



HELSINKI UNIVERSITY OF TECHNOLOGY  
Department of Information and Computer Science  
Adaptive Informatics Research Centre

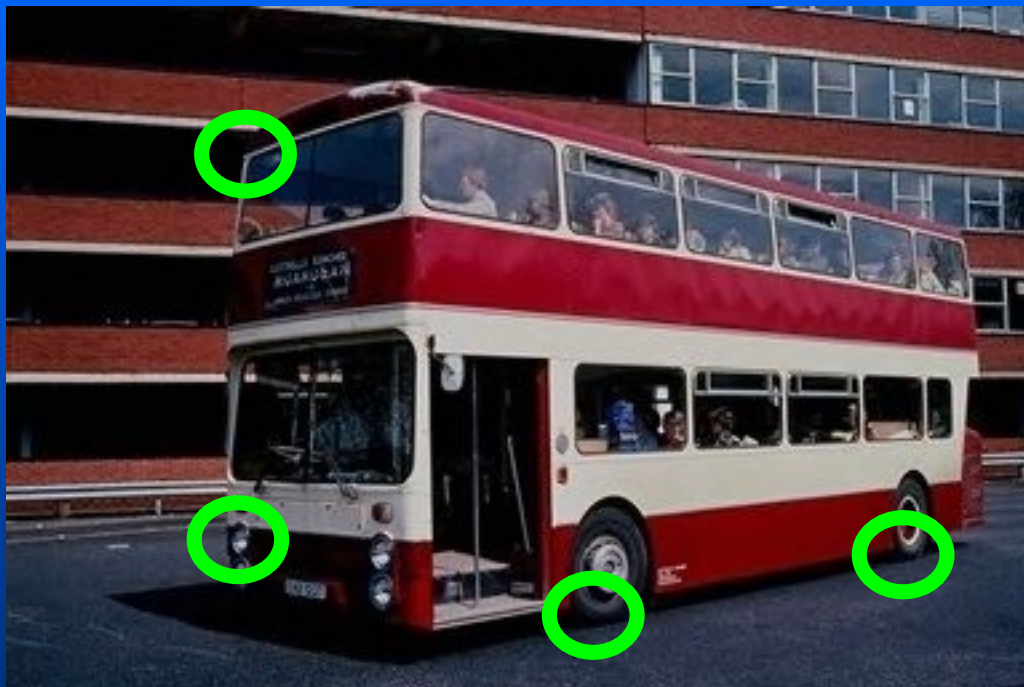
# Experiments on Selection of Codebooks for Local Image Feature Histograms

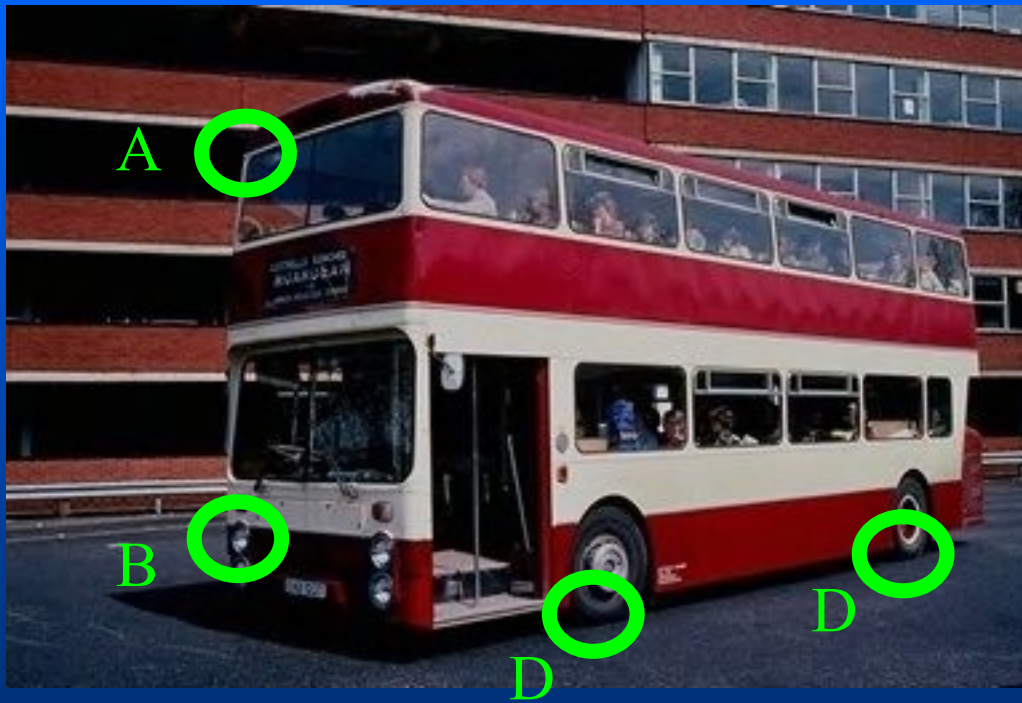
Ville Viitaniemi, 2008-9-11  
paper co-authored with  
Jorma Laaksonen

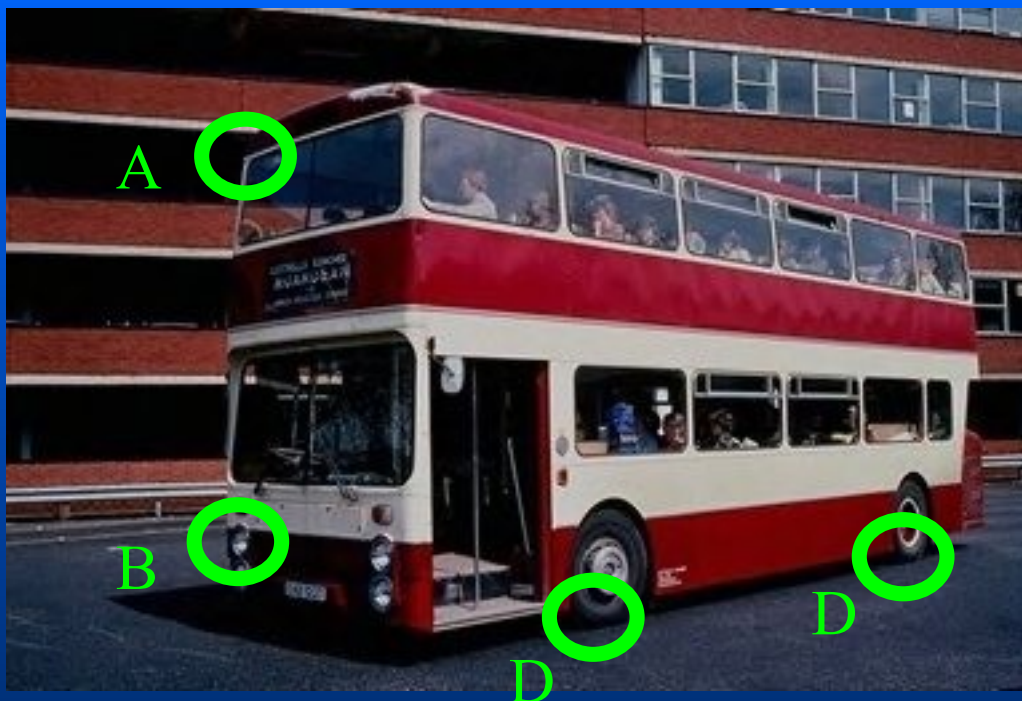
# Background

- Bag-of-Words type representations of images useful in image content analysis tasks
  - histograms of local features
  - e.g. supervised detection of object categories
  - analogous with text IR
- How to define visual words ?

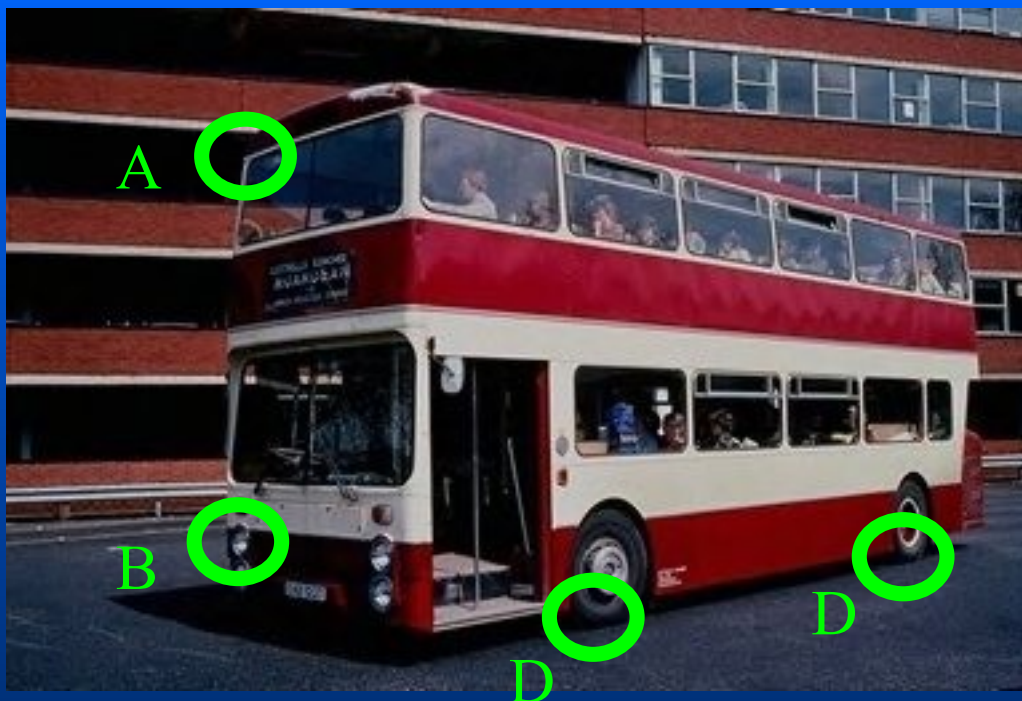








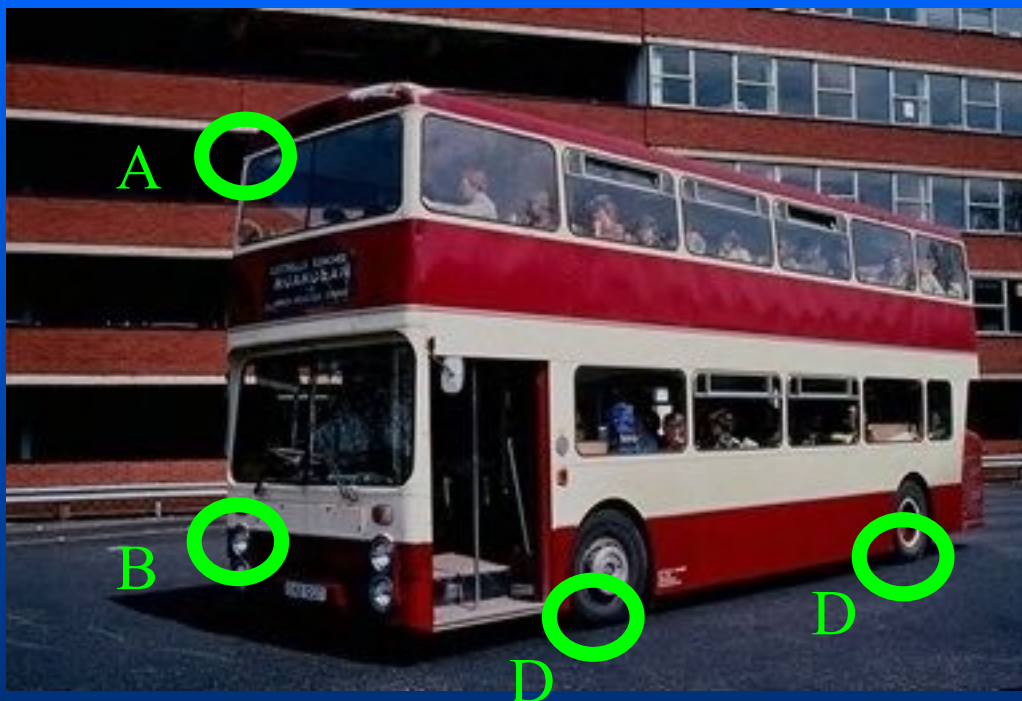
word	count
A	1
B	1
C	0
D	2



word	count
A	1
B	1
C	0
D	2

4D vector

→ (1,1,0,2)



word	count
A	1
B	1
C	0
D	2

4D vector

(1,1,0,2)

SVM

# Considered restrictions to vocabulary selection

- vocabulary by
  - » first finding too fine-grained description of neighbourhoods
  - » then group together several descriptors to one visual word
- further parametrisation: visual words required to be cells in Voronoi tessellation of (Euclidean) descriptor space

# Outline

- In this paper we experiment with several alternatives for choosing the histogram bins (=visual words)
  - post-supervised vs. pre-supervised
  - several popular clustering algorithms
  - range of codebook sizes

# Main findings

- post-supervised -> unsupervised clustering
  - differences between clustering algorithm exist
  - better quantisation error does not always imply better performance in category detection
  - rather large codebooks (~vocabulary size) worked best
- pre-supervised
  - some performance gain over post-supervised with one of the schemes

# Rest of the talk

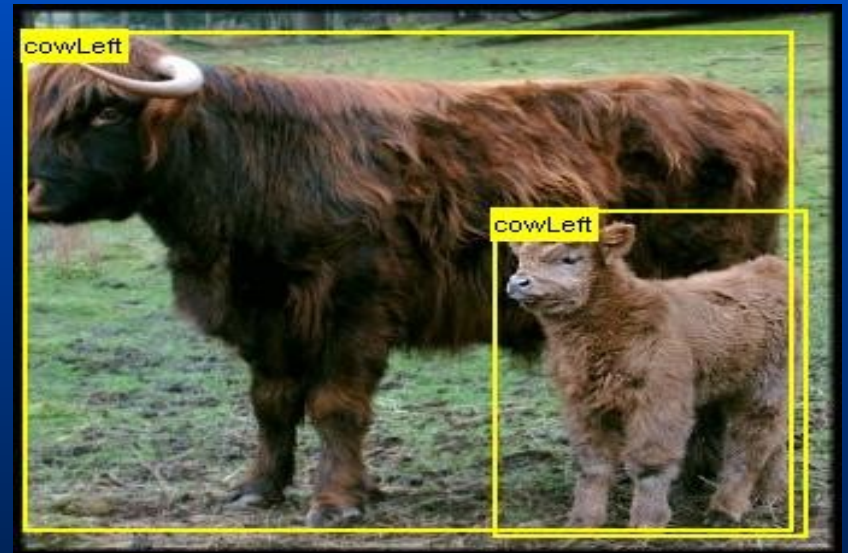
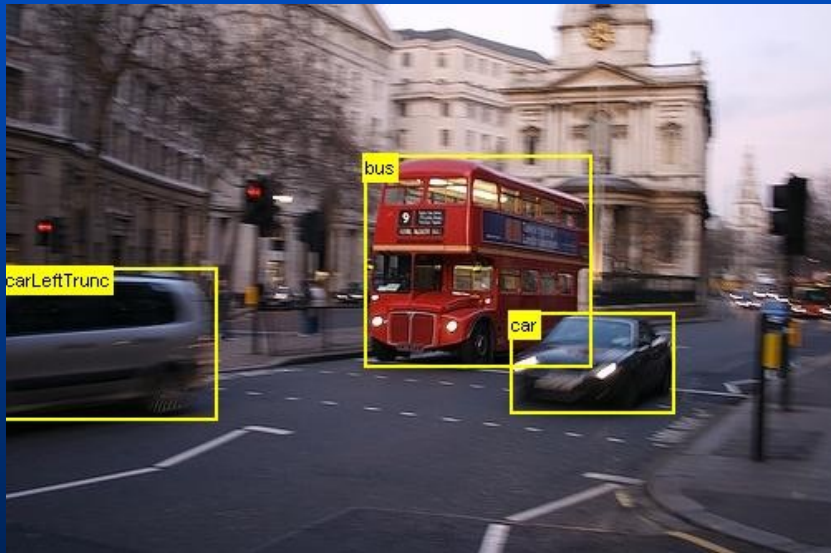
- 1) Image analysis task
- 2) Post-supervised methods
- 3) Pre-supervised methods
- 4) Results
- 5) Conclusions

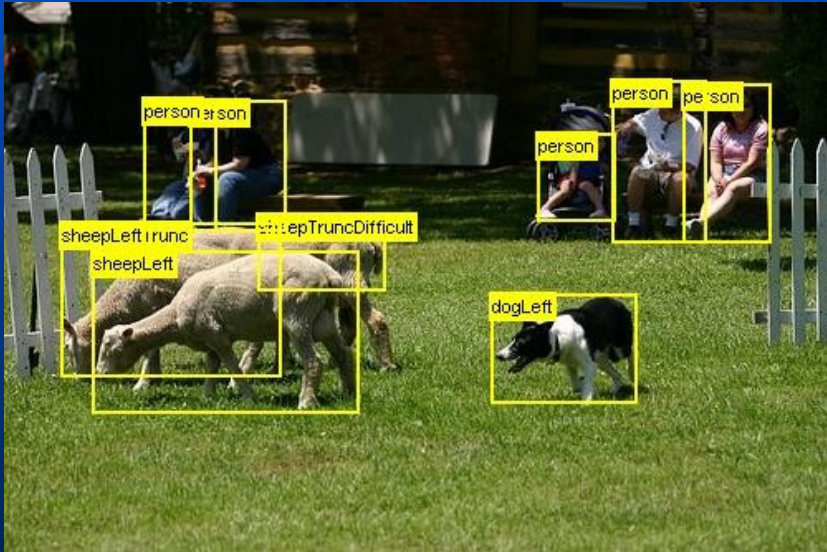
# Image analysis task

- ~ 5000 photographic images from VOC2007 collection
  - annotated for 20 object classes
  - aeroplane, bicycle, bird, bottle, boat, bus, car, cat, chair, cow, dining table,...
- supervised detection of the 20 object classes
  - 1:1 partitioning to training and test images
  - experiments repeated for six different partitionings
- interest points
  - detected with Harris-Laplace detector
  - described with SIFT descriptor
- supervised SVM classification based on histograms

# Image analysis task

- performance reported as mean average precision (MAP) of the 20 object classes





# Post-supervised methods

- -> unsupervised clustering of descriptors of all the interest points
- several popular/convenient clustering algorithms
  - randomly selecting points of the data set as cluster centers
  - k-means
  - Linde-Buzo-Gray (LBG)
  - Self-organising map (SOM)
  - Tree-structured SOM
- codebook sizes in powers of 2: 128,...,8192
- is it useful to normalise the bin counts within a single image (to e.g. unity) ?

# Pre-supervised methods

- descriptors now labelled (e.g. +1/-1)
  - histogram bins can be chosen to be maximally discriminative between classes
- three alternatives tried out
  - two-class (object class vs. background) discriminative codebook trained with Learning Vector Quantisation (LVQ) algorithm
    - » initialisation with LBG
  - two-class codebook by concatenating lists of cluster centers of separate LBG codebooks for the class and background
  - codebook by concatenating lists of cluster centers of LBG codebooks for all the 20 classes
- separate 2-class codebooks needed for each of the 20 object classes

# Results (post-supervised)

- normalisation of the bin counts within images only harmful
  - when using chi-squared kernel
- clustering algorithm
  - differences significant
  - k-means best, closely followed by LBG
  - SOM and TS-SOM worse than random bin selection
    - » despite a clearly smaller quantisation error
- performance seems to peak around codebook size 4096 (for k-means and LBG)

# Results (pre-supervised)

- compared to unsupervised LBG codebook (of comparable size)
  - 2-class LBG codebooks offer some performance advantage
  - 2-class LVQ codebooks do not offer advantage
    - » individual interest points may not be characteristic enough for an object class, but only constellations of points
  - 20-class LBG codebooks do not offer advantage (this result is not as rigorously confirmed as the others)
- increased computational cost in codebook selection phase

Title:map-unsuper-size-annotated.fig  
Creator:fig2dev Version 3.2 Patchlevel 4  
CreationDate:Tue Apr 1 14:25:35 2008

Title: map-unsuper-algo-annotated.fig  
Creator: fig2dev Version 3.2 Patchlevel 4  
CreationDate: Tue Apr 1 14:20:33 2008

# Conclusions

- pronounced and significant MAP difference between best and worst codebooks
  - differences not always predicted by quantisation MSE
- rather large codebooks best in our experiments
  - computationally somewhat cumbersome
  - how would this depend on problem size etc.?
- pre-supervised codebook selection can offer some advantage over post-supervised
  - associated increase in computational cost

# Questions

- what clustering algorithm should be added to the toolbox of tested clustering algorithms?
  - interesting/well-performing/computationally light
- is the testing methodology adequate?
  - at least convinces me
- what next?