Lecture 21
Bayes
Project presentations

May 5  1pm – 2:30pm
2:30pm – 4pm
Complex motion
Dynamic Textures

GIANFRANCO DORETTO
Computer Science Department, University of California, Los Angeles, CA 90095
doretto@cs.ucla.edu

ALESSANDRO CHIUSO
Dipartimento di Ingegneria dell’Informazione, Università di Padova, Italy 35131
chiuso@dei.unipd.it

YING NIAN WU
Statistics Department, University of California, Los Angeles, CA 90095
ywu@stat.ucla.edu

STEFANO SOATTO
Computer Science Department, University of California, Los Angeles, CA 90095
soatto@ucla.edu

Received May 1, 2001; Revised February 21, 2002; Accepted July 3, 2002

\[ \begin{align*}
    x(t + 1) &= Ax(t) + v(t) & v(t) &\sim \mathcal{N}(0, Q) ; & x(0) &= x_0 \\
    y(t) &= Cx(t) + w(t) & w(t) &\sim \mathcal{N}(0, R)
\end{align*} \]
MRF nodes as patches

\[ \Phi(x_i, y_i) \]

\[ \Psi(x_i, x_j) \]
Network joint probability

\[ P(x, y) = \frac{1}{Z} \prod_{i, j} \Psi(x_i, x_j) \prod_i \Phi(x_i, y_i) \]
In order to use MRFs:

• Given observations $y$, and the parameters of the MRF, how infer the hidden variables, $x$?
• How learn the parameters of the MRF?
Derivation of belief propagation

minimum mean square error (MMSE)

\[ x_{1MMSE} = \text{mean}_{x_1} \text{sum}_{x_2} \text{sum}_{x_3} P(x_1, x_2, x_3, y_1, y_2, y_3) \]
The posterior factorizes

\[ x_{1\text{MMSE}} = \underbrace{\text{mean}_{x_1}}_{\text{sum}_{x_2}} \underbrace{\text{sum}_{x_3}}_{P(x_1, x_2, x_3, y_1, y_2, y_3)} \]

\[ \gamma = \underbrace{\text{mean}_{x_1}}_{\text{sum}_{x_2}} \underbrace{\text{sum}_{x_3}}_{\Phi(x_1, y_1)} \]

\[ \Phi(x_2, y_2) \Psi(x_1, x_2) \]

\[ \Phi(x_3, y_3) \Psi(x_2, x_3) \]
Propagation rules

\[ x_{1_{MMSE}} = \text{mean}_{x_1} \sum_{x_2} \sum_{x_3} P(x_1, x_2, x_3, y_1, y_2, y_3) \]

\[ x_{1_{MMSE}} = \text{mean}_{x_1} \sum_{x_2} \sum_{x_3} \Phi(x_1, y_1) \]

\[ \Phi(x_2, y_2) \Psi(x_1, x_2) \]

\[ \Phi(x_3, y_3) \Psi(x_2, x_3) \]

\[ x_{1_{MMSE}} = \text{mean}_{x_1} \Phi(x_1, y_1) \]

\[ \Phi(x_2, y_2) \Psi(x_1, x_2) \]

\[ \Phi(x_3, y_3) \Psi(x_2, x_3) \]
Propagation rules

\( x_{1MMSE} = \text{mean}_{x_1} \Phi(x_1, y_1) \)

\( \sum_{x_2} \Phi(x_2, y_2) \Psi(x_1, x_2) \)

\( \sum_{x_3} \Phi(x_3, y_3) \Psi(x_2, x_3) \)

\( M_1^2(x_1) = \sum_{x_2} \Psi(x_1, x_2) \Phi(x_2, y_2) M_2^3(x_2) \)
Propagation rules

\[ x_{1MMSE} = \text{mean}_{x_1} \Phi(x_1, y_1) \]

\[ \text{sum}_{x_2} \Phi(x_2, y_2) \Psi(x_1, x_2) \]

\[ \text{sum}_{x_3} \Phi(x_3, y_3) \Psi(x_2, x_3) \]

\[ M^2_1(x_1) = \text{sum}_{x_2} \Psi(x_1, x_2) \Phi(x_2, y_2) M^3_2(x_2) \]
Belief Propagation

**BELIEFS:** Approximate posterior marginal distributions

\[
\hat{p}(x_i \mid y) \propto \psi_i(x_i, y) \prod_{k \in \Gamma(i)} m_{ki}(x_i)
\]

\[\Gamma(i) \rightarrow \text{neighborhood of node } i\]

**MESSAGES:** Approximate sufficient statistics

\[
m_{ij}(x_j) \propto \int_{x_i} \psi_{ji}(x_j, x_i)\psi_i(x_i, y) \prod_{k \in \Gamma(i) \setminus j} m_{ki}(x_i) \, dx_i
\]

I. Belief Update (Message Product)

II. Message Propagation (Convolution)
Belief, and message updates

\[ b_j(x_j) = \prod_{k \in \mathcal{N}(j)} M^k_j(x_j) \]

\[ M^j_i(x_i) = \sum_{x_j} \psi_{ij}(x_i, x_j) \prod_{k \in \mathcal{N}(j) \setminus i} M^k_j(x_j) \]
Justifications for BP

• Gives exact marginals for trees
  → *Optimal estimates*
  → *Confidence measures*

• For general graphs, *loopy BP* has excellent empirical performance in many applications

• Recent theory provides some guarantees:
  • Statistical physics: *variational method*  
    *(Yedidia, Freeman, & Weiss)*
  • BP as reparameterization: *error bounds*  
    *(Wainwright, Jaakkola, & Willsky)*
  • Many others…
Belief propagation: the nosey neighbor rule

“Given everything that I know, here’s what I think you should think”

(Given the probabilities of my being in different states, and how my states relate to your states, here’s what I think the probabilities of your states should be)
No factorization with loops!

\[ x_{1_{MMSE}} = \text{mean}_{x_1} \Phi(x_1, y_1) \]

\[ \text{sum}_{x_2} \Phi(x_2, y_2) \Psi(x_1, x_2) \]

\[ \text{sum}_{x_3} \Phi(x_3, y_3) \Psi(x_2, x_3) \Psi(x_1, x_3) \]
References on BP and GBP

• J. Pearl, 1985
  – classic
• Y. Weiss, NIPS 1998
  – Inspires application of BP to vision
• W. Freeman et al learning low-level vision, IJCV 1999
  – Applications in super-resolution, motion, shading/paint discrimination
• H. Shum et al, ECCV 2002
  – Application to stereo
• M. Wainwright, T. Jaakkola, A. Willsky
  – Reparameterization version
• J. Yedidia, AAAI 2000
  – The clearest place to read about BP and GBP.
Interpreting images by propagating Bayesian beliefs

Yair Weiss
Dept. of Brain and Cognitive Sciences
Massachusetts Institute of Technology
E10-120, Cambridge, MA 02139, USA

In this paper we show that an architecture in which Bayesian Beliefs about image properties are propagated between neighboring units yields convergence times which are several orders of magnitude faster than traditional methods and avoids local minima. In particular our architecture is non-iterative in the sense of Marr [5]: at every time step, the local estimates at a given location are optimal given the information which has already been propagated to that location. We illustrate the algorithm’s performance on real images and compare it to several existing methods.
\[ J(Y) = \sum_k w_k (y_k - y_k^*)^2 + \lambda \sum_i (y_i - y_{i+1})^2 \]

Figure 4: a. Local estimate of DOF along the contour. b. Performance of Hopfield gradient descent, relaxation labeling and BBP as a function of time. BBP is the only method that converges to the global minimum. c. DOF estimate of Hopfield net after convergence. d. DOF estimate of BBP after convergence.
Random Fields for segmentation

$I = \text{Image pixels (observed)}$
$h = \text{foreground/background labels (hidden) – one label per pixel}$
$\theta = \text{Parameters}$

$$p(h \mid I, \theta)$$

Posterior

1. Generative approach models joint
   $\rightarrow \text{Markov random field (MRF)}$

2. Discriminative approach models posterior directly
   $\rightarrow \text{Conditional random field (CRF)}$
Generative Markov Random Field

\[ p(h, I | \theta) = \frac{p(I | h, \theta) p(h | \theta)}{Z(\theta)} \]

\[ = \frac{1}{Z(\theta)} \left[ \prod_i \phi_i(I | h_i, \theta_i) \prod_{ij} \psi_{ij}(h_i, h_j | \theta_{ij}) \right] \]

- Likelihood
- MRF Prior

\( h \) (labels) \( \in \{ \text{foreground}, \text{background} \} \)

\( I \) (pixels)

Image Plane

Prior has no dependency on \( I \)
Conditional Random Field

Discriminative approach

\[
p(h \mid I, \theta) = \frac{1}{Z(I, \theta)} \left[ \prod_i \phi_i(h_i, I \mid \theta_i) \prod_{ij} \psi_{ij}(h_i, h_j, I \mid \theta_{ij}) \right]
\]

- Dependency on I allows introduction of pairwise terms that make use of image.

- For example, neighboring labels should be similar only if pixel colors are similar \(\rightarrow\) Contrast term

\[e.g\, \text{Kumar and Hebert 2003}\]
$p(h | \Omega, I, \theta) \propto \prod_i \phi_i^1(I | h_i, \theta_i) \phi_i^2(h_i | \Omega) \prod_{ij} \psi_{ij}^1(h_i, h_j | \theta_{ij}) \cdot \psi_{ij}^2(I | h_i, h_j, \theta_{ij})$

- $\Omega$ is a shape prior on the labels from a Layered Pictorial Structure (LPS) model

- Segmentation by:
  - Match LPS model to image (get number of samples, each with a different pose)
  - Marginalize over the samples using a single graph cut [Boykov & Jolly, 2001]
OBJCUT:
Shape prior - $\Omega$ - Layered Pictorial Structures (LPS)

- Generative model
- Composition of parts + spatial layout

Parts in Layer 2 can occlude parts in Layer 1

Kumar, et al. 2004, 2005
OBJCUT: Results

Using LPS Model for Cow

In the absence of a clear boundary between object and background
Generative models

Two big families:

• Grammar based models

• Topic models

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center’s share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.
Grammars

“A common framework for visual knowledge representation and object categorization. Grammars, studied mostly in language, are known for their expressive power in generating a very large set of configurations or instances, i.e. their language, by composing a relatively much smaller set of words, i.e. shared and reusable elements, using production rules.”

A Stochastic Grammar of Images
Song-Chun Zhu and David Mumford
Object → Bag of ‘words’
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 reach the brain from our eyes. For a long time it was thought that the retinal image was transmitted point by point to visual centers in the brain; the cerebral cortex was a movie screen, so to speak, upon which the image in the eye was projected. Through the discoveries of Hubel and Wiesel we now know that behind the origin of the visual perception in the brain there is a considerably more complicated course of events. By following the visual impulses along their path to the various cell layers of the optical cortex, Hubel and Wiesel have been able to demonstrate that the message about the image falling on the retina undergoes a step-wise analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.

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% jump in exports to $750bn, compared with a 18% rise in imports to $660bn. The figures are likely to further annoy the US, which has long argued that Chinese exports are unfairly undervalued. Beijing agrees the surplus is too high, but says the yuan is only one factor. Bank of China governor Zhou Xiaochuan said the country also needed to do more to boost domestic demand so more goods stay within the country. China increased the value of the yuan against the dollar by 2.1% in July and permitted it to trade within a narrow band, but the US wants the yuan allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.
Related works

• Early “bag of words” models: mostly texture recognition

• Hierarchical Bayesian models for documents (pLSA, LDA, etc.)
  – Hoffman 1999; Blei, Ng & Jordan, 2004; Teh, Jordan, Beal & Blei, 2004

• Object categorization
  – Csurka, Bray, Dance & Fan, 2004; Sivic, Russell, Efros, Freeman & Zisserman, 2005; Sudderth, Torralba, Freeman & Willsky, 2005;

• Natural scene categorization
  – Vogel & Schiele, 2004; Fei-Fei & Perona, 2005; Bosch, Zisserman & Munoz, 2006
Hierarchical Topic Models

- Topic models typically use a “bag of words” approx.:
  - Learning topics allows transfer of information within a corpus of related documents
  - Mixing proportions capture the distinctive features of particular documents

Latent Dirichlet Allocation (LDA)

*Blei, Ng, & Jordan, JMLR 2003*
Analogy: Discovering topics in text collections

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center’s share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

<table>
<thead>
<tr>
<th>“Arts”</th>
<th>“Budgets”</th>
<th>“Children”</th>
<th>“Education”</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEW</td>
<td>MILLION</td>
<td>CHILDREN</td>
<td>SCHOOL</td>
</tr>
<tr>
<td>FILM</td>
<td>TAX</td>
<td>WOMEN</td>
<td>STUDENTS</td>
</tr>
<tr>
<td>SHOW</td>
<td>PROGRAM</td>
<td>PEOPLE</td>
<td>SCHOOLS</td>
</tr>
<tr>
<td>MUSIC</td>
<td>BUDGET</td>
<td>CHILD</td>
<td>EDUCATION</td>
</tr>
<tr>
<td>MOVIE</td>
<td>BILLION</td>
<td>YEARS</td>
<td>TEACHERS</td>
</tr>
<tr>
<td>PLAY</td>
<td>FEDERAL</td>
<td>FAMILIES</td>
<td>HIGH</td>
</tr>
<tr>
<td>MUSICAL</td>
<td>YEAR</td>
<td>WORK</td>
<td>PUBLIC</td>
</tr>
<tr>
<td>BEST</td>
<td>SPENDING</td>
<td>PARENTS</td>
<td>TEACHER</td>
</tr>
<tr>
<td>ACTOR</td>
<td>NEW</td>
<td>SAYS</td>
<td>BENNETT</td>
</tr>
<tr>
<td>FIRST</td>
<td>STATE</td>
<td>FAMILY</td>
<td>MANIGAT</td>
</tr>
<tr>
<td>YORK</td>
<td>PLAN</td>
<td>WELFARE</td>
<td>NAMPHY</td>
</tr>
<tr>
<td>OPERA</td>
<td>MONEY</td>
<td>MEN</td>
<td>STATE</td>
</tr>
<tr>
<td>THEATER</td>
<td>PROGRAMS</td>
<td>PERCENT</td>
<td>PRESIDENT</td>
</tr>
<tr>
<td>ACTRESS</td>
<td>GOVERNMENT</td>
<td>CARE</td>
<td>ELEMENTARY</td>
</tr>
<tr>
<td>LOVE</td>
<td>CONGRESS</td>
<td>LIFE</td>
<td>HAITI</td>
</tr>
</tbody>
</table>

Blei, et al. 2003
Visual analogy

document - image

word - visual word

topics - objects
2 generative models

1. Naïve Bayes classifier
   - Csurka Bray, Dance & Fan, 2004

2. Hierarchical Bayesian text models (pLSA and LDA)
   - Background: Hoffman 2001, Blei, Ng & Jordan, 2004
   - Natural scene categorization: Fei-Fei et al. 2005
First, some notations

- $w_n$: each patch in an image
  - $w_n = [0,0,…1,…,0,0]^T$
- $w$: a collection of all $N$ patches in an image
  - $w = [w_1,w_2,…,w_N]$
- $d_j$: the $j^{th}$ image in an image collection
- $c$: category of the image
- $z$: theme or topic of the patch
Documents collection

Co-occurrence table:

Number of times word i appears on document/image j
Case #1: the Naïve Bayes model

\[ c^* = \arg \max_c p(c \mid w) \propto p(c) p(w \mid c) = p(c) \prod_{n=1}^{N} p(w_n \mid c) \]

Object class decision
Prior prob. of the object classes
Image likelihood given the class

Csurka et al. 2004
Our in-house database contains 1776 images in seven classes\(^1\): faces, buildings, trees, cars, phones, bikes and books. Fig. 2 shows some examples from this dataset.

\(^1\) The exact number can vary slightly due to the nature of the dataset.
Table 1. Confusion matrix and the mean rank for the best vocabulary ($k=1000$).

<table>
<thead>
<tr>
<th>True classes</th>
<th>faces</th>
<th>buildings</th>
<th>trees</th>
<th>cars</th>
<th>phones</th>
<th>bikes</th>
<th>books</th>
</tr>
</thead>
<tbody>
<tr>
<td>faces</td>
<td>76</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>buildings</td>
<td>2</td>
<td>44</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>trees</td>
<td>3</td>
<td>2</td>
<td>80</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>cars</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>75</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>phones</td>
<td>9</td>
<td>15</td>
<td>1</td>
<td>16</td>
<td>70</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>bikes</td>
<td>2</td>
<td>15</td>
<td>12</td>
<td>0</td>
<td>8</td>
<td>73</td>
<td>0</td>
</tr>
<tr>
<td>books</td>
<td>4</td>
<td>19</td>
<td>0</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>69</td>
</tr>
</tbody>
</table>

| Mean ranks   | 1.49  | 1.88      | 1.33  | 1.33 | 1.63   | 1.57  | 1.57  |

Csurka et al. 2004
Case #2: Hierarchical Bayesian text models

Probabilistic Latent Semantic Analysis (pLSA)

Latent Dirichlet Allocation (LDA)

Hoffman, 2001

Blei et al., 2001
Case #2: Hierarchical Bayesian text models

Probabilistic Latent Semantic Analysis (pLSA)

Sivic et al. ICCV 2005
Case #2: Hierarchical Bayesian text models

Latent Dirichlet Allocation (LDA)

“beach”

Fei-Fei et al. ICCV 2005
Case #2: the pLSA model
Case #2: the pLSA model

\[ p(w_i \mid d_j) = \sum_{k=1}^{K} p(w_i \mid z_k) p(z_k \mid d_j) \]
Case #2: Recognition using pLSA

\[ z^* = \arg \max_z p(z \mid d) \]

\[ \text{Slide credit: Josef Sivic} \]
Case #2: Learning the pLSA parameters

$$L = \prod_{i=1}^{M} \prod_{j=1}^{N} P(w_i | d_j)^{n(w_i, d_j)}$$

Maximize likelihood of data using EM

M … number of codewords

N … number of images

Observed counts of word $i$ in document $j$

Slide credit: Josef Sivic
Two bag-of-words classifiers

ICCV 2005 short courses on Recognizing and Learning Object Categories

A simple approach to classifying images is to treat them as a collection of regions, describing only their appearance and ignoring their coordinates. Such methods have been successfully used in the text community for analyzing documents and are known as "bag-of-words" models, since each document is represented as a distribution over fixed vocabulary(s). Using such a representation, methods such as probabilistic latent semantic analysis (pLSA) [1] and Latent Dirichlet Allocation (LDA) [2] are able to extract coherent topics within document collections in an unsupervised manner.

Recently, Fei-Fei et al. [3] and Sivic et al. [4] have applied such methods to the visual domain. The demo code implements pLSA, and for comparison, a Naive Bayes classifier is also provided which requires labelled training data, unlike pLSA.

The code consists of Matlab scripts (which should run under both Windows and Linux) and a couple of 32-bit Linux binaries for doing the actual classification. Hence the whole system will need to be run on Linux. The code is for teaching/research purposes only. If you find a bug please point it out to us.

Download

Download the code and datasets (32 Mb bytes)

Operation of code

To run the demo,
From Images to Features

- Pixels are very sensitive to changes in lighting & pose
- Instead represent image as affine covariant regions:
  - Harris affine invariant regions (corners & edges)
  - Maximally stable extremal regions (segmentation)

Software provided by Oxford Visual Geometry Group
Sample Detected Features
Describing Feature Appearance

- **SIFT**: Scale Invariant Feature Transform
- Normalized histogram of orientation energy in each affinely adapted region (128-dim.)
A Discrete Feature Vocabulary

- Using all training images, build a dictionary via K-means clustering (~1000 words)
- Map each SIFT descriptor to nearest word

\[ w_{ji} \rightarrow \text{appearance of feature } i \text{ in image } j \]

\[ y_{ji} \rightarrow \text{2D position of feature } i \text{ in image } j \]
Form dictionary

Build visual vocabulary by k-means clustering
SIFT descriptors (K~2,000)
Example regions assigned to the same dictionary cluster

Cluster 1

Cluster 2

Slide credit: Bryan Russell & Josef Sivic
Polysemy

In English, “bank” refers to:
1. a institution that handle money
2. the side of a river

Regions that map to the same visual word:
Representing an image with visual words

Sivic & Zisserman '03

Interest regions

Visual words

Slide credit: Bryan Russell & Josef Sivic
System overview

Input image

Compute visual words

Discover visual topics

Slide credit: Bryan Russell & Josef Sivic
Bag of words

Interest regions

Stack visual word histograms as columns in matrix

Throw away spatial information!

Slide credit: Bryan Russell & Josef Sivic
Latent Dirichlet Allocation (LDA)

Blei, et al. 2003

- LDA model assumes exchangeability
- Order of words does not matter

\[ w_{ij} \] - words

\[ z_{ij} \] - topic assignments

\[ \mu_i \] - topic mixing weights

\[ \Phi_k \] - word mixing weights

\[ p(w_{ij}) \propto \sum_{k=1}^{K} p(w_{ij} | z_{ij} = k, \phi_k) p(z_{ij} = k | \theta_i) \]

Slide credit: Bryan Russell & Josef Sivic
Inference

\[ w_{ij} \] - words

\[ z_{ij} \] - topic assignments

\[ \mu_i \] - topic mixing weights

\[ \hat{\alpha}_k \] - word mixing weights

Use Gibbs sampler to sample topic assignments

\[
z_{ij} \sim p(z_{ij} = k | w_{ij} = v, w_{\setminus(ij)}, z_{\setminus(ij)}, \alpha, \beta)
\]

- Only need to maintain counts of topic assignments
- Sampler typically converges in less than 50 iterations
- Run time is less than an hour

[Griffiths & Steyvers 2004]
Apply to Caltech 4 + background images

Faces 435
Motorbikes 800
Airplanes 800
Cars (rear) 1155
Background 900
Total: 4090

Slide credit: Bryan Russell & Josef Sivic
Most likely words given topic

Topic 1

Word 1

Word 2

Topic 2

Word 1

Word 2

Slide credit: Bryan Russell & Josef Sivic
Most likely words given topic

Topic 3

Word 1

Word 2

Topic 4

Word 1

Word 2

Slide credit: Bryan Russell & Josef Sivic
\[ p(w_{ij} | d_i) \]

\[ p(z_{ij} | d_i) \]

\[ w \rightarrow p(w_{ij} | d_i) \]

\[ z \rightarrow \sim \]

\[ w \downarrow \]

\[ p(w_{ij} | z_{ij}) \]

\[ z \downarrow \]

\[ d \rightarrow \]

\[ p(z_{ij} | d_i) \]

Slide credit: Bryan Russell & Josef Sivic
Image clustering

Confusion matrices:

Average confusion:

<table>
<thead>
<tr>
<th>Expt.</th>
<th>Categories</th>
<th>T</th>
<th>LDA</th>
<th>pLSA</th>
<th>KM baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>(1)</td>
<td>4</td>
<td>4</td>
<td>97</td>
<td>98</td>
<td>72</td>
</tr>
<tr>
<td>(2)</td>
<td>4 + bg</td>
<td>5</td>
<td>78</td>
<td>78</td>
<td>56</td>
</tr>
<tr>
<td>(2)*</td>
<td>4 + bg</td>
<td>6</td>
<td>84</td>
<td>76</td>
<td>–</td>
</tr>
<tr>
<td>(2)*</td>
<td>4 + bg</td>
<td>7</td>
<td>78</td>
<td>83</td>
<td>–</td>
</tr>
<tr>
<td>(2)*</td>
<td>4 + bg-fxd</td>
<td>7</td>
<td>90</td>
<td>93</td>
<td>–</td>
</tr>
</tbody>
</table>
Image as a mixture of topics (objects)
Hierarchical DP Object Model

Slide credit: Erik Sudderth
Scenes of Fixed Sets of Objects

Pr(object | scene) → β → ζ → ψ

Assumes a fixed number of object instances

Pr(part | object) → o → z → w → y

Reference positions (ONE PER OBJECT)

covariance

context

Slide credit: Erik Sudderth
Street Scene Segmentations
TDP for 3D Scenes

**Global Density**
- Object category
- Part size & shape
- Transformation prior

**Transformed Densities**
- Object category
- Part size & shape
- Transformed locations

**3D Scene Features**
- Object category
- 3D Location

**2D Image Features**
- Appearance Descriptors
- 2D Pixel Coordinates
Single-Part Office Scene Model

Background  Bookshelves  Computer Screen  Desk

Slide credit: Erik Sudderth