Gesicht- und Handerkennung

Michael Schnell

25.06.2005

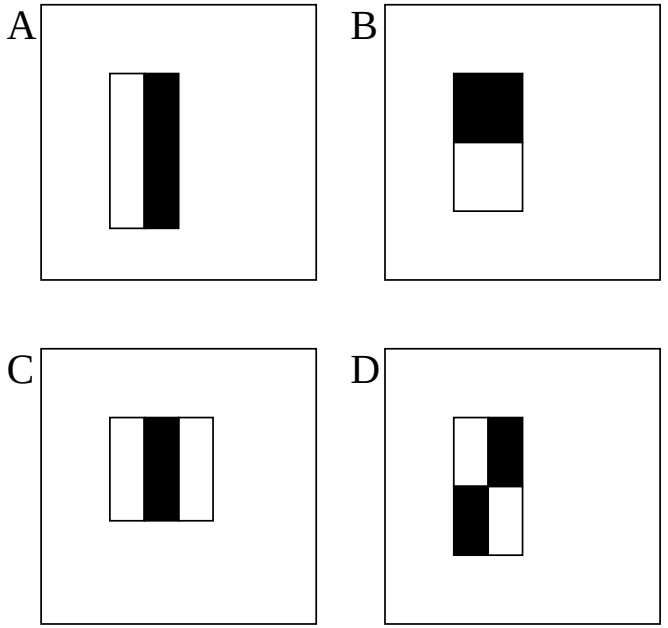
Seminar Gruppenaktionen in dynamischen, unsicheren Umgebungen

Übersicht

- Objekterkennung von Paul Viola und Michael Jones
  - Integral Image
  - AdaBoost: Auswahl der Merkmale, trainieren der Klassifikatoren
  - Kaskade aus Klassifikatoren
  - Gesichtserkennung
- Handerkennung von Mathias Kölsch und Matthew Turk
  - Schätzfunktion: Wie gut lassen sich Objekte klassifizieren?

Einfache Merkmale

- Rechtecke: Summen von Grauwerten
- Merkmale: Differenzen der Rechtecksummen
- bei 24x24 Pixel: 45.396 solcher Merkmale



A,B: Typ 1, C: Typ 2,

D: Typ 3

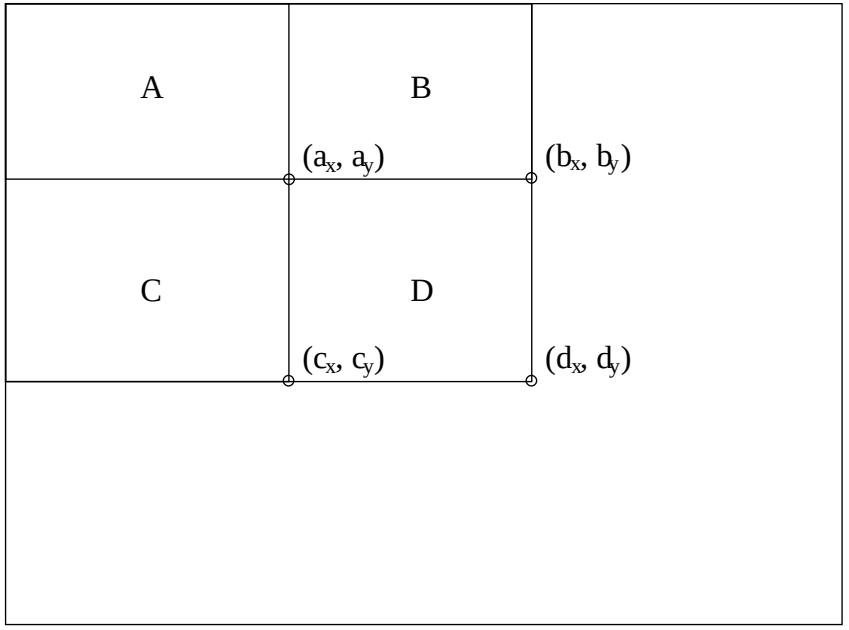
Integral Image

$$ii(x,y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

Linearer Aufwand:

$$s(x,y) = s(x,y-1) + i(x,y)$$
$$ii(x,y) = ii(x-1,y) + s(x,y)$$

$$s(x,-1) := 0 \quad ii(-1,y) = 0.$$

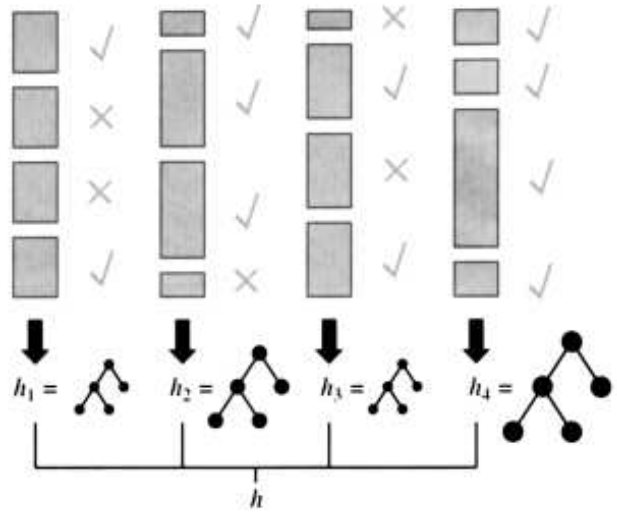


Merkmale effizient berechenbar:
O(1)

$$D = ii(d_x, d_y) - ii(b_x, b_y) - ii(c_x, c_y) + ii(a_x, a_y)$$

AdaBoost, das Original

- Erkennungsleistung von einfachen Lernern verbessern
- Kombiniert mehrere schwache Klassifi katoren zu einem starken
- Jedes Beispiel hat ein Gewicht
- Gewichte werden in mehrere Runden angeglichen, so daß falsch klassifi zierte Beispiele höheres Gewicht bekommen
- Jeder schwache Klassifi kator bekommt ein Gewicht (Abhängig von der Korrektheit)
- Starker (Meta-)Klassifi kator ist eine gewichtete Kombination der schwachen.



Meta-Klassifi kator h setzt sich aus $h_1, \cdots, h_n$ zusammen

AdaBoost von Viola und Jones

- 45.396 Merkmale bei 24x24 Pixel

- Wählt Merkmale und erstellt den Meta-Klassifikator
- schwacher Klassifikator $h_j(x)$ besteht aus
  - einem einzigen Merkmal $f_j(x)$
  - einem Schwellwert θ_j
  - einem Vorzeichen $p_j \in \{1, -1\}$
- $h_j(x) = \begin{cases} 1 & \text{wenn } p_j \cdot f_j(x) < p_j \cdot \theta_j \\ 0 & \text{sonst} \end{cases}$

AdaBoost von Viola und Jones

- Gegeben: Beispiele $(x_1, y_1), \cdots, (x_n, y_n) \quad y_i \in \{0, 1\}$.
- Initialisiere die Gewichte $\omega_{1,i} = \frac{1}{2l}, \frac{1}{2m}$ für $y_i = 1, 0$
- Für t aus $1, \cdots, T$:
  - Normalisiere die Gewichte: $\omega_{t,i} \leftarrow \frac{\omega_{t,i}}{\sum_{j=1}^{n} \omega_{t,j}}$
  - Für jedes Merkmal j
    - Trainiere einen Klassifikator h_j (ein Merkmal)
    - Berechne Fehler $\epsilon_j = \sum_i \omega_i |h_j(x_i) - y_i|$
  - Wähle h_t mit dem kleinstem Fehler ϵ_t
  - Gewichte: $\omega_{t+1,i} = \omega_{t,i} \left(\frac{\epsilon_t}{1-\epsilon_t} \right)^{1-e_i}$ mit $e_i = 0$ für $h_t(x_i) = y_i$, sonst 1

AdaBoost von Viola und Jones

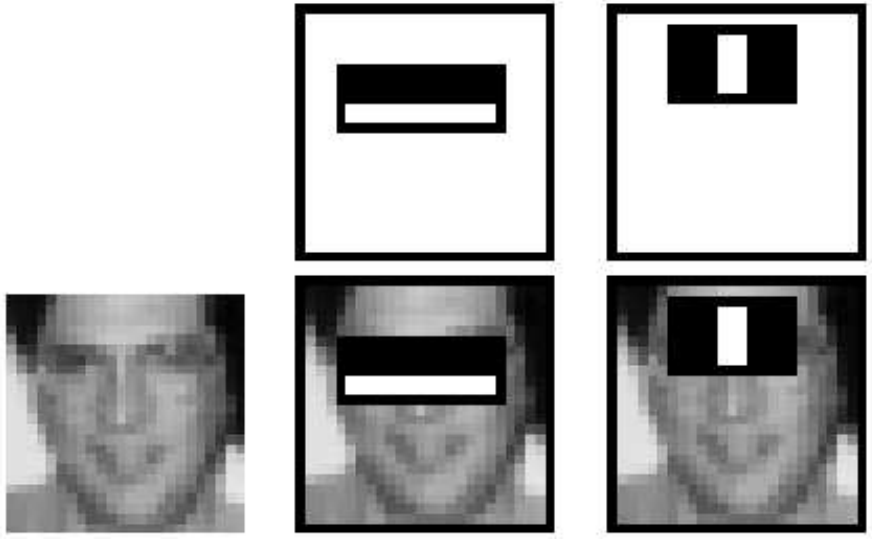
- Der endgültige Klassifikator ist:

$$h(x) = \begin{cases} 1 & \text{wenn } \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=0}^{T} \alpha_t \\ 0 & \text{sonst} \end{cases}$$

Wobei $\alpha_t = \log \frac{1-\epsilon_t}{\epsilon_t}$ die Gewichte der schwachen Klassifikatoren sind.

Ein erster Test von AdaBoost

- Klassifikator aus 200 Merkmalen
- Erkennungsrate: 95%
- *false positiv rate*: $7,1 \cdot 10^{-5}$ oder 1 aus 14084
- 0,7 Sekunden für ein Bild von 384x288 Pixel



Objekte in Bildern erkennen

Bild zur Klassifikation

- an jeder Position
- zu jeder Skalierung

untersuchen.

$\Rightarrow$ Sehr viele Teilbilder

MIT+CMU:

130 Bilder, 507 Gesichter

Schrittgröße 1 Pixel

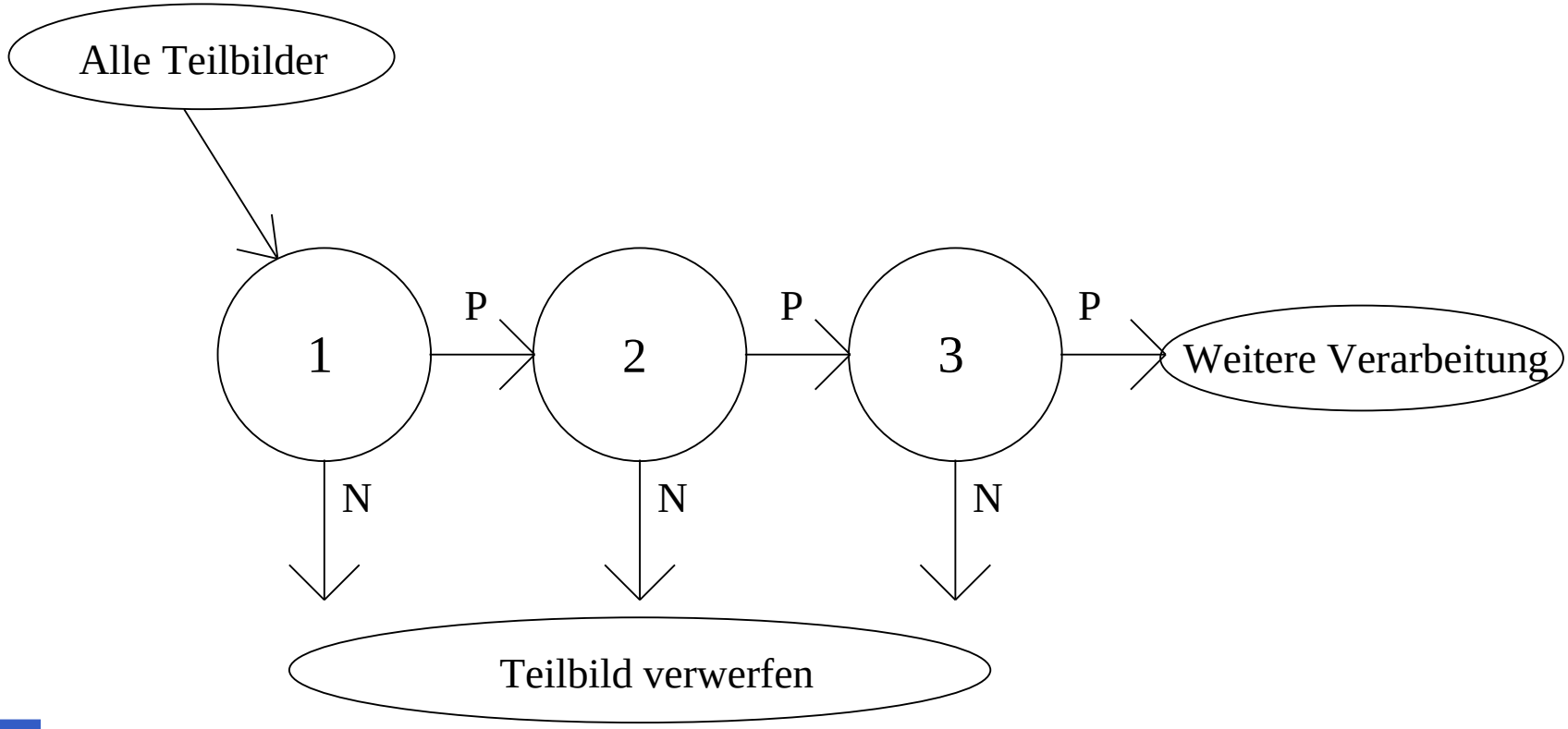
Skalierungsfaktor 1,25

75.081.800 Teilbilder

Kaskade

Idee:

Schnelle einfache Klassifikatoren zur Vorsortierung



Kaskade trainieren

$$D = \prod_{i=0}^{K} d_i \qquad F = \prod_{i=0}^{K} f_i$$

- Bei 10 Ebenen und Zielerkennungsrate von 90%:
  $D = 0.9 \approx 0.99^{10} \Rightarrow$ Jede Ebene mindestens 99%
- Aber: $F = 0.3^{10} \approx 5,9 \cdot 10^{-6}$
  $\Rightarrow$ *false positiv rate* von bis zu 30% auf jeder Ebene
- extrem hohe Erkennungsraten auf den Ebenen nötig.
  $\Rightarrow$ Schwellwert von AdaBoost anpassen

Kaskade trainieren

- Gegeben: d, f, F_{target}
- P = {positive Beispiele}, N = {negative Beispiele}
- $F_0 = 1.0;\ D_0 = 1.0;\ \ i = 0$
- while $F_i > F_{target}$
  - $i \leftarrow i + 1;\ \ F_i = F_{i-1};\ \ n_i = 0$; //Anzahl der Merkmale
  - while $F_i > f \times F_{i-1}$
    - $n_i \leftarrow n_i + 1$ (Nehme ein Merkmal mehr)
    - Klassifi kator mit n_i Merkmalen durch AdaBoost mit N und P trainieren
    - F_i und D_i der momentanen Kaskade bestimmen
    - Schwellwert des aktuellen Klassifi kators heruntersetzen, bis die aktuelle Kaskade eine Erkennungsrate von mindestens $d \times D_{i-1}$ erreicht. (Beeinflusst auch F_i.)
  - $N \leftarrow \emptyset$
  - Falls $F_i > F_{target}$ so mache alle Bilder in N, die keine Gesichter enthalten, aber von der aktuellen Kaskade als positiv erkannt werden.

Beispiel: Gesichtserkennung

- Positive: 4916 Bilder aus dem Internet
- von Hand markiert und auf 24x24 Pixel skaliert.
- 1. Ebene:
  2 Merkmale, $\approx$ 100% erkannt, 60% verworfen
- 2. Ebene:
  5 Merkmale, $\approx$ 100% erkannt, 80% verworfen
- 3 Ebenen mit 20 Merkmalen
- 2 Ebenen mit 50 Merkmalen
- 5 Ebenen mit 100 Merkmalen
- 20 Ebenen mit 200 Merkmalen

Beispiel: Gesichtserkennung

- Trainieren dauerte Wochen (466 MHz AlphaStation XP900)
- Klassifikator schaffte 384x288 Pixelbild in 0,067 Sekunden (auf Pentium III 700MHz)
- MIT+CMU: Teilbild wurde im Schnitt auf 8 Merkmale untersucht.

- Klassifikator unempfindlich auf kleine Änderungen in Skalierung und Position
  $\Rightarrow$ mehrere Treffer für das gleiche Gesicht zu einem Treffer vereinen.

Handerkennung

- Forschungsgebiet: Gestensteuerung, basiert auf Videobildern.
- Welche Handstellungen sind am besten geeignet?
- Schätzfunktion

Schätzfunktion

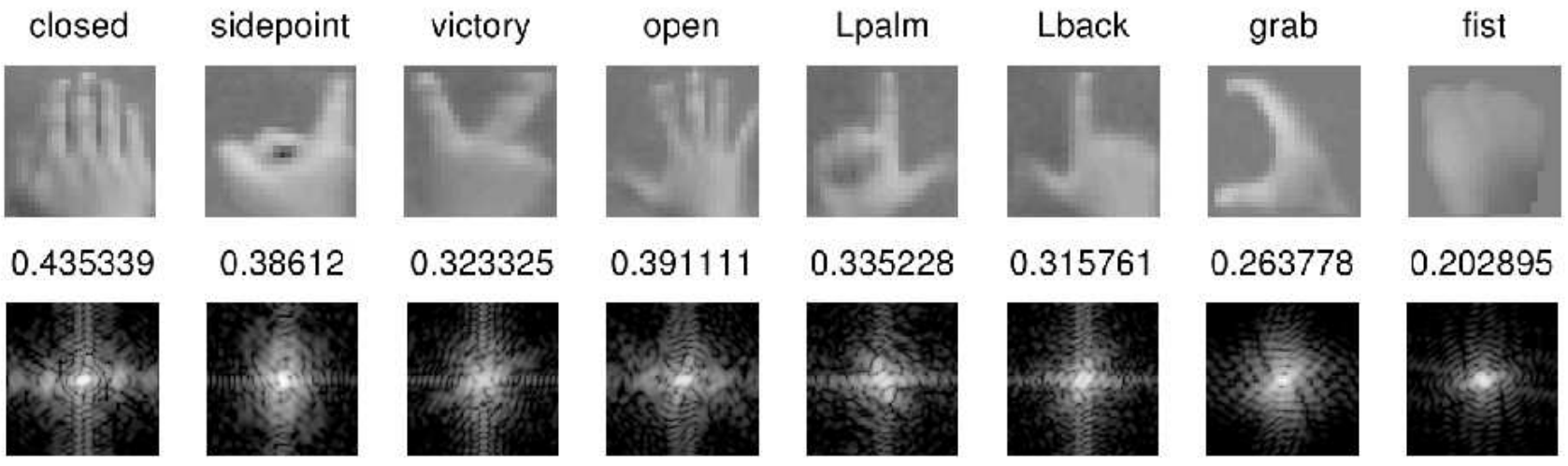
$$F(u,v) = \frac{1}{25 \cdot 25} \sum_{m=0}^{24} \sum_{n=0}^{24} I(m,n) e^{-i2\pi(\frac{mu}{25} + \frac{nv}{25})}$$

$$P(u,v) = \frac{1}{25 \cdot 25} \sum_{m=0}^{24} \sum_{n=0}^{24} \frac{1}{2} e^{-i2\pi(\frac{mu}{25} + \frac{nv}{25})}$$

$$D(u,v) = \log |F(u,v) - P(u,v)|$$

$$s = e^{\frac{1}{k} \cdot \sum_{u,v} D(u,v)}$$

Handstellungen

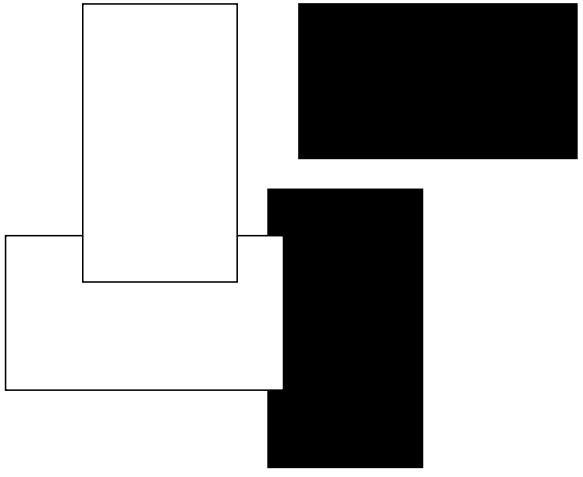


Nicht kaskadierter Klassifikator

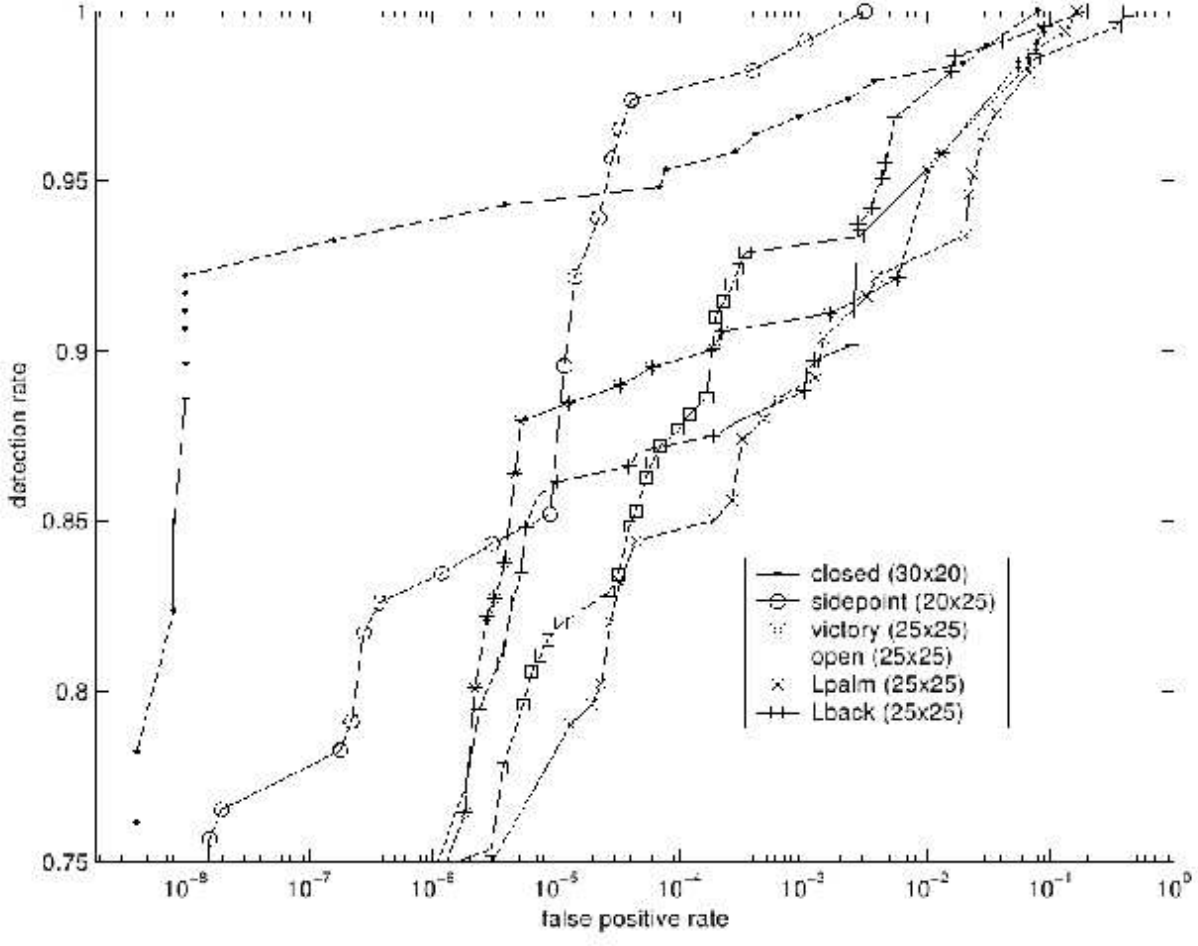
- 2300 Bilder von Händen
- unterschiedliche Hintergründe, Personen, Beleuchtung
- auf 25x25 Pixel skaliert
- 23.000 Negative.

- Schätzfunktion nur Teilweise bestätigt.

Neues Merkmal



Kaskadierter Klassifikator



Zusammenfassend

- Methode von Viola und Jones sehr schnell
- Integral Image $\rightarrow$ Klassifikator statt Bild skalieren
- Variante von AdaBoost, wählt Merkmale und trainiert den Klassifikator
- Kaskade: Schnelle Klassifikatoren sortieren viele Nagative aus
- Weiteres Merkmal von Kölsch und Turk hinzugenommen
- Schätzfunktion erspart viel Rechenzeit bei der Auswahl von Handstellungen

Literatur

[1] Paul Viola und Michael Jones (2001). Robust Real-time Object Detection. In *Second International Workshop on Statistical and Computational Theories of Vision - Modeling, Lerning, Computing and Sampling*, Cambridge.

[2] Mathias Kölsch und Matthew Turk, Robust Hand Detection. University of California, Santa Barbara.