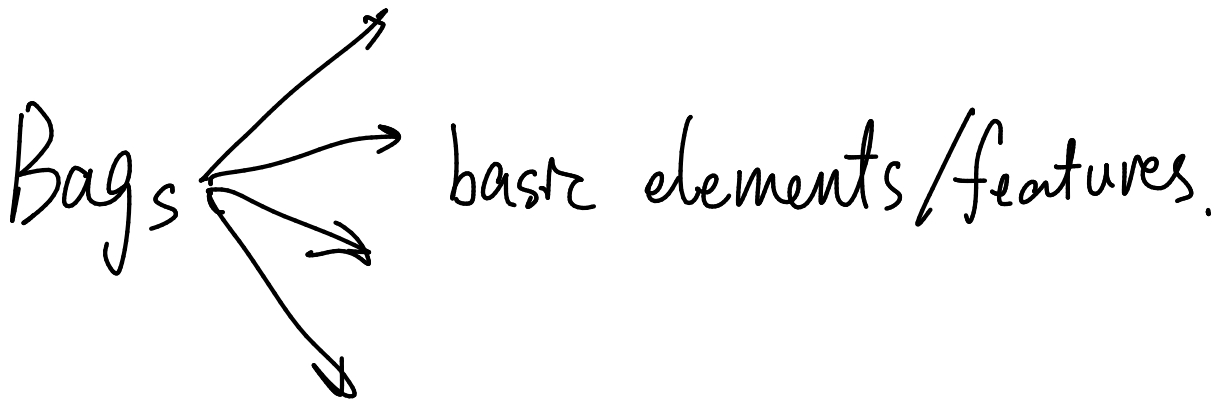
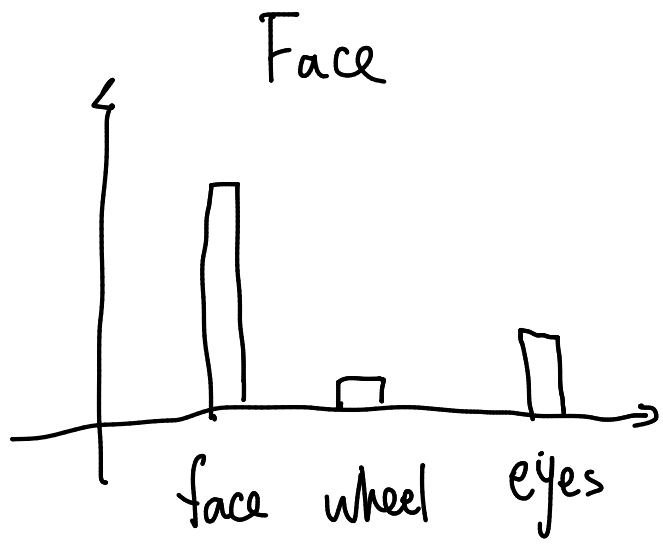
Lect 11 History of Image Categorization

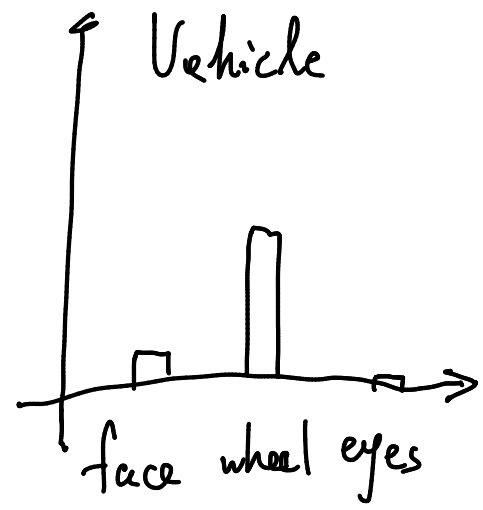
1. 2006. Bag of words model.



- Independent features
- Histogram representation

Say





① feature detection

Use all methods to detect features.

Image → [features set] (N # of features; length by feature description)

② Codeword Dictionary formation.

features → face features
features → vehicle features
features → other features

by clustering methods (k-mean, community detection)

↓

assign each feature to nearest/most likely category

use 1- hot key. / soft quantization.

↓

get a score for each category by summation.

Now.

Image → score vector [k×1] (k: # of categories in bag.)

Working Pipeline

Size of dictionary

Image → SIFT → descriptor → k-Dim score → SVM

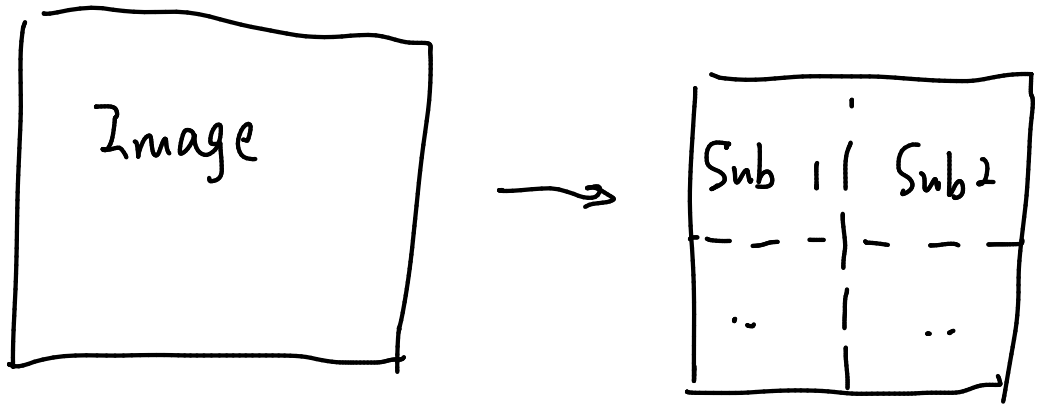
Problem:

No spacial Info.

② Spacial Pyramid Matching (SPM) 2006

– still Bag of Word method.

but to sub-images.



each sub-image has a score vector

③ Histogram of Orientation of Gradient

2. Dataset

→ Caltech 101

101 categories. roughly 50 images of each.

→ PASCAL / LABEL ME / IMAGNET ·····

3 Competition.

Give out the label for each image

Turns out CNN is better than any of previous methods,

Hythothesis of CNN & Traditional Method.

① Edge detectors. blobs ··· can be viewed as Convolution

② Pyramid method, histgram ··· can be viewed as pooling.

CNN can simulate previous methods.

But, filter taps are learnt from a huge dataset. rather than given by human.

So, for cases where images set is small, CNN may not work pretty well, because too many params.

4. Transfer Learning.

→ Train the overall network on some very large dataset

→ Fix all weights, but the last layer, then train it.

Or if image set is a little larger, can free some higher layers, say last 2-layers.

Intuition.

a first several layer may be universal, say the first layer is doing edge detector, blob detector. So, even our ta et dataset is not large, we can borrow the setting from other model