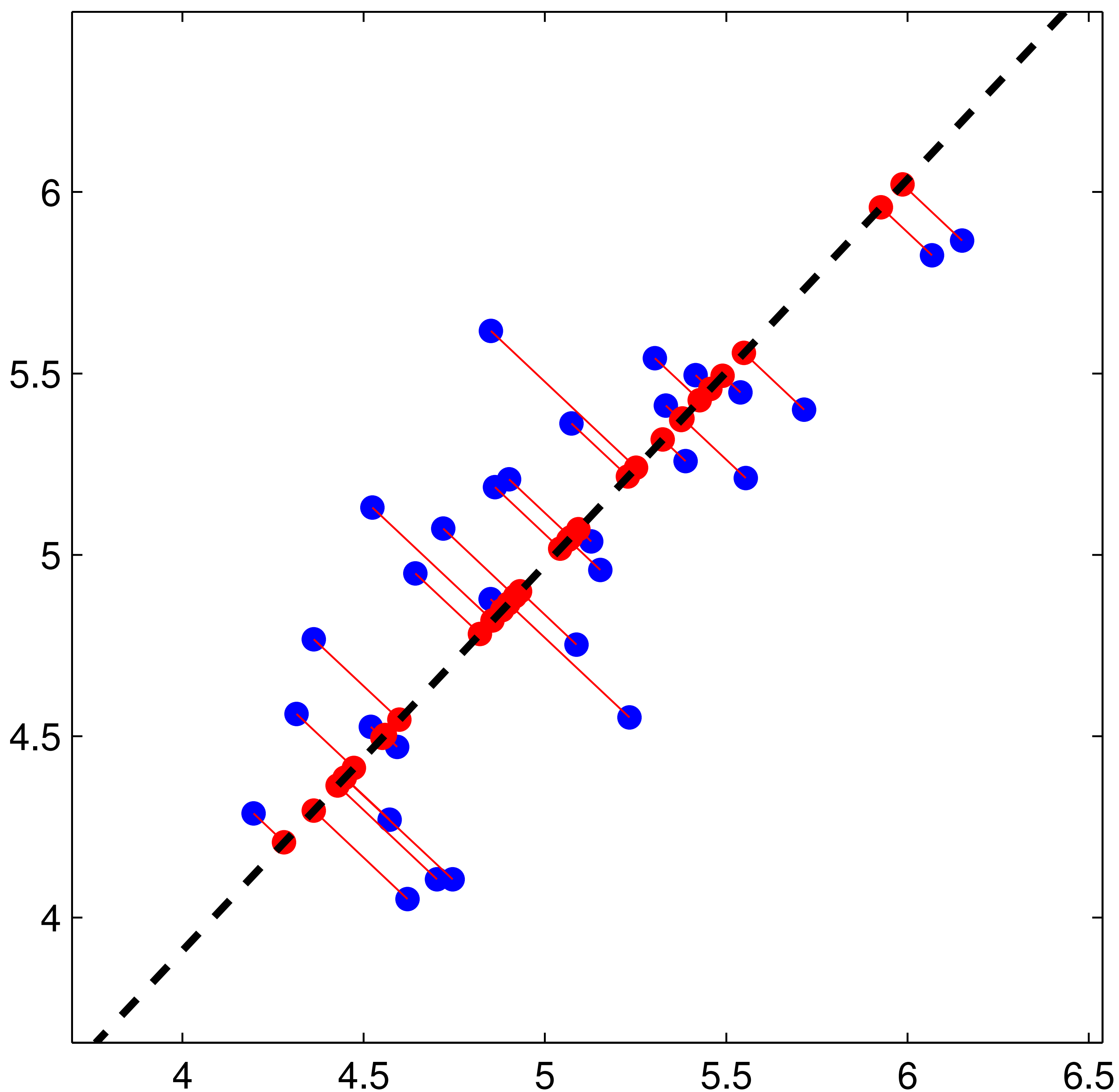





1
5
9
13
2
6
10
14
3
7
11
15
4
8
12
16
1
2
3
16

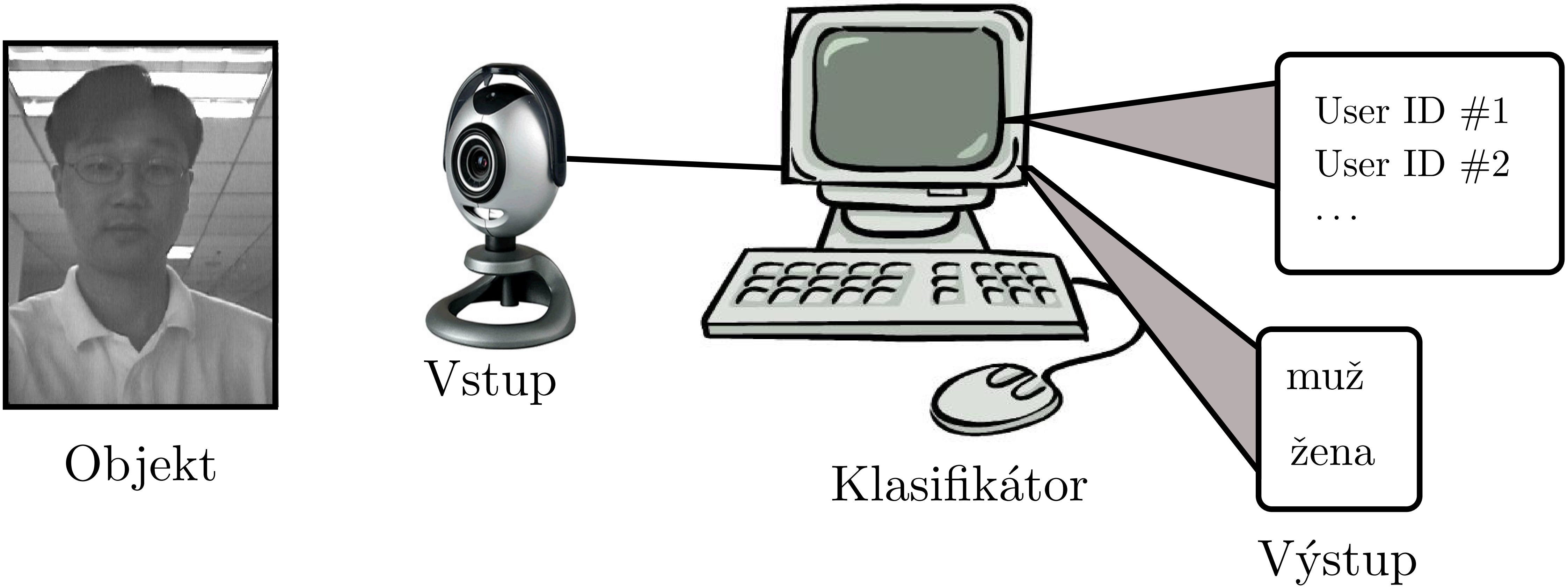

0
1
2
3
c
4
5
6
7

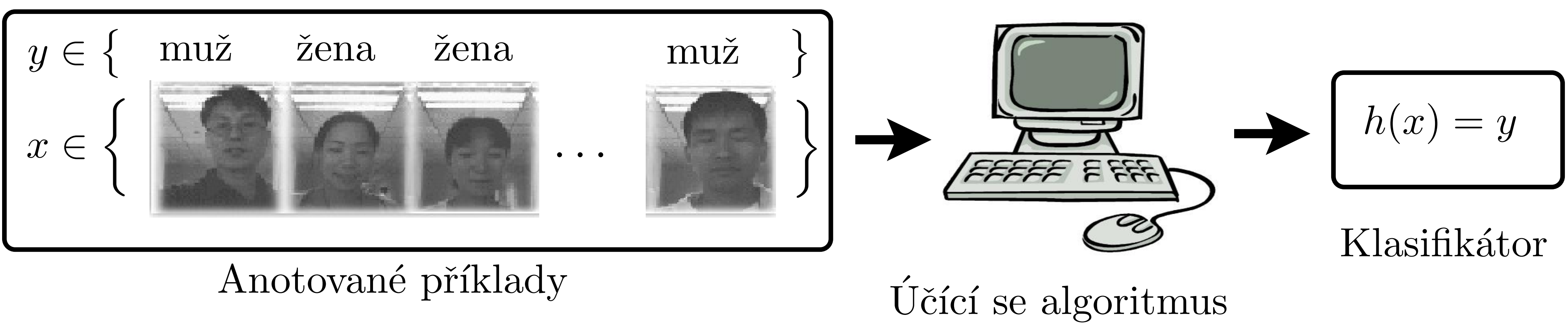

LBP
LBP
Face image
The face image is divided into blocks
LBP histogram from each block
Feature histogram



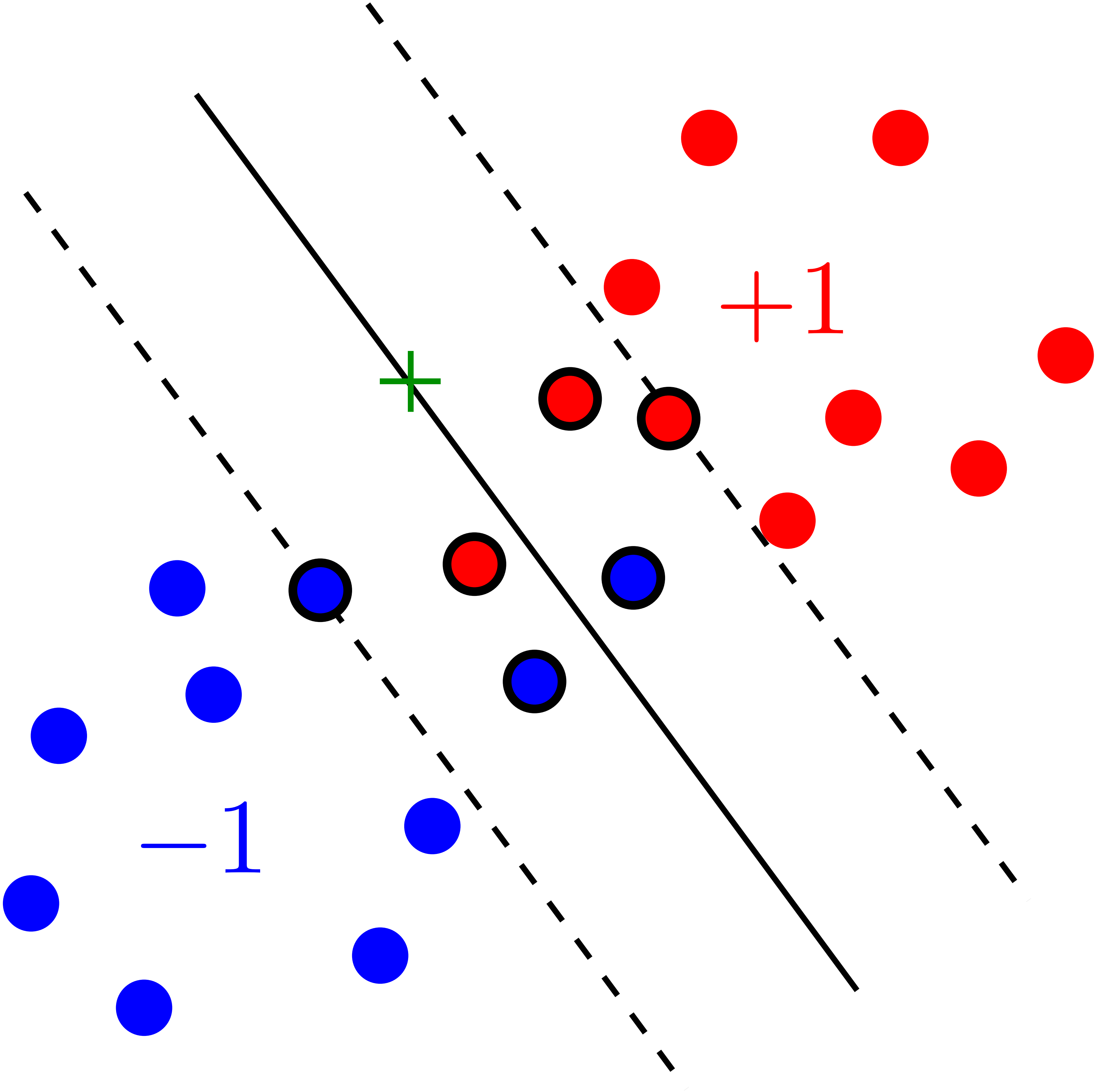
6
5.5
5
4.5
4
4
4.5
5
5.5
6
6.5





User ID #1
User ID #2
. . .
muž
žena
Objekt
Vstup
Klasifikátor
Výstup



$y \in \{$ muž žena žena muž $\}$
$x \in \{ \dots \}$
Anotované příklady
Účící se algoritmus
$h(x) = y$
Klasifikátor



+1
−1