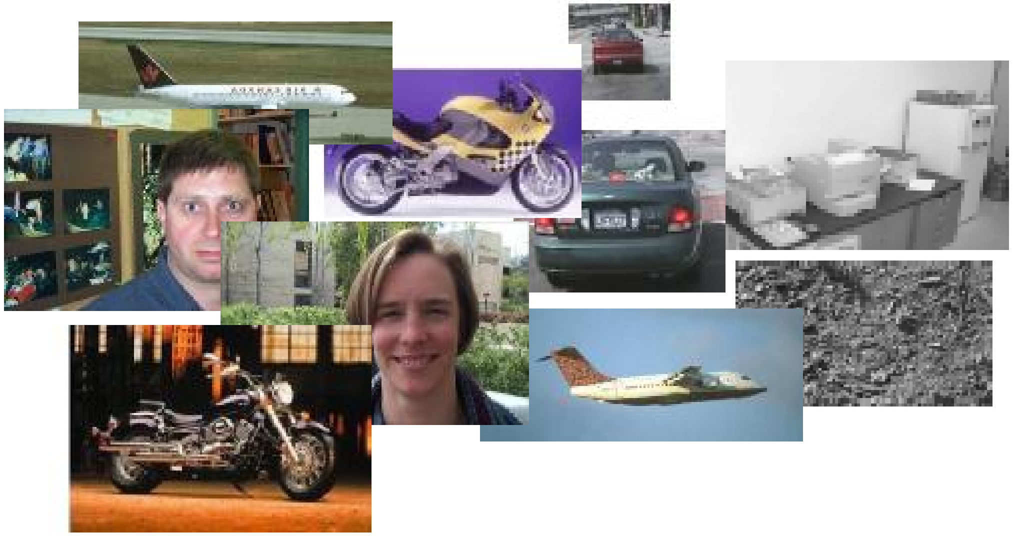
Using Multiple Segmentations to Discover Objects and their Extent in Image Collections

Carolina Galleguillos

Introduction

Goal: Given a collection of unlabelled images, discover visual object categories and their segmentation automatically.



Approach: 1) Produce multiple segmentations of each image.

2) Discover clusters of similar segments.

3) Score all segments by how well they fit object cluster.

Background

The task of discovering objects and scene categories

[Fei-Fei & Perona, 2005] [Quelhas et al, 2005] and [Sivic et all, 2005]

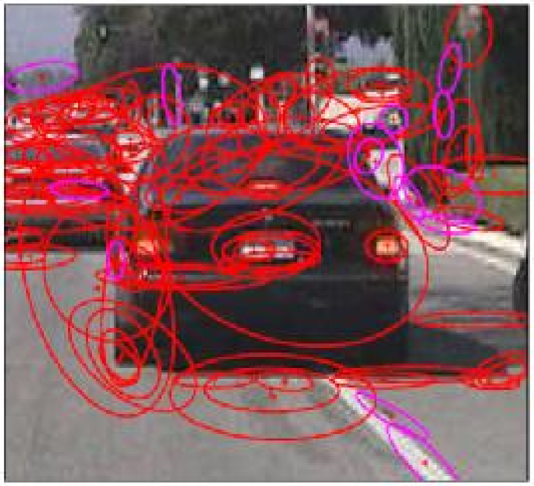
→ Borrowing tools from the statistical text analysis community (pLSA and LDA) that use bag of words approach.

MAPPING ONTO VISUAL DOMAIN:

§ Images are treated as **documents**.

§ Cluster affine invariant point descriptors as visual **words**.

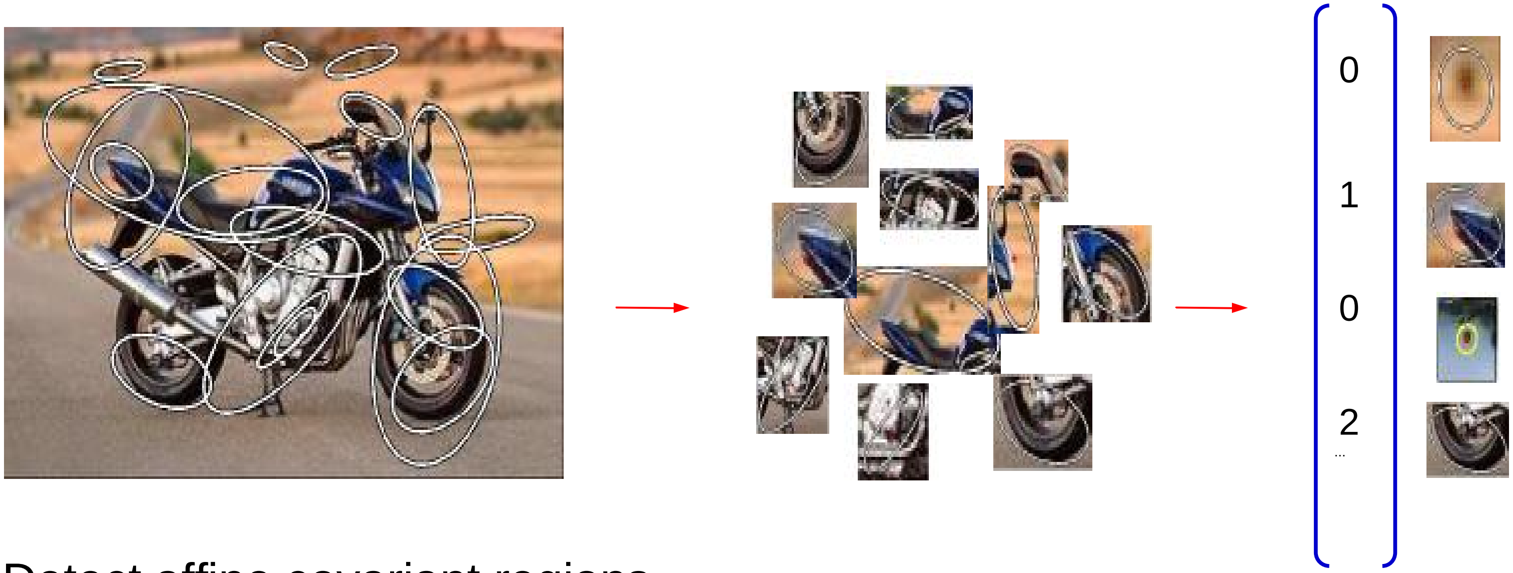
§ Each Image is represented by a histogram of visual words.



Issues: Visual words are not always as descriptive as text (visual phonemes or visual letters).

Background: Bag-of-words Approaches

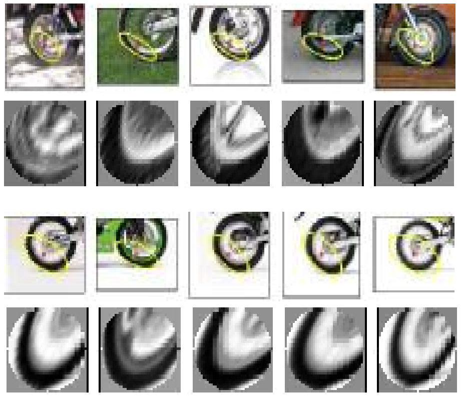
Represent an image as a histogram of “visual words”



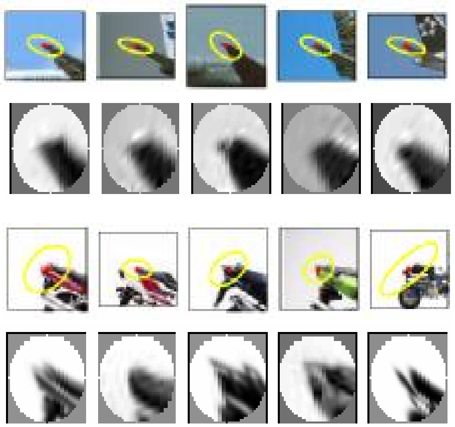
- Detect affine covariant regions.
- Represent each region by a SIFT descriptor.
- Build visual vocabulary by k-means clustering (K~1,000).
- Assign each region to the nearest cluster centre.

Visual word shortcomings

Visual Synonyms: Two different visual words representing a similar part of an object (wheel of a motorbike).

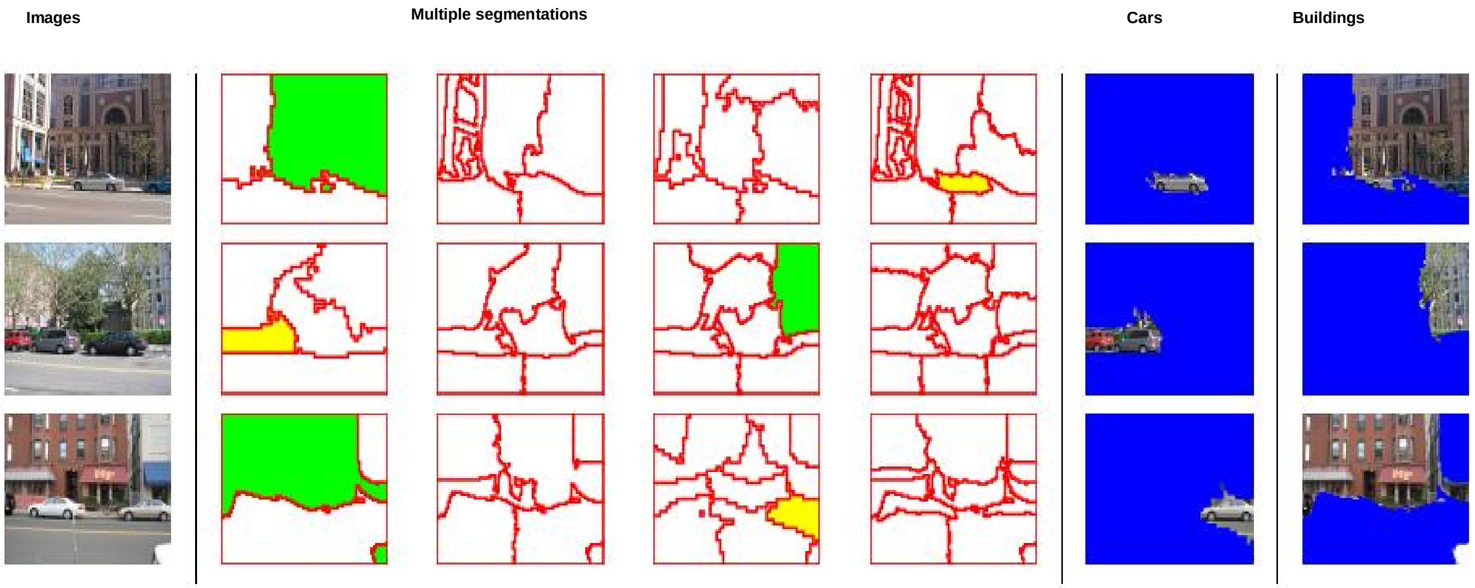


Visual Polysemy: Single visual word occurring on different (but locally similar) parts on different object categories.



If the object and its background are highly correlated, modelling the entire image can actually help recognition.

Multiple segmentations for to produce groups of visual words



Intuition #1: *All segmentations are wrong, but some segments are good*

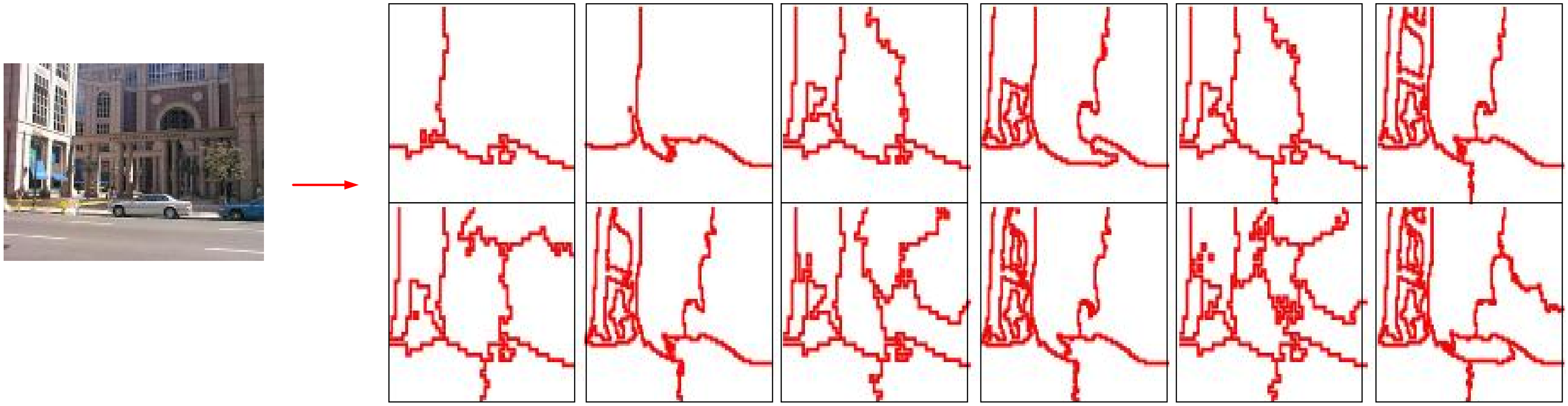
Intuition #2: *All good segments are alike, each bad segment is bad in its own way.*

The Algorithm

Given a large collection of unlabeled images:

1. For each image, compute multiple candidate segmentations using Normalized-Cuts.
2. For each segment, compute histograms of visual words.
3. Perform topic discovery, treating each segment as a document, using LDA over all segments in the collection.
4. For each topic sort segments using KL divergence.

Multiple segmentations

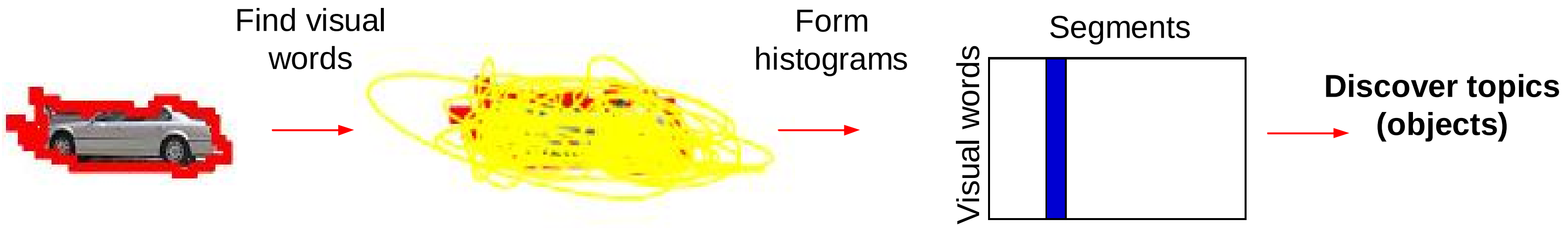


We use Normalized Cuts, varying parameter settings:

# segments and image scale.

Discovering Objects

Representing Segments:



Find visual words

Form histograms

Segments

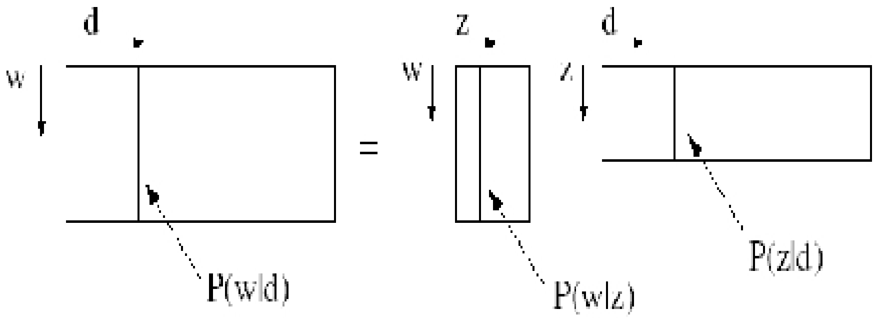
Visual words

Discover topics (objects)

Finding coherent segment clusters (topics):

w … visual words d … documents (images) z … topics (‘objects’)

Use statistical text analysis techniques such as Latent Semantic Analysis (LSA), Probabilistic LSA [Hofmann ’99] or Latent Dirichlet Allocation (LDA) [Blei et al. ’03]. Here we chose LDA.



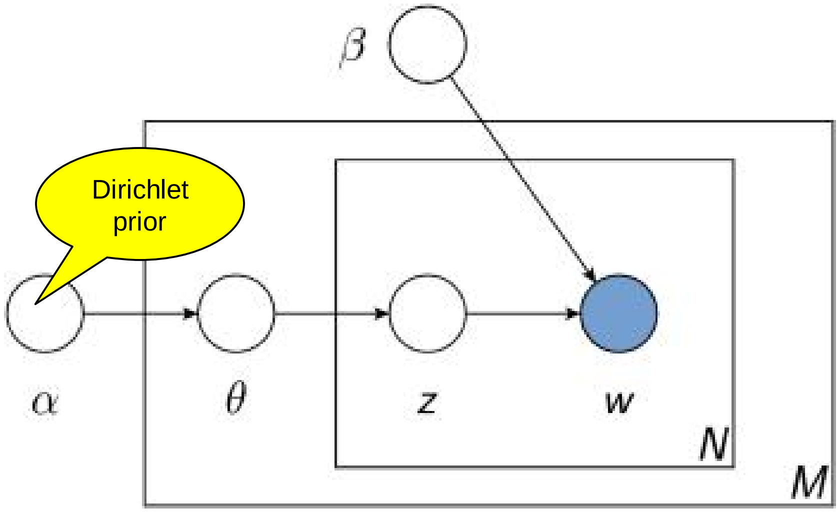
P(w|d), P(z|d) and P(w|z) are multinomial distributions

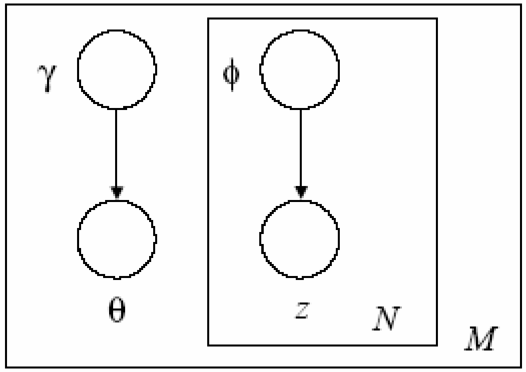
Latent Dirichlet Allocation [Blei et al, 2003]

Generative probabilistic model for collections of discrete data such as text corpora.

LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics.

Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities.



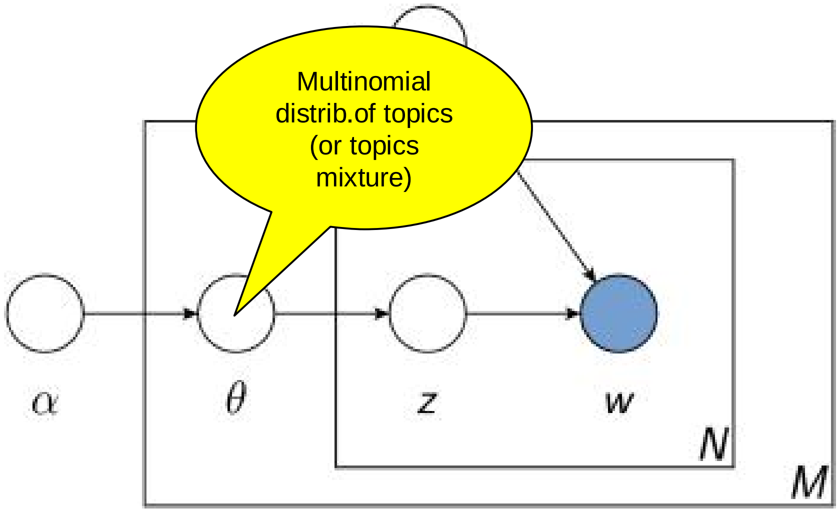


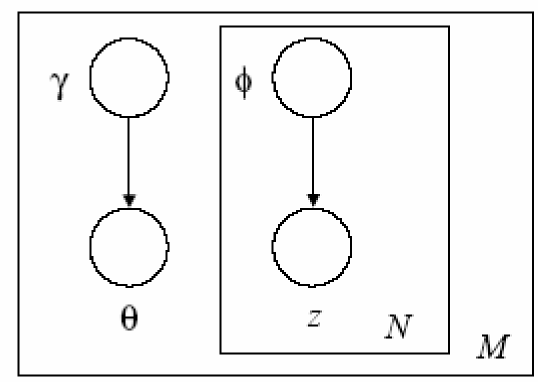
Latent Dirichlet Allocation [Blei et al, 2003]

Generative probabilistic model for collections of discrete data such as text corpora.

LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics.

Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities.



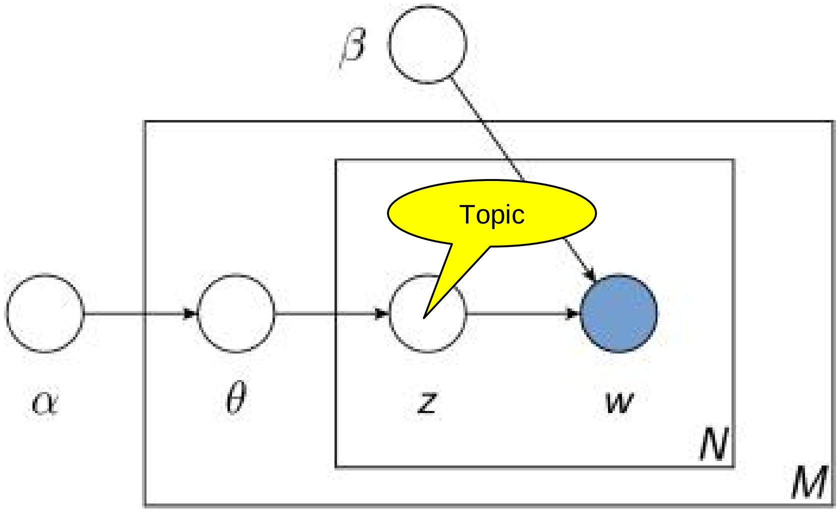


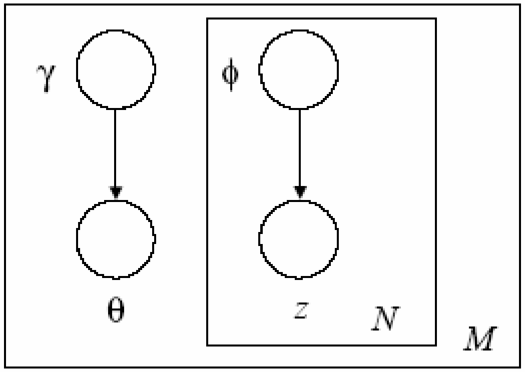
Latent Dirichlet Allocation [Blei et al, 2003]

Generative probabilistic model for collections of discrete data such as text corpora.

LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics.

Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities.



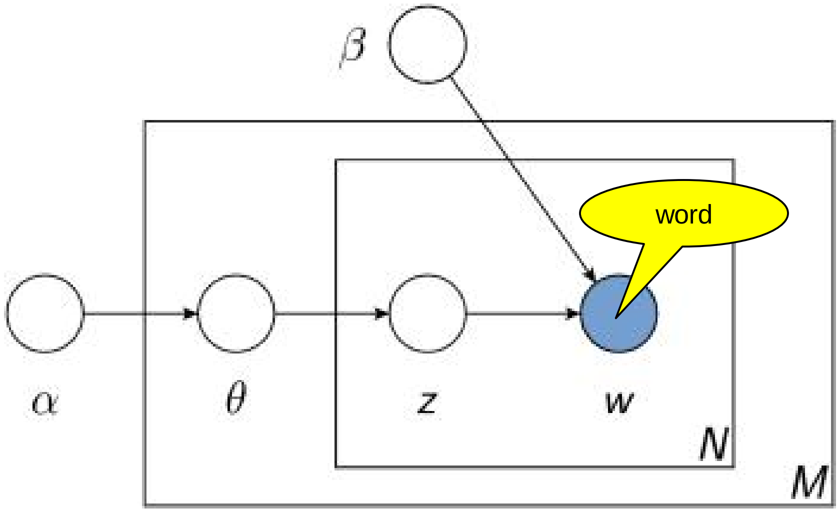


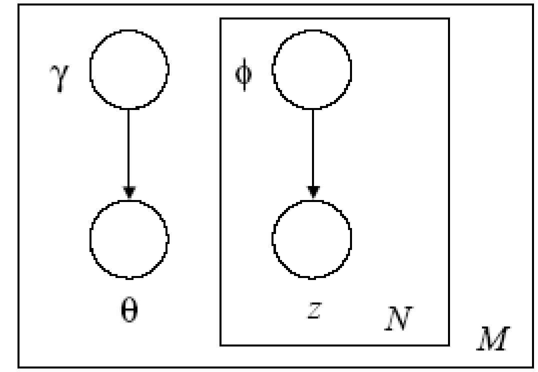
Latent Dirichlet Allocation [Blei et al, 2003]

Generative probabilistic model for collections of discrete data such as text corpora.

LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics.

Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities.



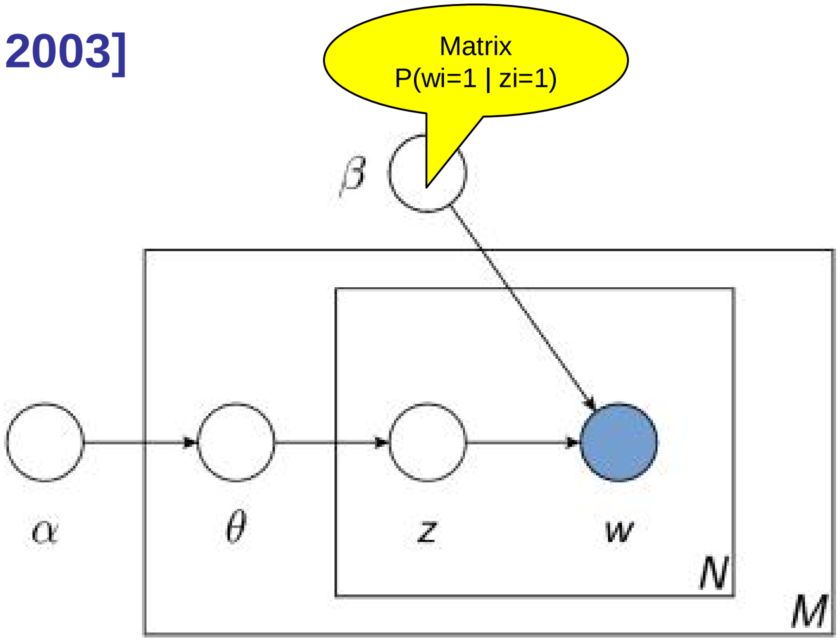


Latent Dirichlet Allocation [Blei et al, 2003]

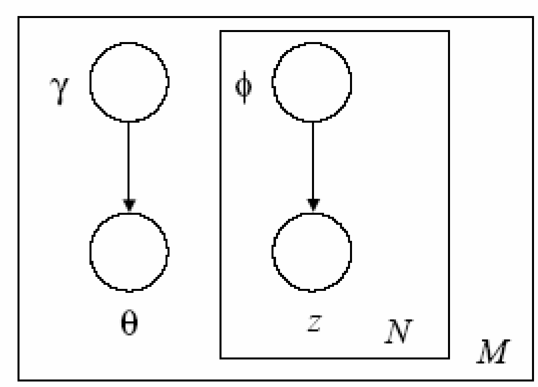
Generative probabilistic model for collections of discrete data such as text corpora.

LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics.

Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities.



Matrix
P(wi=1 | zi=1)

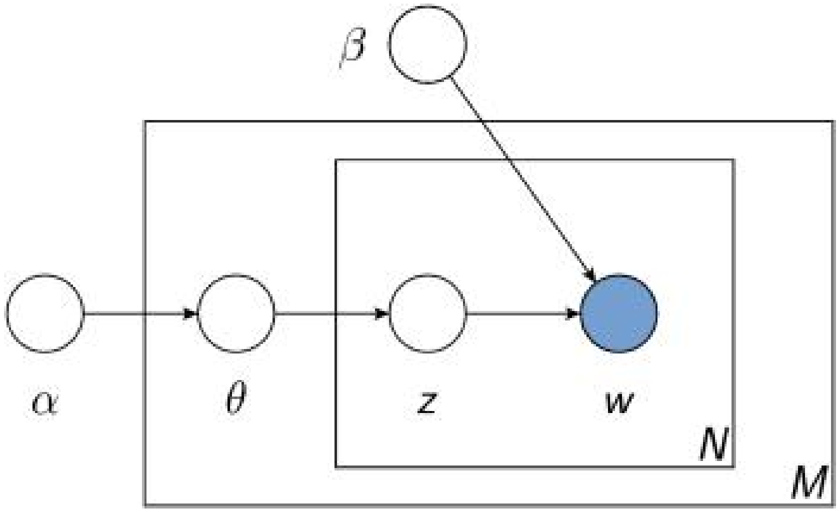


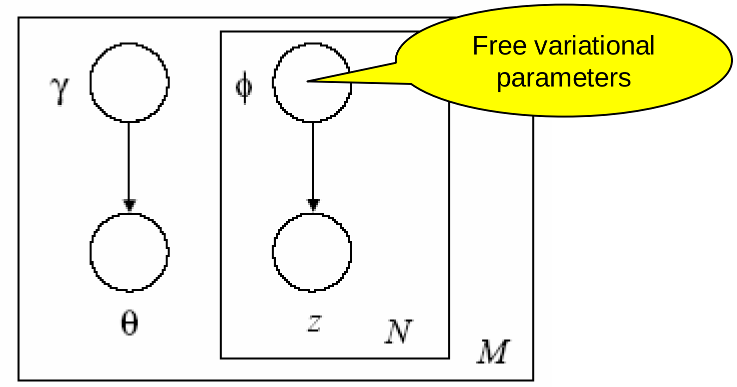
Latent Dirichlet Allocation [Blei et al, 2003]

Generative probabilistic model for collections of discrete data such as text corpora.

LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics.

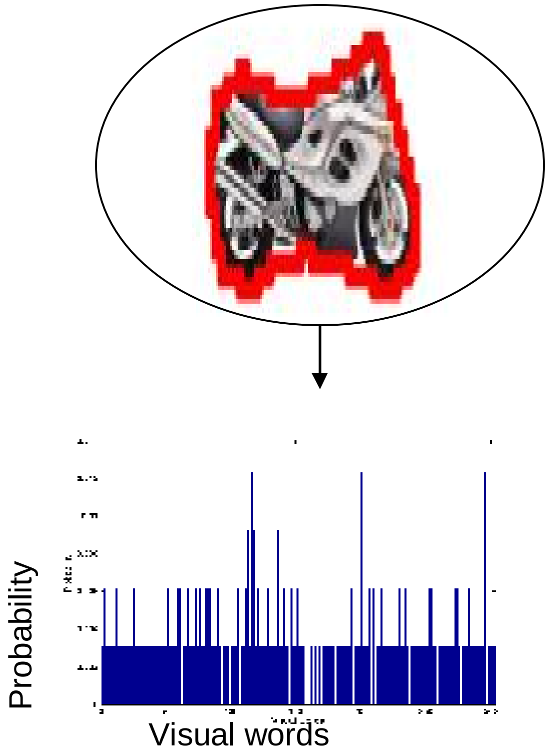
Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities.



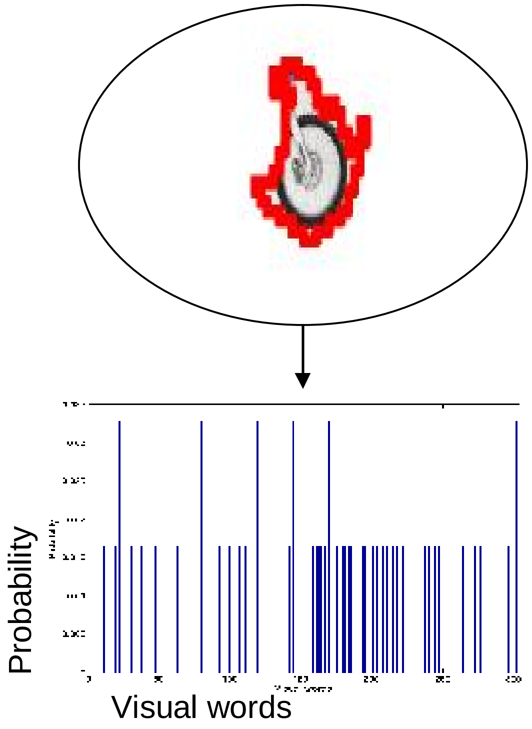


Segment scoring

Compare segment distributions against learned topic distribution over visual words using KL divergence

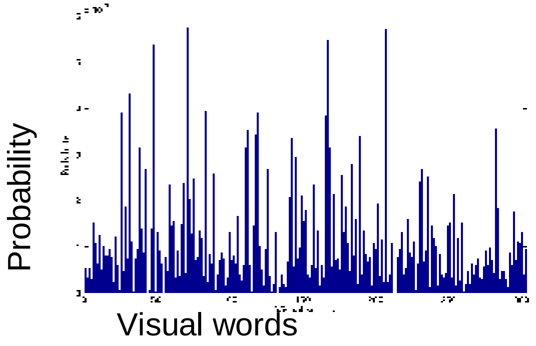


KL divergence: 1.89

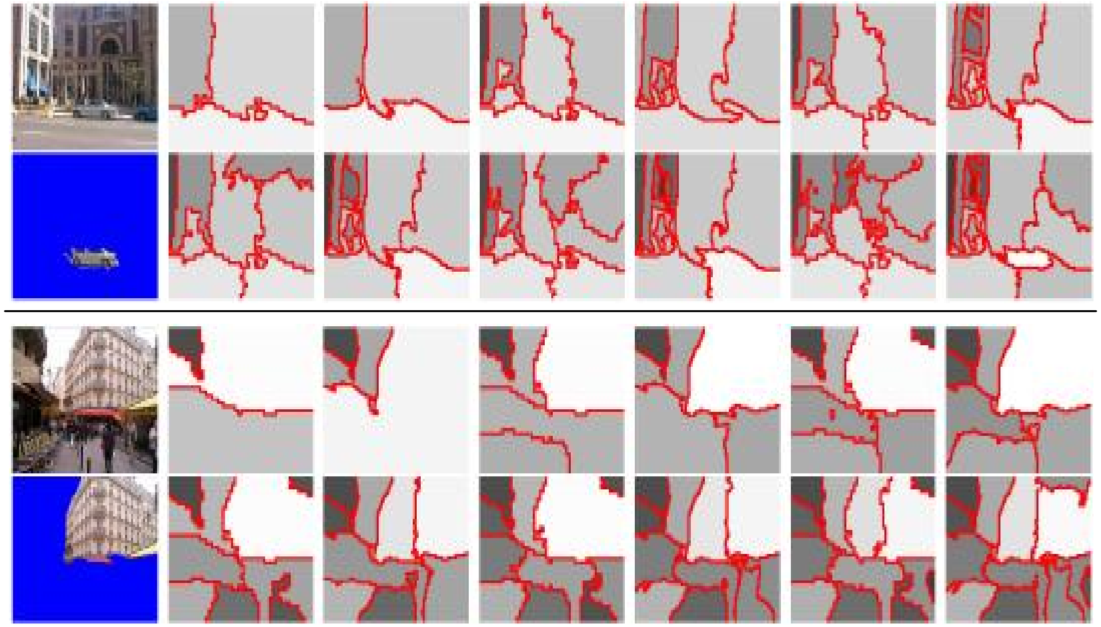


KL divergence: 2.90

Learned topic distribution



Segmentations and their KL divergence

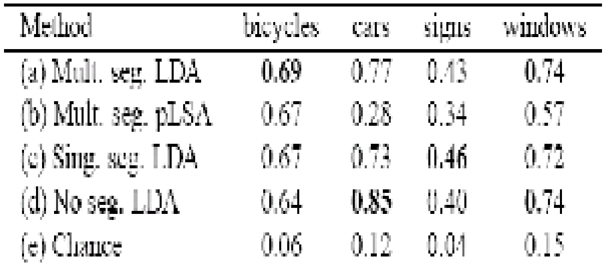


KL divergence

$$D_{\mathrm{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

$$\mathrm{D}(q(\theta, \mathbf{z} \,|\, \gamma, \phi) \;\|\; p(\theta, \mathbf{z} \,|\, \mathbf{w}, \alpha, \beta))$$

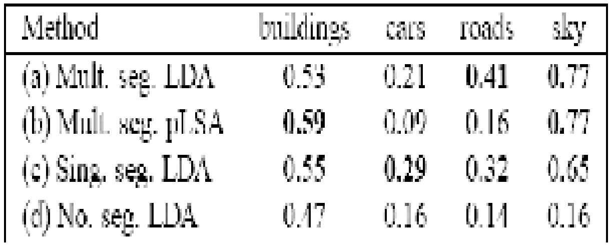
Retrieval accuracy



Method	bicycles	cars	signs	windows
(a) Mult. seg. LDA	**0.69**	0.77	0.43	**0.74**
(b) Mult. seg. pLSA	0.67	0.28	0.34	0.57
(c) Sing. seg. LDA	0.67	0.73	**0.46**	0.72
(d) No seg. LDA	0.64	**0.85**	0.40	**0.74**
(e) Chance	0.06	0.12	0.04	0.15

Average precision for MSRC

Segmentation accuracy



Method	buildings	cars	roads	sky
(a) Mult. seg. LDA	0.53	0.21	**0.41**	**0.77**
(b) Mult. seg. pLSA	**0.59**	0.09	0.16	**0.77**
(c) Sing. seg. LDA	0.55	**0.29**	0.32	0.65
(d) No. seg. LDA	0.47	0.16	0.14	0.16

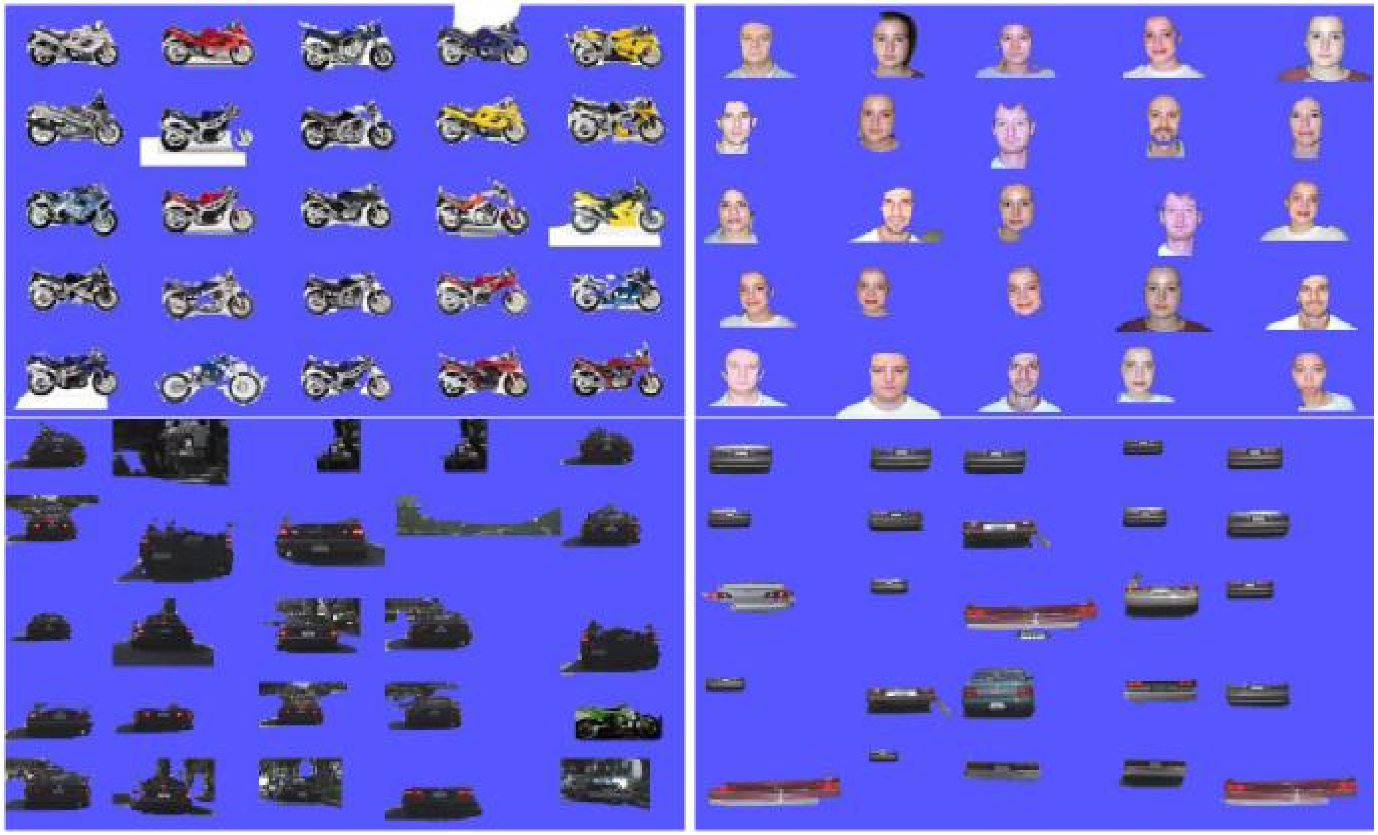
Average overlap area score for LabelMe

Results

Montages of top segments given a discovered object category

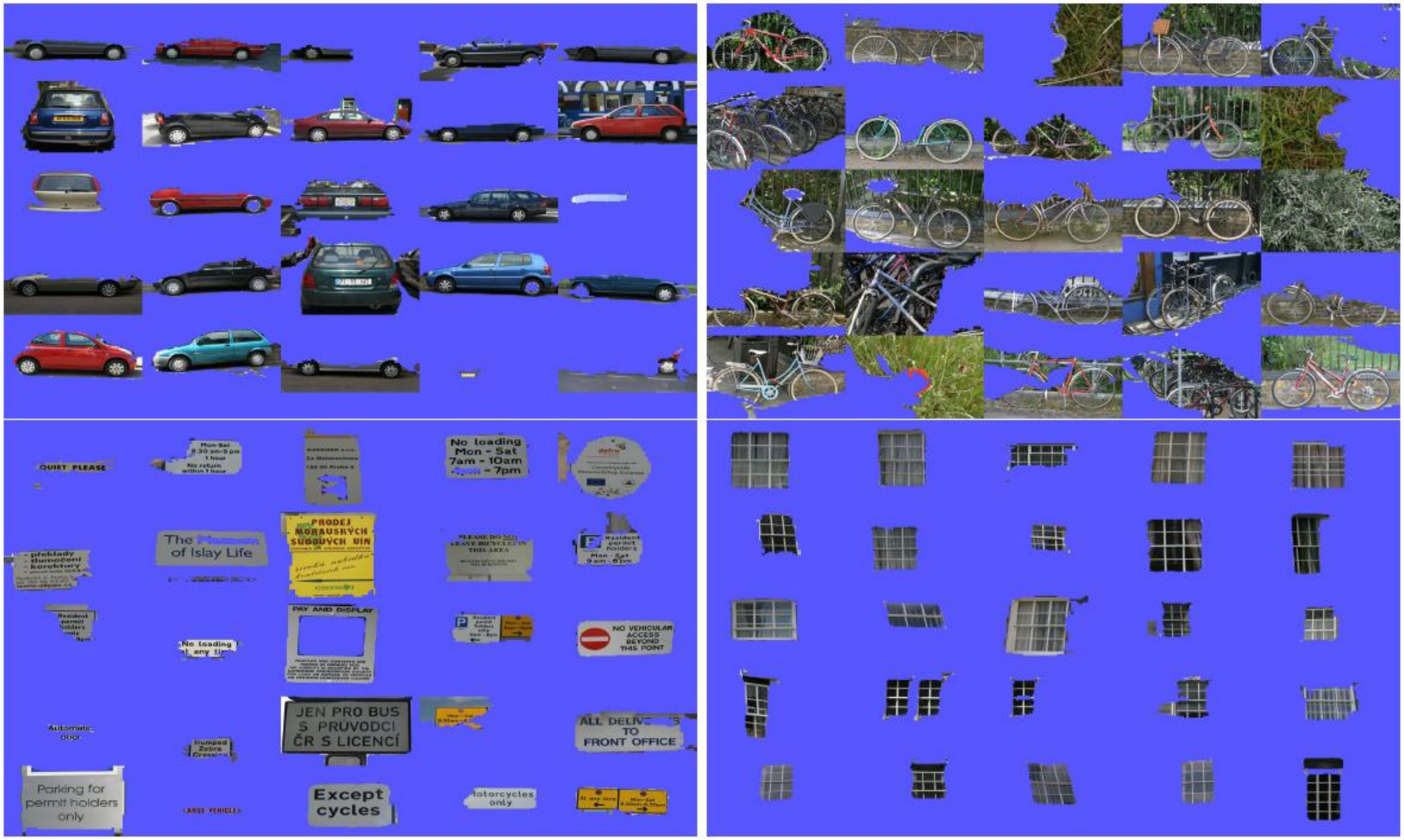
Sort segments based on their KL divergence score computed against the learned visual word distribution for a given topic

Caltech 5: 10 topics, 4090 images

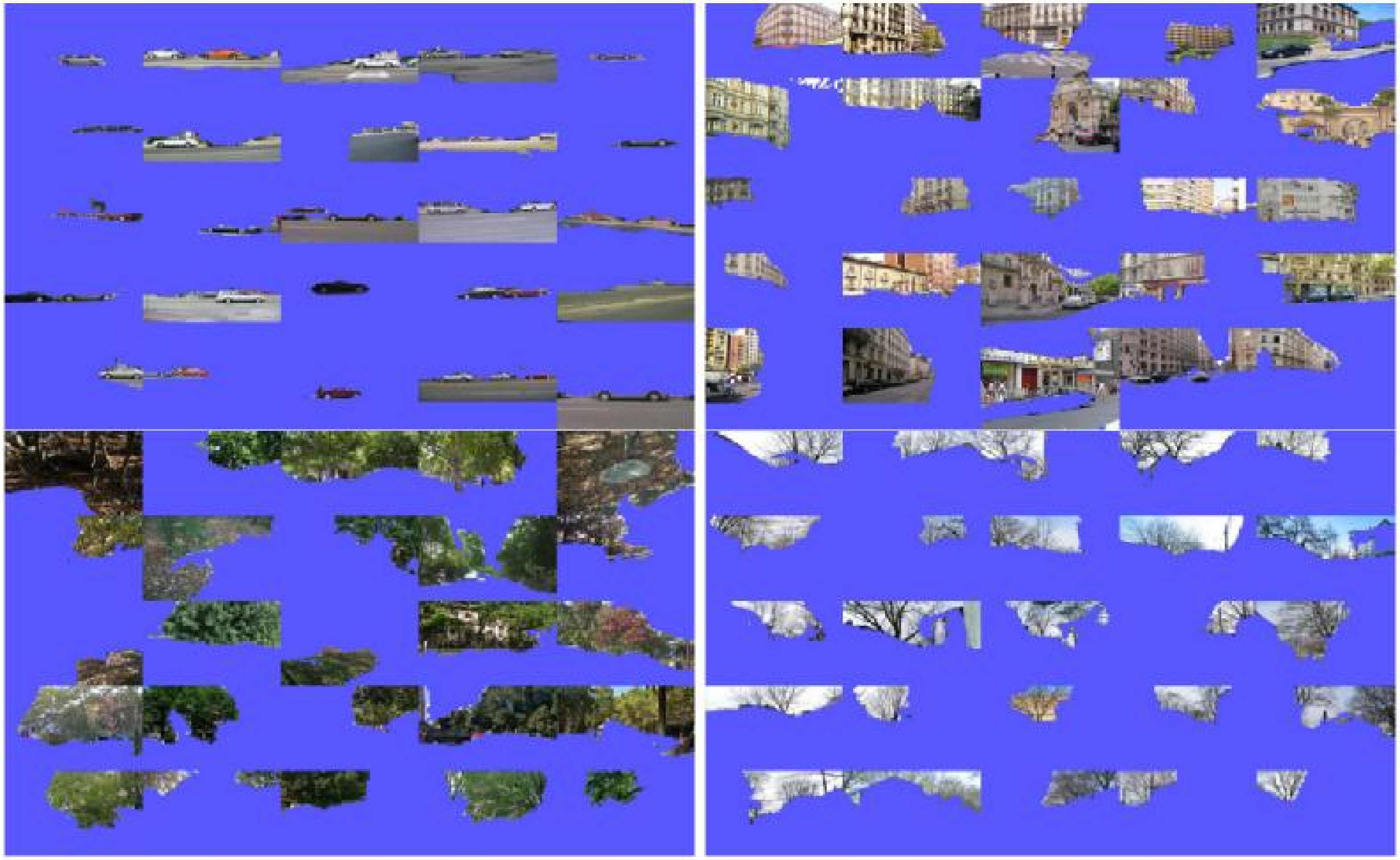


Microsoft research Cambridge (MSRC) set:

25 topics, 4325 images


QUIET PLEASE
Mon-Sat 8.30 am-5 pm 1 hour No return within 1 hour
No loading Mon - Sat 7am - 10am - 7pm
- překlady - tlumočení - korektury
The of Islay Life
PRODEJ MORAVSKÝCH SUDOVÝCH VÍN
PLEASE DO NOT LEAVE BICYCLES IN THIS AREA
Resident permit holders Mon - Sat 9am - 6pm
Resident permit holders only
No loading at any time
PAY AND DISPLAY
NO VEHICULAR ACCESS BEYOND THIS POINT
Automatic door
Humped Zebra Crossing
JEN PRO BUS S PRŮVODCÍ ČR S LICENCÍ
ALL DELIV TO FRONT OFFICE
Parking for permit holders only
LARGE VEHICLES
Except cycles
otorcycles only

Dataset: LabelMe 20 topics, 1554 images



References

Russell, B.C.; Freeman, W.T.; Efros, A.A.; Sivic, J.; Zisserman, A., "Using Multiple Segmentations to Discover Objects and their Extent in Image Collections," *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* , vol.2, no.pp. 1605- 1614, 2006

D. Blei, A. Ng, and M. Jordan. *Latent Dirichlet allocation*. Journal of Machine Learning Research, January 2003

T. Hofmann. Probabilistic latent semantic indexing. In 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, CA, USA, 1999.

The images in this presentation were taken from the CVPR06 poster “Using Multiple Segmentations to Discover Objects and their Extent in Image Collections”