Computer Vision: Bag of Visual Words

D. Schlesinger – TUD/INF/KI/IS

Punktmerkmale → Bildmerkmale

Grundlage:

Input: Bild

→ Interessante Punkte (Harris, MSER, Laplace ...)

→ Deskriptoren, ein pro Punkt (Haar, SIFT, Bildfragmente ...)

Output: Ein „Tupel“ von Vektoren pro Bild.

Was ist im Bild zu sehen (Erkennung, Klassifikation)?

Die Deskriptoren an sich sind nicht aussagekräftig genug – es sind „nur Zahlen“.

Die Menge aller Deskriptoren wird geclustert,
der Deskriptor (numerisch) wird wehrend der Erkennung durch Clusternummer ersetzt
– Klassifikation (Interpretation des Wertes) – eine „etwas semantischere“ Bedeutung

⇒ **Output**: Eine Liste der Besondere Punkte mit den entsprechenden Clusternummern.

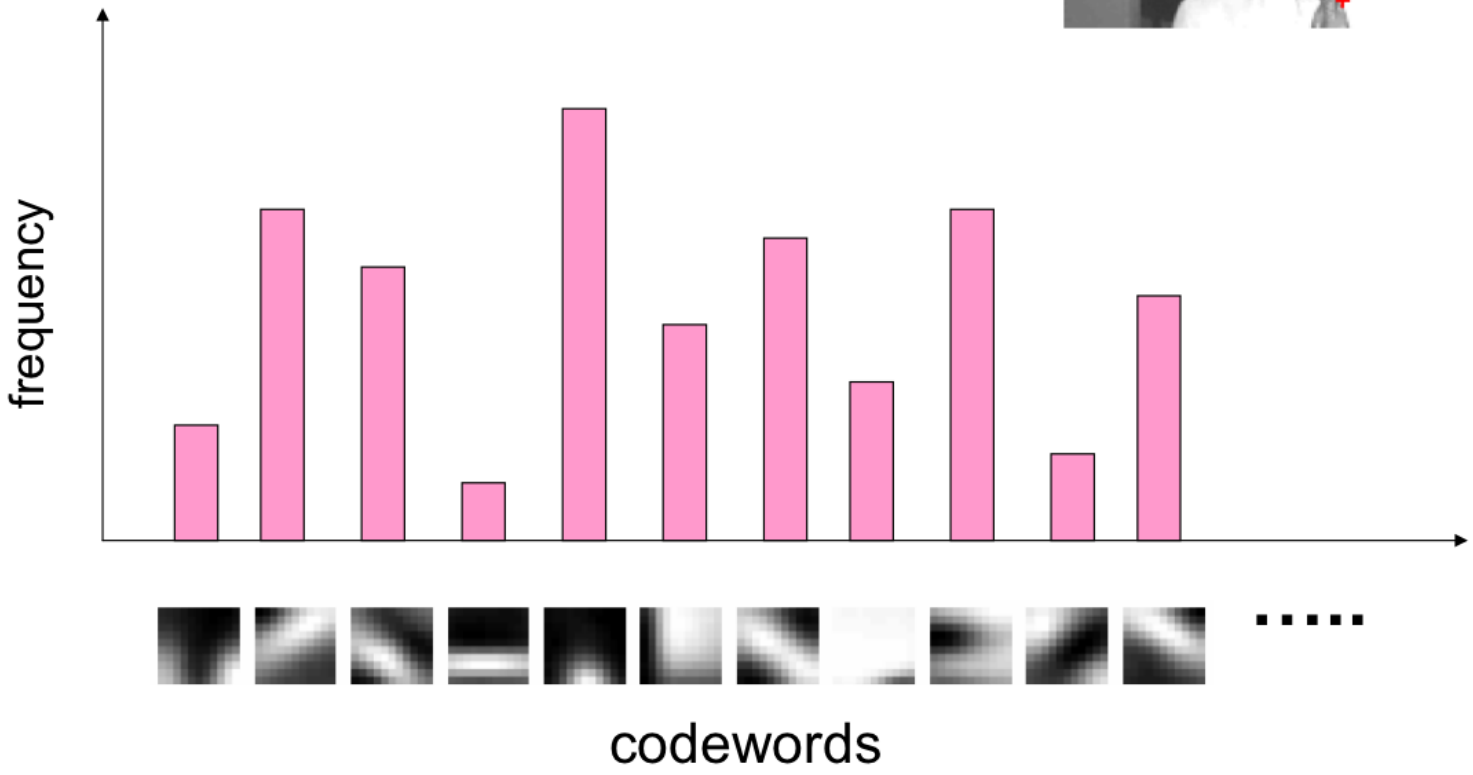
Was ist im Bild zu sehen?

Die Idee:
die Häufigkeiten des Vorkommens der Clusternummer sind für die Klassifikation relevant
Beispiel: „sehr viele vertikale Kanten im Bild“ ⇒ (höchstwahrscheinlich) „Zebra“

Bildmerkmal ist das Histogramm des Vorkommens der Clusternummer.

Image representation – normalized histogram

- detect interest point features
- find closest visual word to region around detected points
- record number of occurrences, but not position



9

Erkennung

Was ist im Bild zu sehen?

$$\text{Bild} \quad \rightarrow \quad \text{Merkmal} \quad \stackrel{?}{\rightarrow} \quad \text{Klasse}$$

Ein **Klassifikator** ist eine Abbildung $e : \mathbb{R}^n \rightarrow K$,
die jedem **Muster** $x \in \mathbb{R}^n$ eine **Klasse** $k \in K$ zuordnet.

Lernaufgabe:
Gegeben sei eine Familie $\mathcal{E}$ der Klassifikatoren
Gegeben sei eine **annotierte Lernstichprobe** $\Big((x_1, k_1), (x_2, k_2) \ldots (x_L, k_L)\Big)$
Man suche nach dem Klassifikator $e \in \mathcal{E}$ so, dass $e(x_l) = k_l$ für alle l gilt
(wenn es mehrere gibt, finde den „besten“)

Mashine Learning, Mustererkennung

Empfehlenswert:

`http://people.csail.mit.edu/torralba/shortCourseRLOC/index.html`