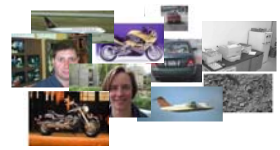
Using Multiple Segmentations to Discover Objects and their Extent in Image Collections

Bryan C. Russell1, Alexei A. Efros2, Josef Sivic3, William T. Freeman1 and Andrew Zisserman3

Presented by Priyanka Godbole
4/10/2007

Goal

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



Approach

The multiple segmentation approach to automatically discover objects and thus the visual category uses a 4 step algorithm:

- Normalized Cuts algorithm - compute multiple segmentations for each image.
- For each segment compute a histogram of visual words.
- Perform topic discovery on all segments, treating each segment as a document.
- Sort the segments based on similarity between visual words of segment and topic.

Why multiple Segmentation ?

- Choice of a seg. algo. NOT CRITICAL
- We do not rely on FULL segmentation to be correct
- NOT STABLE as the output changes when the parameters are changed.
- So we use a multiple seg. approach

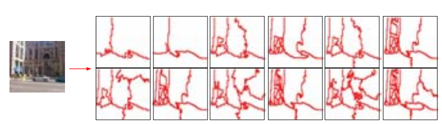
Normalized Cuts Algorithm

- It produces a **global segmentation** such that large segments could be objects.
- To produce multiple segmentations, vary 2 parameters –
  - # of segments K (K =3, 5, 7, 9)
  - Size of input image (2 image scales – 50 and 100 pixels across)
  - For LabelMe dataset, K= 11, 13 also used
  - For MSRC dataset, image scale = 150 pixel across also used.

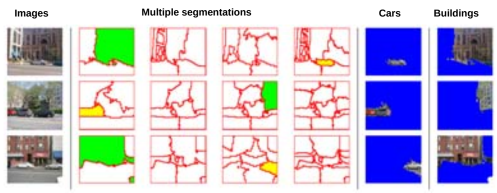
Multiple segmentations

We use Normalized Cuts, varying parameter settings:

segments and image scale



Intuitions



- 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*

Obtaining visual words

- Due to imperfection in segmentation → representation is tolerant to partial occlusion and clutter

Bag-of-words Approaches

Represent 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

Topic discovery

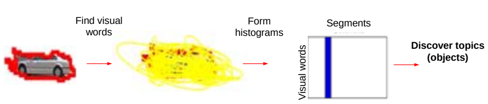
It partitions the segmented objects into visual object classes.

It uses:

- probabilistic Latent Semantic Analysis (pLSA)
- Latent Dirichlet Allocation (LDA).
- It uses unordered "bag of words" representation of documents.

Topic Discovery

- Representing Segments

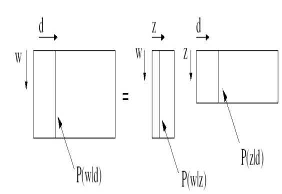


Topic Discovery

- Finding coherent segment clusters (topics):

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

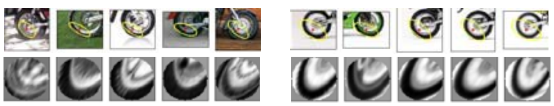
Use statistical text analysis techniques such as Latent Semantic Analysis (LSA), Probabilistic LSA [Hofmann ’99] or Latent Dirichlet Allocation (LDA) [Blei et al. ’03]. Here we chose LDA.



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

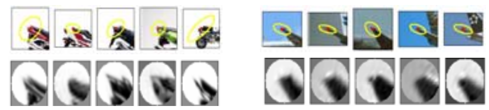
Visual word shortcomings - 1

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



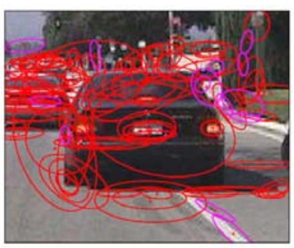
Visual word shortcomings - 2

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

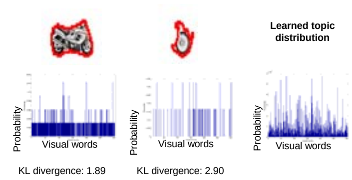


Visual word shortcomings - 3

- Lack of hard segmentation

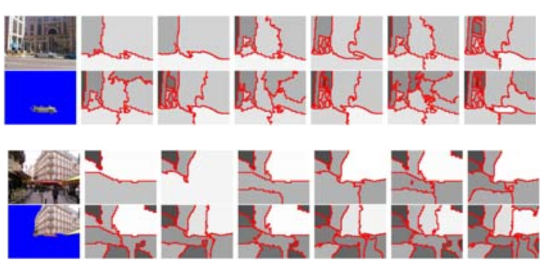


Segment Scoring



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

Segments and their KL divergence



White indicates low KL divergence

Results

Retrieval Accuracy: Average precision for MSRC

- For bicycles and windows the method performs on par or better than the other methods.
- Precision recall curves are evaluated and average precision is reported.

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

Results

Segmentation Accuracy: Average overlap area score for Label Me

The LabelMe dataset is more difficult as the images are taken in the natural habitat.

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

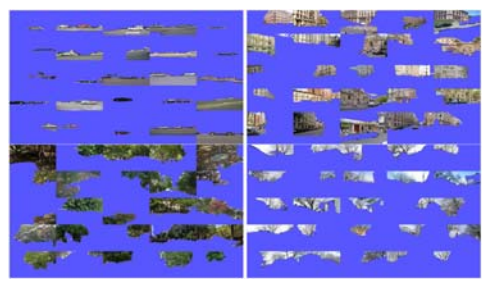
Results: Top segments Montages

Caltech 5 10 topics, 4090 images

Results: Top segments Montages

MSRC Set 25 topics, 4325 images

Results: Top segments Montages



Label Me 20 topics, 1554 images

Thank You