Learning Object Categories from Google's Image Search

R. Fergus et al

Present by Jie Xiao

Dept. of Computer Science

Univ. of Texas at San Antonio

Outline

- Motivation
- “Bag of words” Model
- Approaches (pLSA, ABS-pLSA, TSI-pLSA)
- Dataset
- Experiment
- Conclusion

Motivation

- Current approaches of object categorization require manual labeled dataset as training set.
- Collecting data is time-consuming, involved in numerous human work.
- Finding good examples is another concern.

Bag of Words Model

Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that r our eyes. For a long ti retinal image wa isual centers i s a movie s image discove know th perceptio y more com following the ath to the various c ortex, Hubel and Wiesel ha demonstrate that the *message abou image falling on the retina undergoes p- wise analysis in a system of nerve cell stored in columns. In this system each c has its specific function and is responsible a specific detail in the pattern of the retinal image.*

sensory, brain, visual, perception, retinal, cerebral cortex, eye, cell, optical nerve, image Hubel, Wiesel

China is forecasting a trade surplus of $90bn (£51bn) to $100bn this year, a threefold increase on 2004's $32bn. The Commerce Ministry said the surplus would be created by a predicted 30% $750bn, compared w $660bn. T annoy th China's deliber agrees yuan is governo also need demand so country. China e yuan against the dol nd permitted it to trade within a narrow but the US wants the yuan to be allowed de freely. However, Beijing has made it cl t it will take its time and tread carefully be allowing the yuan to rise further in value.

China, trade, surplus, commerce, exports, imports, US, yuan, bank, domestic, foreign, increase, trade, value

Slide credit: Rob Fergus

Bag of Words Model

- LSA:
- U and V are orthonormal matrices

$$X = U\Sigma V^T$$

- A singular value decomposition(SVD) process

- pLSA

Bag of Words Model -- pLSA

- D: set of documents
- W: visual words
- Z: topics
- Latent variable z is associate with w and d.

- Matrix $N_{M \times N}$:co-occurrence of words and doc

- N (w,d) : the number of word w appears in document d.

Bag of Words Model – pLSA (Cont.)

$P(w|z)$ co-occurrence of words within a topic

$P(z|d)$ density of topic on a given document

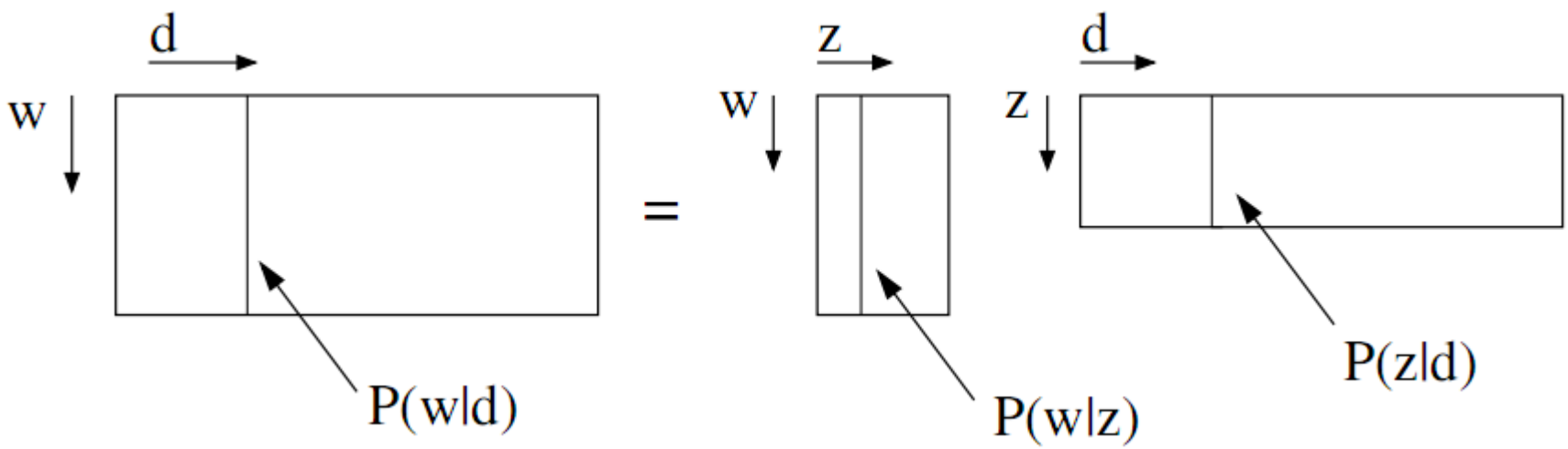
$$P(w_i|d_j) = \sum_{k=1}^{K} P(z_k|d_j)P(w_i|z_k)$$

$$P(w,d) = \sum_{z=1}^{Z} P(w|z)P(z|d)P(d) \qquad (1)$$

Bag of Words Model – pLSA (Cont.)

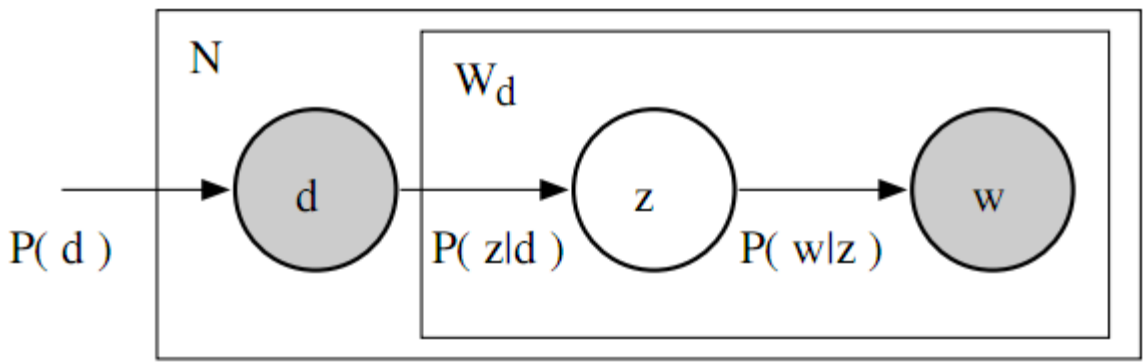
- $P(w|z)$ topic specific word distribution
- $P(z|d)$ document specific mixing proportion

$$P(w_i|d_j) = \sum_{k=1}^{K} P(z_k|d_j)P(w_i|z_k)$$



Bag of Words Model – pLSA (Cont.)

$$P(w,d) = \sum_{z=1}^{Z} P(w|z)P(z|d)P(d)$$



Bag of Words Model – pLSA (Cont.)

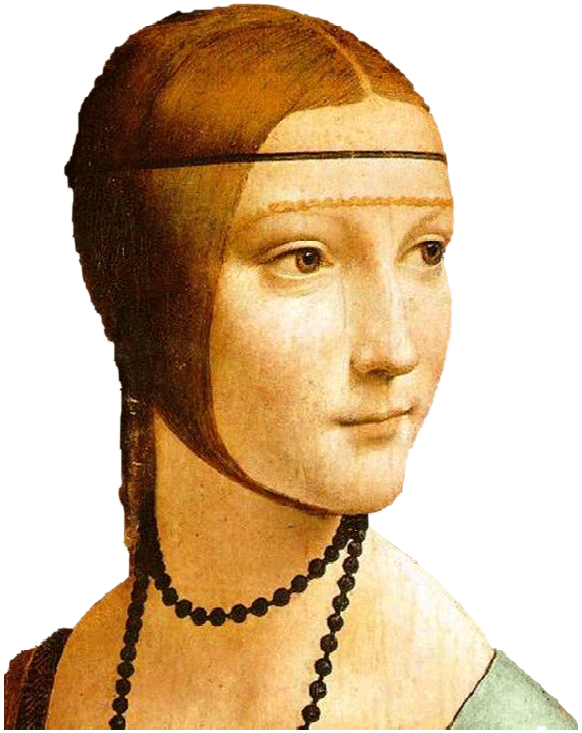
- Calculating by EM
- E step: $P(z|w,d)$
- M step: $P(d)\ P(z|d)\ P(w|z)$

$$P(w,d) = \sum_{z=1}^{Z} P(w|z)P(z|d)P(d)$$

$$L = \prod_{d=1}^{D} \prod_{w=1}^{W} P(w,d)^{n(w,d)}$$

Bag of Words Model (Cont.)

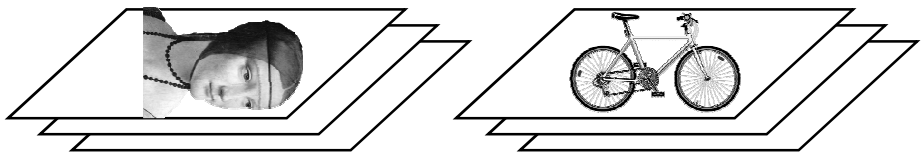
Object → **Bag of words**



Slide credit: Rob Fergus

Bag of Words Model (Cont.)

1. Representation



feature detection
& representation

2. codewords dictionary

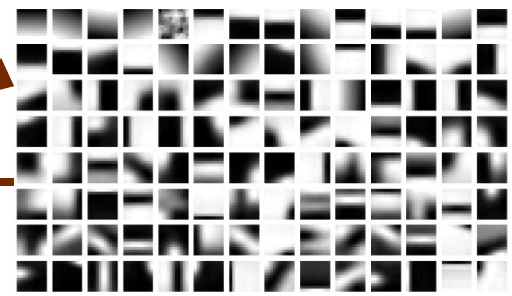
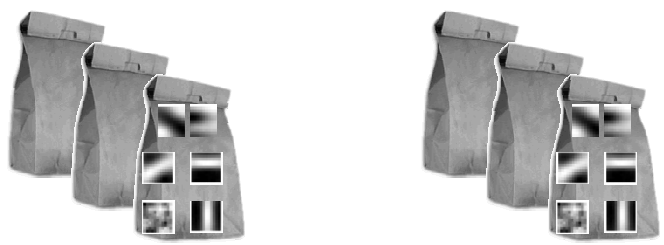


image representation

3.



Slide credit: Rob Fergus

Approach

- ABS-pLSA

  Quantize the location within the image into one of X bins

Use $P(w, x, d) = \sum_{z=1}^{Z} P(w, x|z)P(z|d)P(d)$

Instead of

$$P(w, d) = \sum_{z=1}^{Z} P(w|z)P(z|d)P(d)$$

Approach (Cont.)

- TSI-pLSA

  Introducing latent variable, c, represents the centriod of the object.

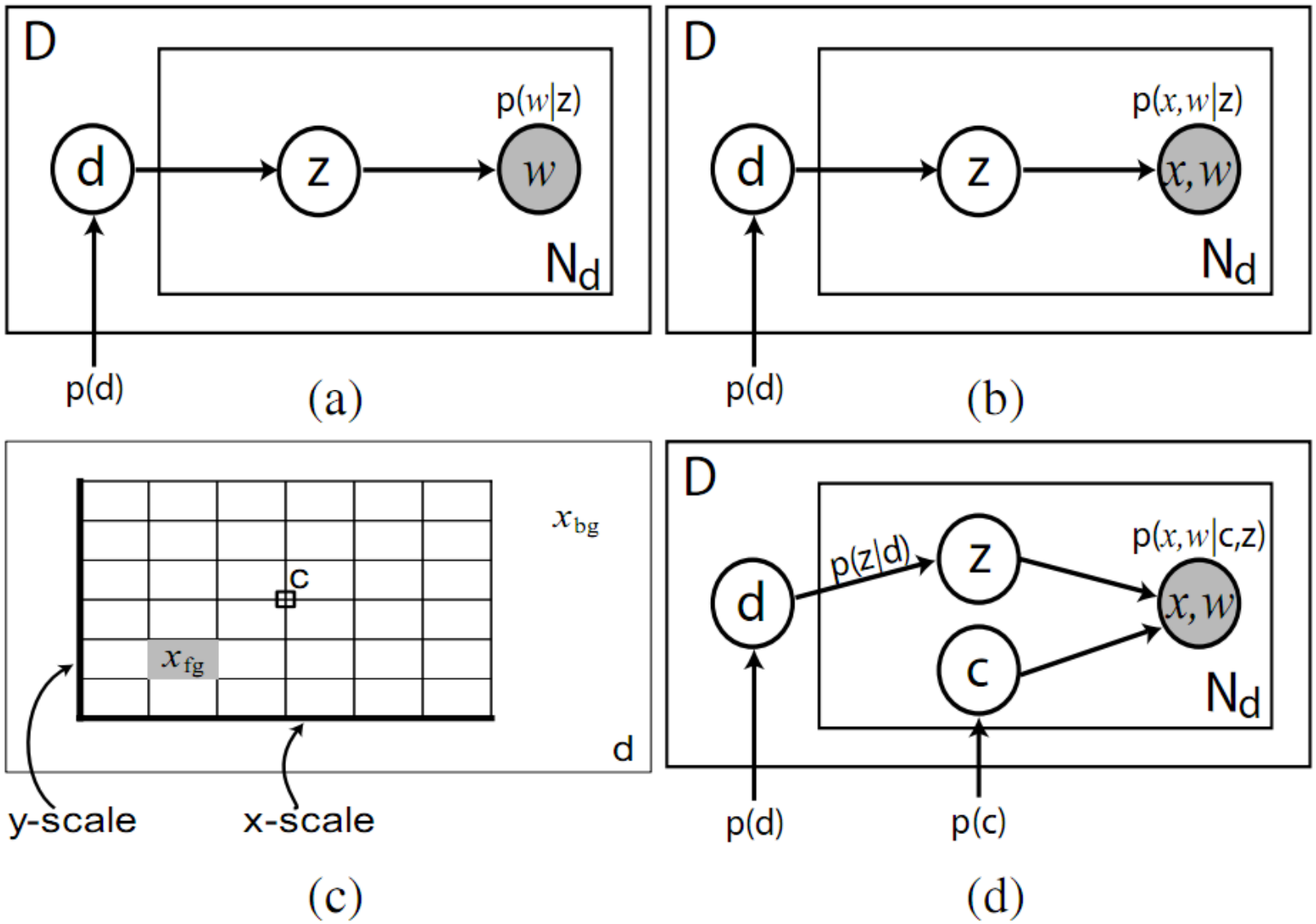
  X_{fg} foreground bins

  x_{bg} background bin

  $$X = X_{fg} + 1$$

$$P(w, x|z) = \sum_c P(w, x, c|z) = \sum_c P(w, x|c, z)P(c)$$

Approach (Cont.)



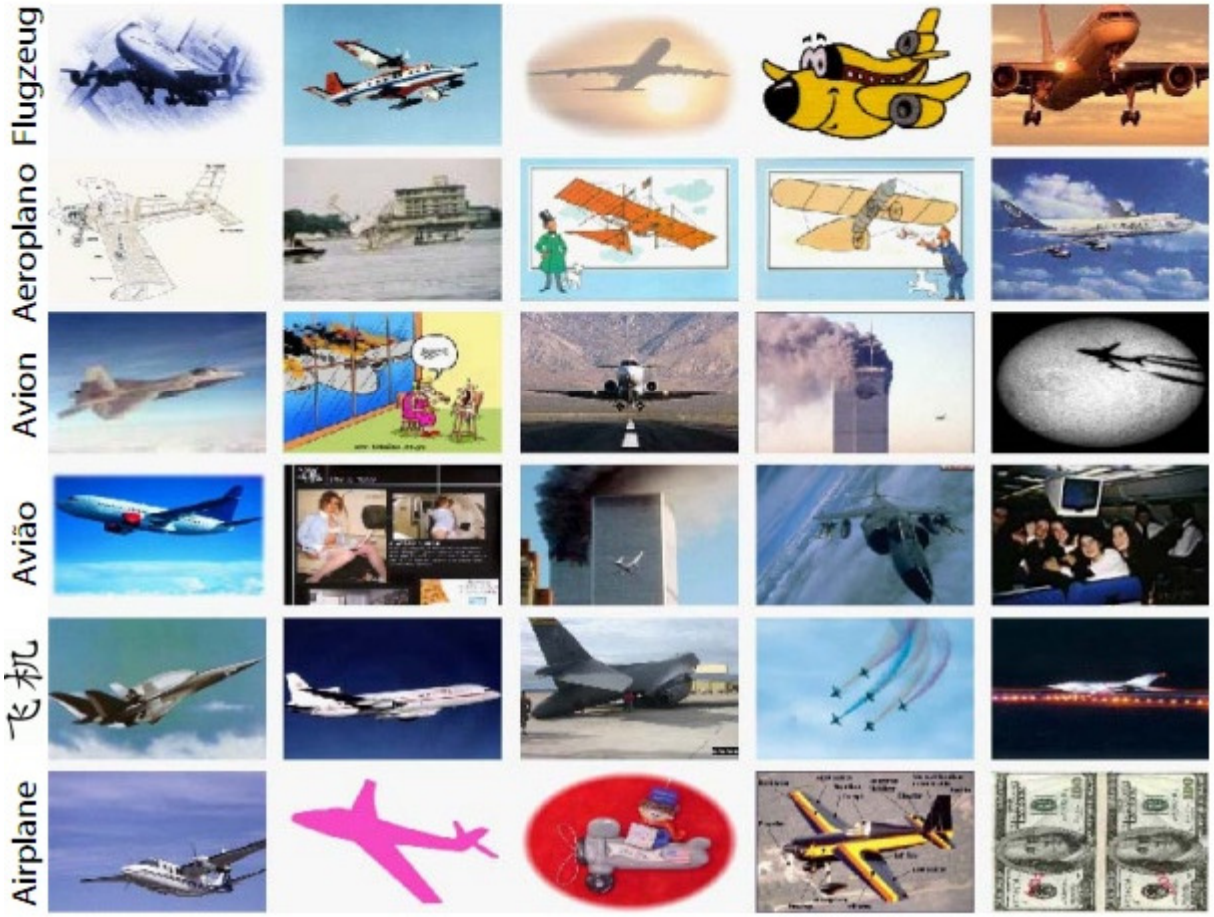
Datasets

- PT: prepared training set, manually gathered
- P: prepared test set
- G: raw download data from Google image.
  - Good image: good examples, related to keyword category
  - Intermediate images: related to keyword category, low quality than good image
  - Junk images: totally unrelated to the keyword category

Datasets (Cont.)

- V: Google validation set.
  - Assume the images from first pages are positive examples.
  - Cross language collections

Datasets (Cont.)



Datasets (Cont.)

- statistics

	Size of Dataset				Distrib. of Google Images (%)		
Category	PT	P	V	G	Good	Inter.	Junk
Airplane	400	400	30	874	18.1	8.6	73.3
Cars Rear	400	400	30	596	32.2	12.9	54.9
Face	217	217	30	564	24.3	21.3	54.4
Guitar	450	450	25	511	25.3	30.5	44.2
Leopard	100	100	15	516	19.6	27.5	52.9
Motorbike	400	400	30	688	33.4	29.8	36.8
Wrist watch	180	181	35	342	63.4	13.8	22.8
PASCAL Cars	272	275	-	-	-	-	-
PASCAL Cars Bg.	412	412	-	-	-	-	-
PASCAL Motorbike	214	202	-	-	-	-	-
PASCAL Motorbike Bg.	570	754	-	-	-	-	-
Caltech Bg.	400	400	-	-	-	-	-
Cars Rear Bg.	400	400	-	-	-	-	-

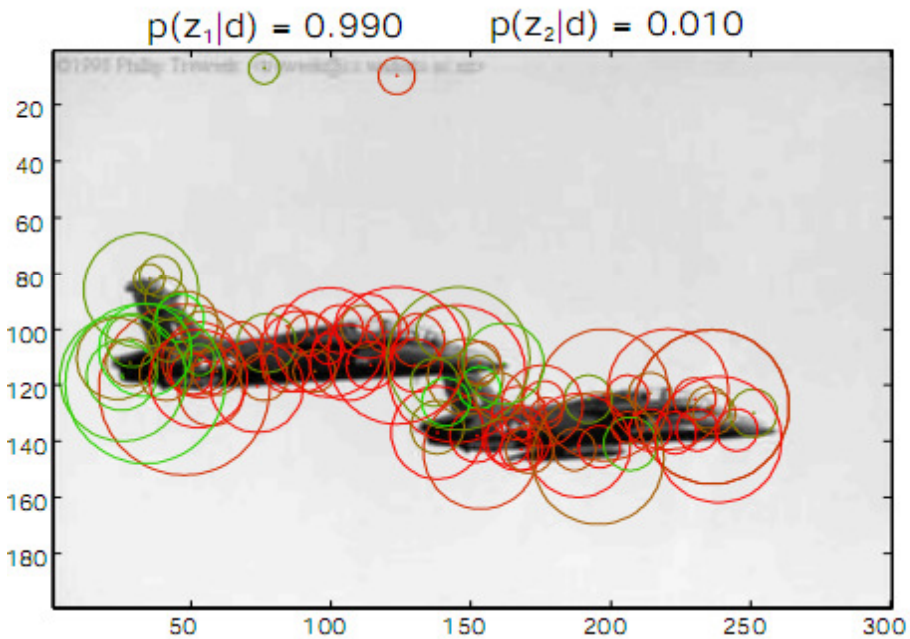
Experiments

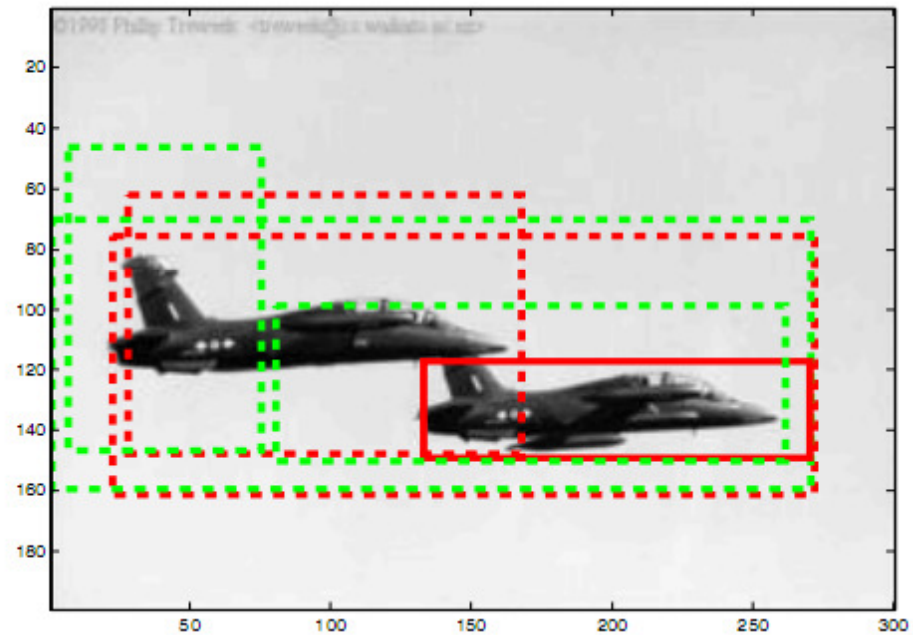
- Region detectors:
  - Convert to grayscale
  - Resize to a moderate size
  - Detect region
  - Represent by SIFT descriptor
  - Quantize descriptor vector

Experiments – region detector

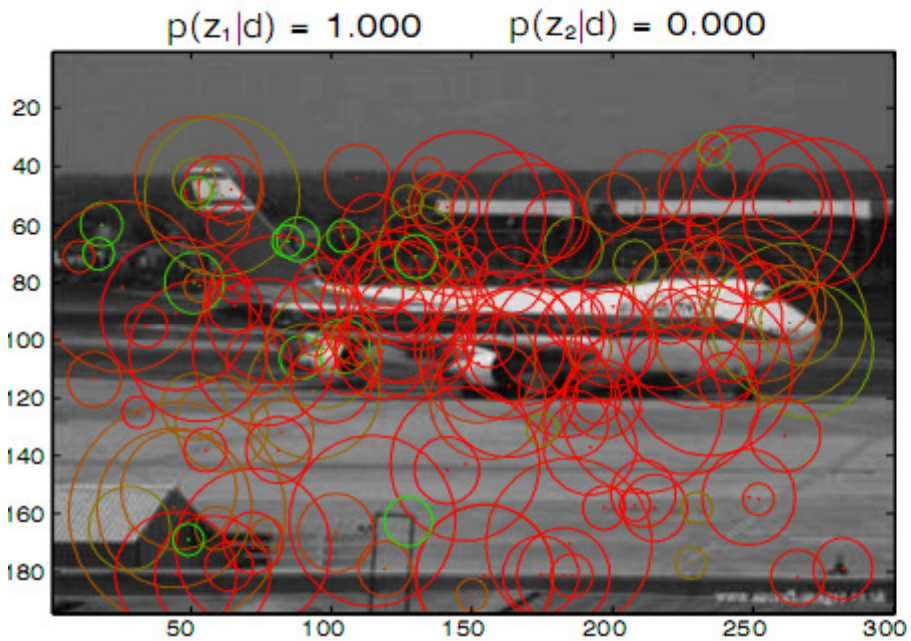
- Region detectors:
  - Kadir & Brady saliency operator
  - Multi-scale Harris detector
  - Difference of Gaussian
  - Edge based operator

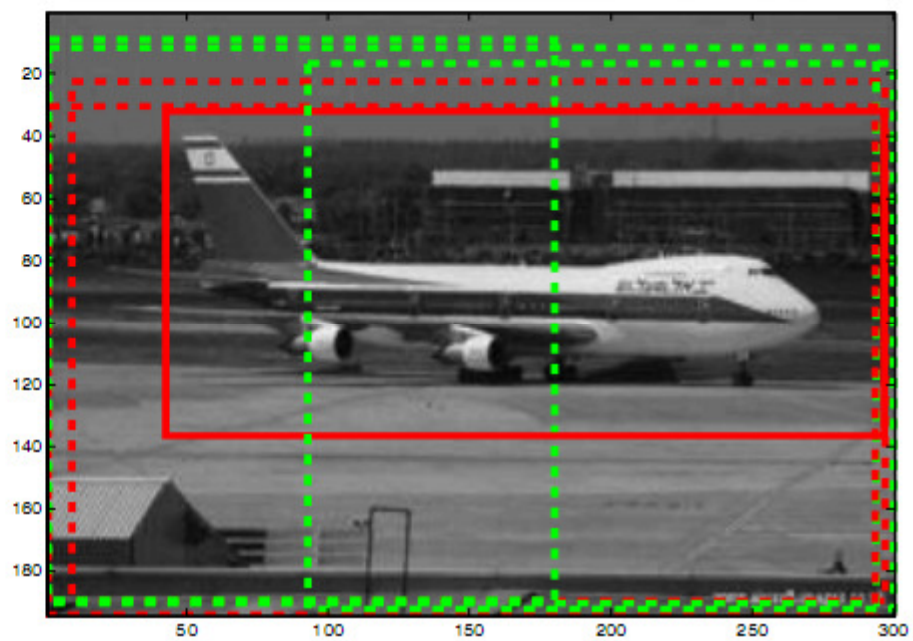
Experiments (Cont.)



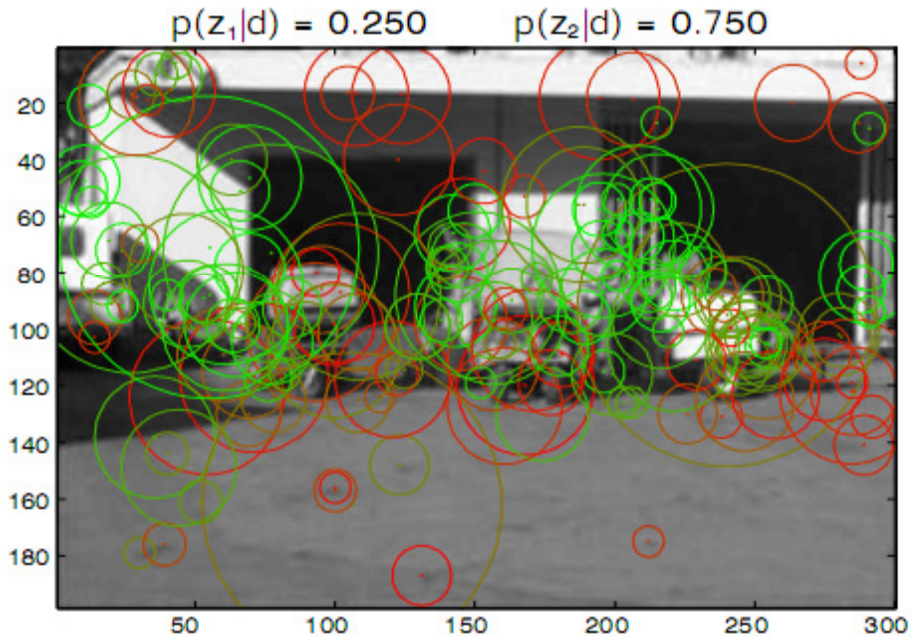


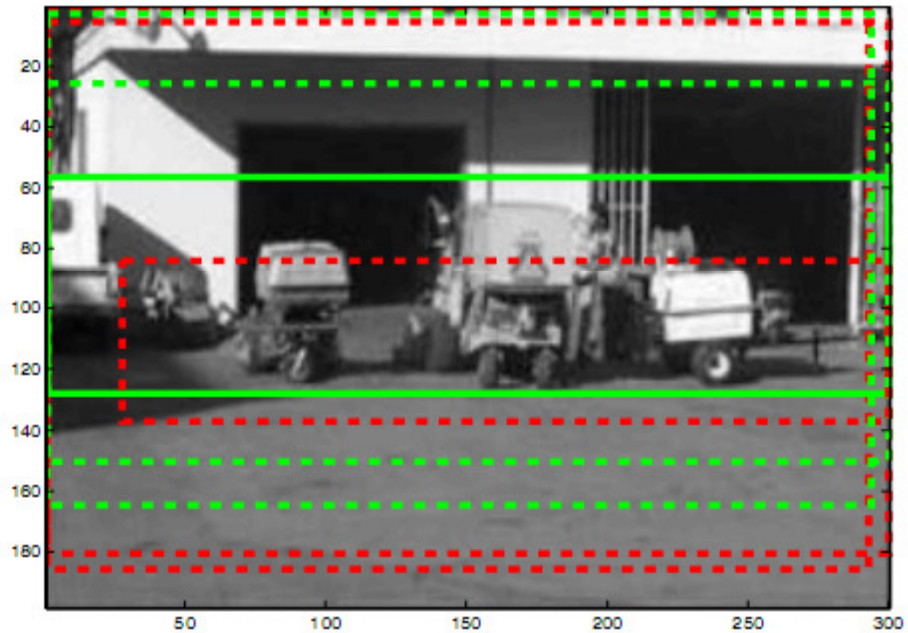
Experiments (Cont.)



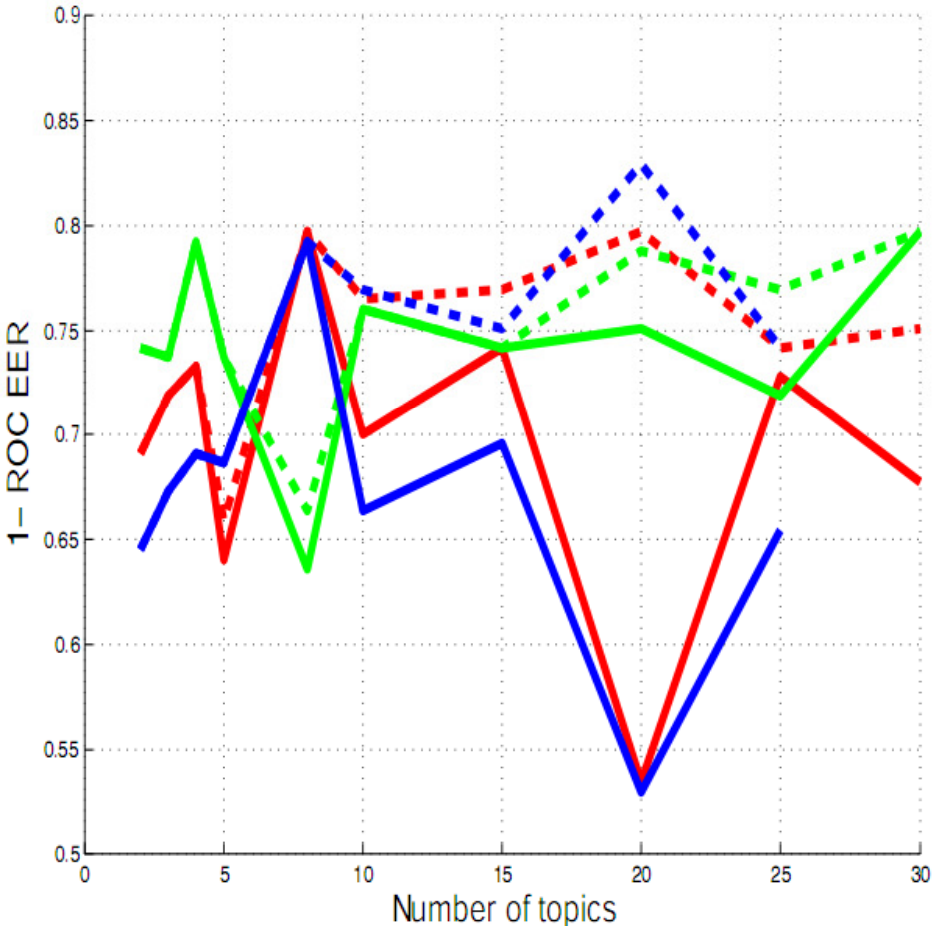


Experiments (Cont.)





Experiments (Cont.)



- Red: pLSA
- Green: ABS-pLSA
- Blue: TSI-pLSA
- Solid line: performance of automatically chosen topic within model
- Dashed line: performance of best topic within model

Discussion

- Limited categories
- Prior knowledge about number of categories
- Image background
- Similar visual word

Conclusion

- Introduce spatial information in pLSA.
- Learn object category by category name.

Thank you!