Computer Vision Systems Programming VO

Object Category Recognition

Christopher Pramerdorfer

Computer Vision Lab, Vienna University of Technology

Topics

Scene classification using the bag of words model

Fast face detection using boosted Haar features

Convolutional neural networks for large-scale problems

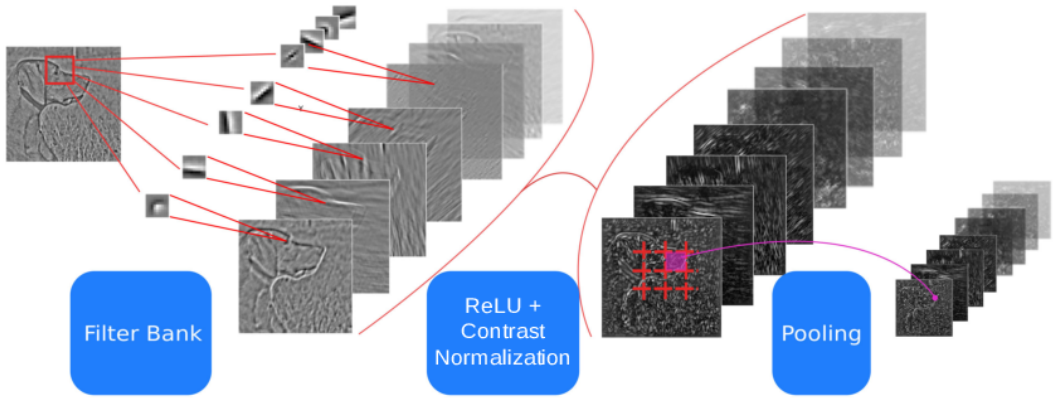


Image adaoted from Kavukcuoglu 2011

Scene Classification

We want to distinguish between c scene categories

- So $w \in \{0, \ldots, c-1\}$ (classification problem)

Street Scenes

Sea Scenes

Image adapted from Prince 2012

Scene Classification
Bag of Visual Words

We represent an image as a collection of visual words

- Images can be compared based on visual word distribution

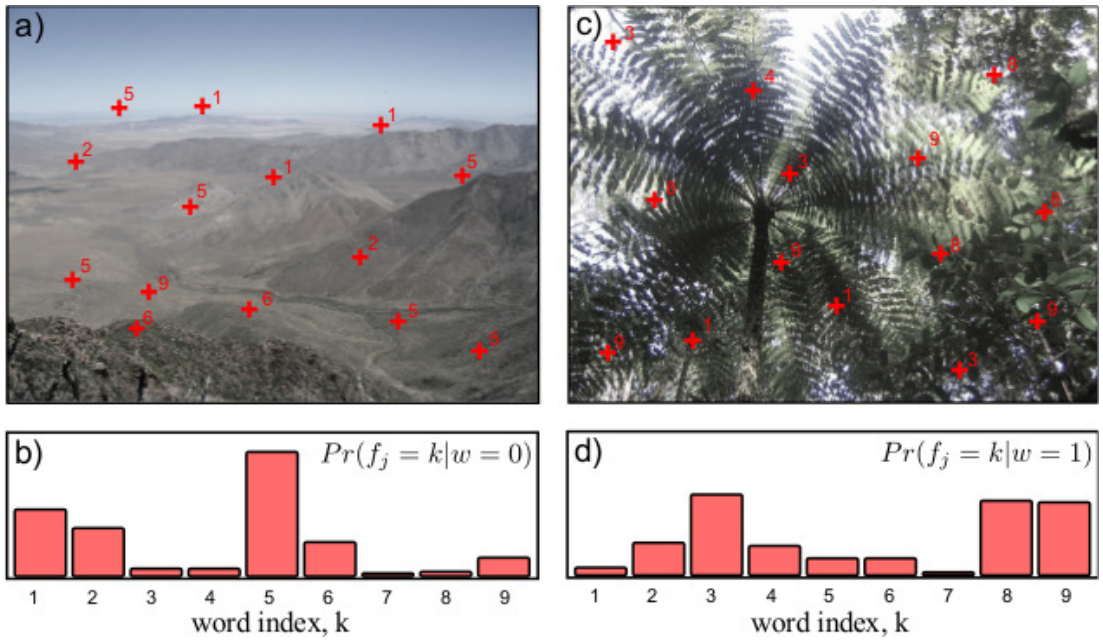


Image from Prince 2012

Scene Classification
Bag of Visual Words

Visual words are learned from an image collection

- Compute (SIFT) keypoints and descriptors for all images
- Cluster descriptors into k clusters using k-means
- k cluster means represent visual words

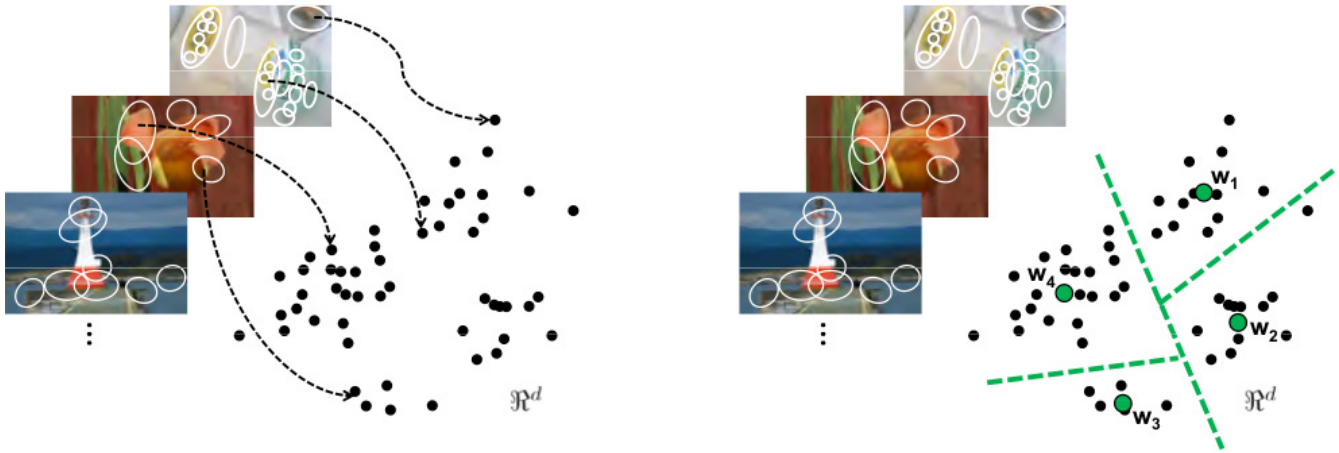


Image from Grauman and Leibe 2011

Scene Classification
Bag of Visual Words

Visual word distribution $\mathbf{x} \in \mathbb{N}^k$ of image obtained by

- Computing keypoints and descriptors
- Assigning each feature to closest visual word
- Summing up the assignment counts for each visual word

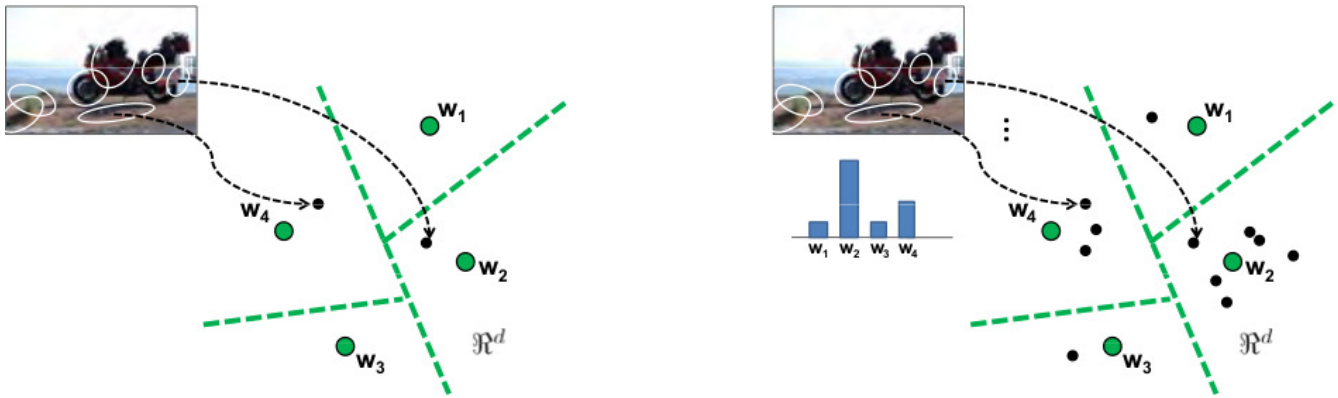


Image from Grauman and Leibe 2011

This image representation is called bag of (visual) words

Now that we have $\mathbf{x}$ we can select and learn a suitable model

- SVMs are often used in the literature
- For a probabilistic alternative see Prince 2012

Scene Classification
Bag of Visual Words – Remarks

Many improvements to this model exist

- Better clustering schemes
- Fuzzy assignment to visual words
- Spatial information (constellation model)

Popular and can work well, but no longer state of the art

Face Detection

Image from olympus-europa.com

Face Detection

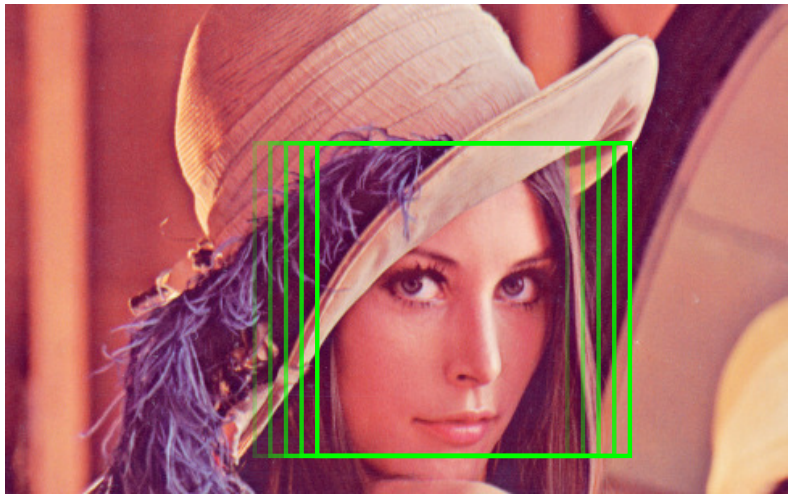
Many applications

- Smart cameras (autofocus on faces)
- Security (preprocessing step to face recognition)
- Augmented reality

Face Detection

We don't know where the faces are so we

- Slide a fixed-size window over the image
- Compute $\Pr(w|\mathbf{x})$ for each window ($w = 1$ if face, 0 if not)



Face Detection

Many windows, so computing $\mathbf{x}$ and $\Pr(w|\mathbf{x})$ must be fast

We focus on the popular method from Viola and Jones 2001

Bibliography I

Grauman, Kristen and Bastian Leibe (2011). *Visual object recognition*. Morgan & Claypool.

Kavukcuoglu, Koray (2011). *Learning feature hierarchies for object recognition*. PhD thesis.

Prince, S.J.D. (2012). *Computer Vision: Models Learning and Inference*. Cambridge University Press.

Viola, Paul and Michael Jones (2001). *Rapid object detection using a boosted cascade of simple features*.