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

Bryan C. Russell[1], Alexei A. Efros[2], Josef Sivic[3], William T. Freeman[1] and Andrew Zisserman[3]
[1]MIT [2]Carnegie Mellon University [3]Oxford University

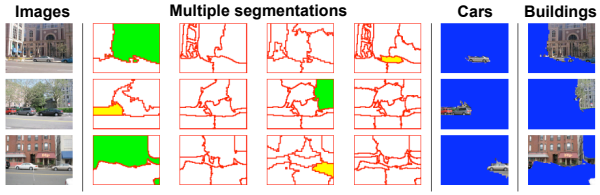
Introduction

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

Approach: 1) Produce multiple segmentations of each image

2) Discover clusters of similar segments

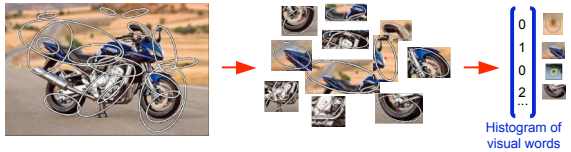
3) Score all segments by how well they fit object cluster



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*

Background: Bag-of-words Approaches

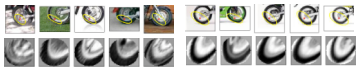
Represent an image as a histogram of "visual words"



- 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

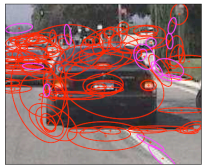
Mikolajczyk and Schmid'02, Schaffalitzky and Zisserman'02, Matas et al. '02, Lowe'99, Sivic and Zisserman'03

Visual word shortcomings



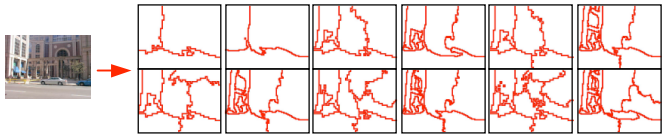
Visual Synonyms: Two different visual words representing a similar part of an object (wheel of a motorbike).

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



Lack of hard segmentation

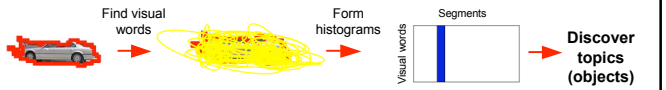
Multiple segmentations



We use Normalized Cuts, varying parameter settings: # segments and image scale

Discovering Objects

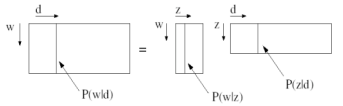
Representing Segments:



Finding coherent segment clusters (topics):

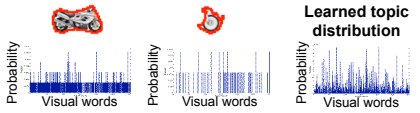
w … visual words d … documents (images) z … topics ('objects')

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

Segment scoring

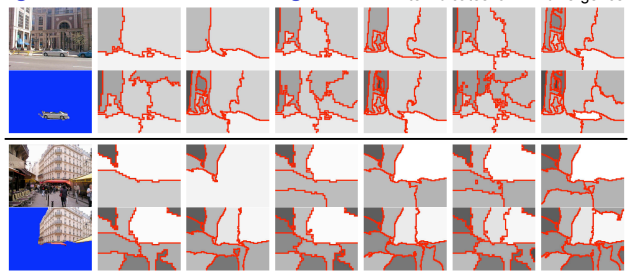


KL divergence: 1.89 KL divergence: 2.90

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

Segmentations and their KL divergence

White indicates low KL divergence



Retrieval accuracy

Average precision for MSRC

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

Segmentation accuracy

Average overlap area score for LabelMe

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

Overlap score ratio: Intersection / Union

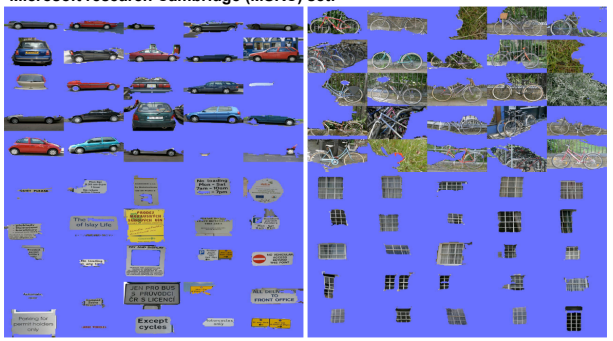
Results

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



LabelMe: 20 topics, 1554 images