Real-Time Computerized Annotation of Pictures

Jia Li James Z. Wang

The Pennsylvania State University

Email: jiali@psu.edu, jwang@ist.psu.edu

How Visible Are Web Images?

Keukenhof photos




ALIPR: Automatic Linguistic Indexing for Pictures - Real Time

plant, flower,

landscape,

people, tulip

tree, plant,

people,

water, garden

flower,

plant,lake,

rural,

building

animal,

people,

wild-life, dog,

landscape

Architecture for Training

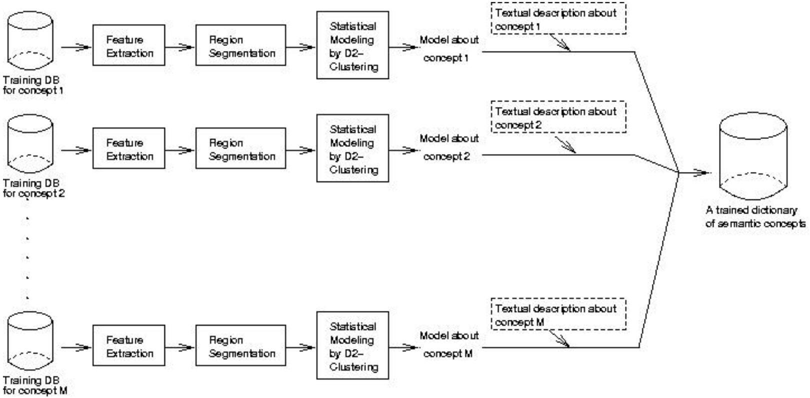
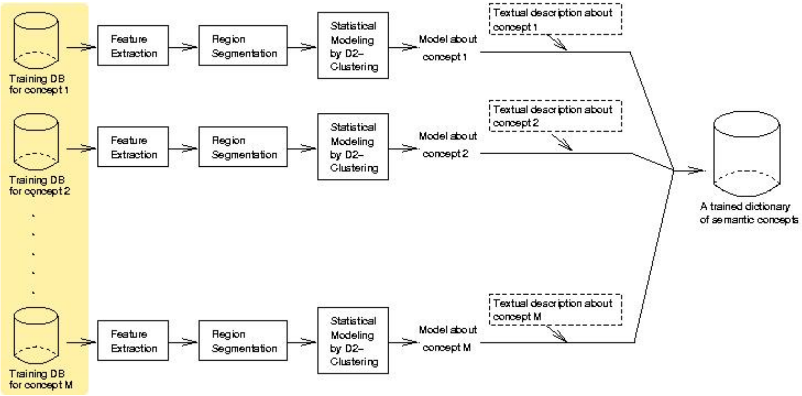


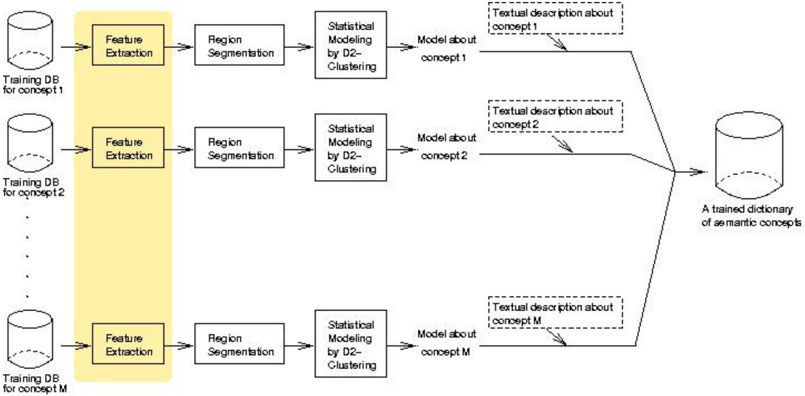
Image “Knowledge Base”



Six Hundred Semantic Categories

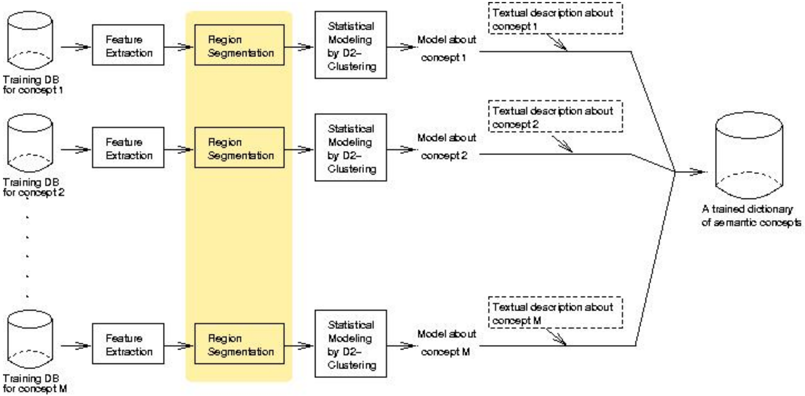
- Corel image database
  - 80 images per category.
  - Each category is described by several words: `‘‘autumn, tree, landscape, lake’’`.
  - A total of 332 distinct words.

Feature Extraction



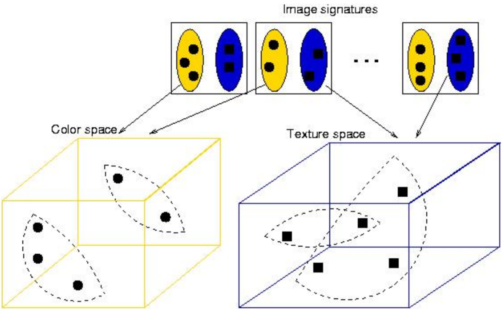
- Color components: LUV
- Texture features: wavelet coefficients

Region Segmentation and Signature Formulation

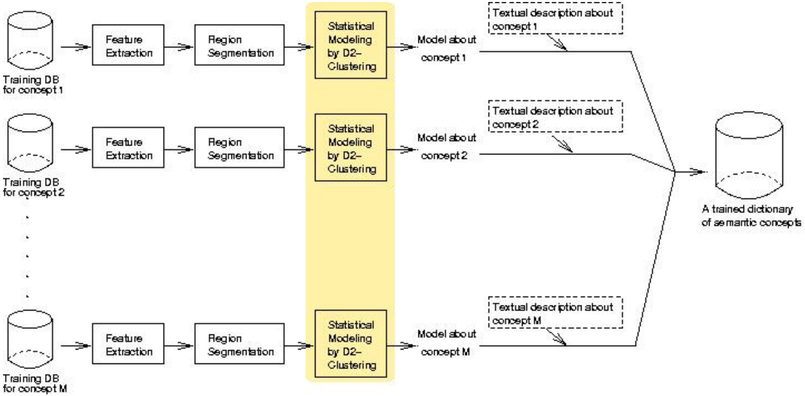


Region Segmentation and Signature Formulation

- An image signature resides in $\Omega = \Omega_1 \times \Omega_2$.
- Color distribution: $\beta_{i,1} \in \Omega_1$.
- Texture distribution: $\beta_{i,2} \in \Omega_2$.
- $\beta_{i,j} = \{(v_{i,j}^{(1)}, p_{i,j}^{(1)}), ..., (v_{i,j}^{(m_{i,j})}, p_{i,j}^{(m_{i,j})})\}$.

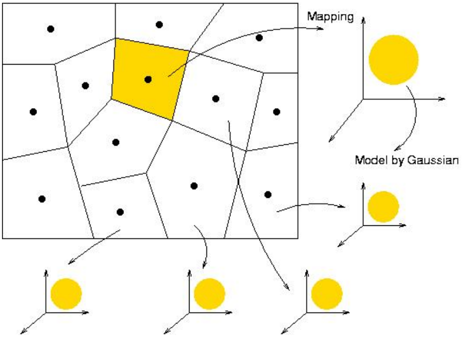


Profiling Image Concepts via Mixture Models

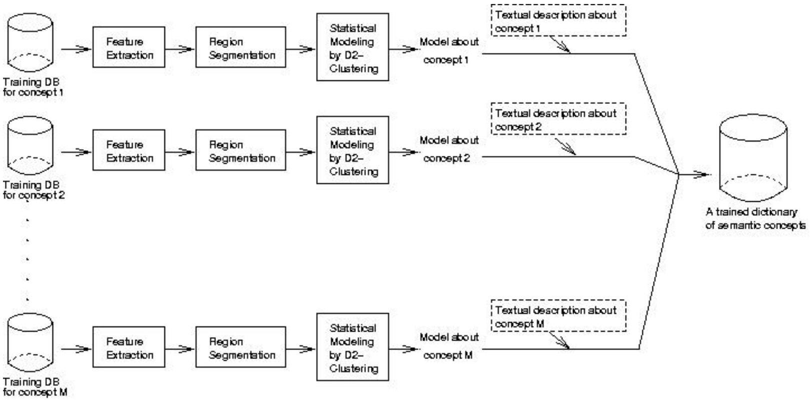


Mixture Modeling via Local Mapping

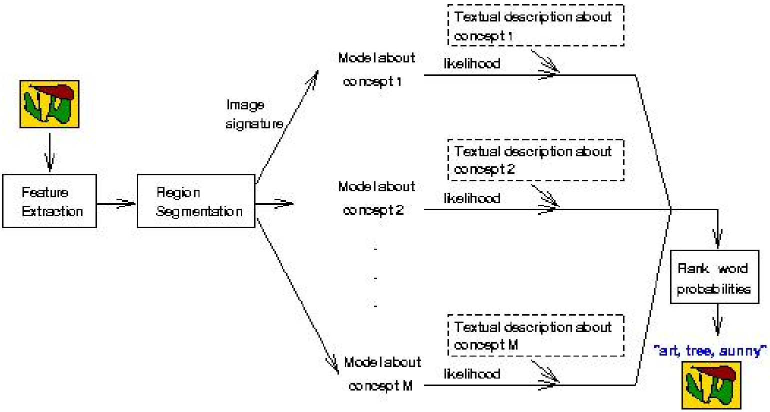
- Mixture modeling for space Ω
  - Carve Ω into cells by clustering.
  - Map each cell to an Euclidean space, preserving pairwise distances.
  - Model the mapped points by Gaussian.
- Images: a grid of feature vectors
  - Gaussian mixture
  - 2-D HMM



Architecture for Training



Architecture for Annotation



Word Probabilities

- Total word list: $\mathcal{W} = \{w_1, w_2, ..., w_K\}$.
- Semantic categories containing word w_i: $\mathcal{C}(w_i)$.
- Model of category m: $\mathcal{M}_m$, $m = 1, ..., M$.
- Prior on categories: ρ_m (set uniform).

Category prob. given signature β

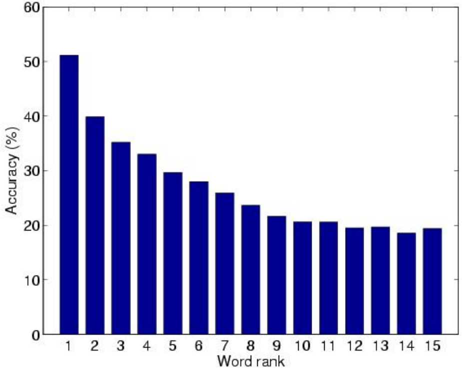
$$p_m(\beta) = \frac{\rho_m f(\beta \mid \mathcal{M}_m)}{\sum_{l=1}^{M} \rho_l f(\beta \mid \mathcal{M}_l)}$$

Word probability

$$q(\beta, w_i) = \sum_{m: m \in \mathcal{C}(w_i)} p_m(\beta) \ .$$

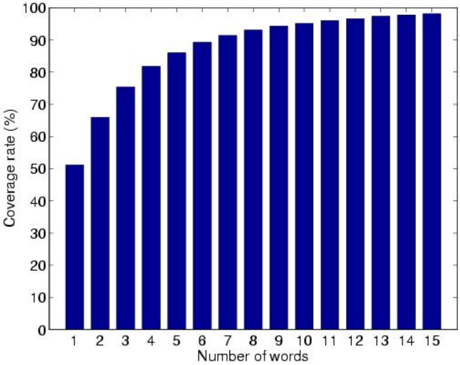
Human Evaluation on flickr.com Images

- Manual evaluation on 5, 411 flickr.com images.
- Accuracy of the first word: 51.17%.



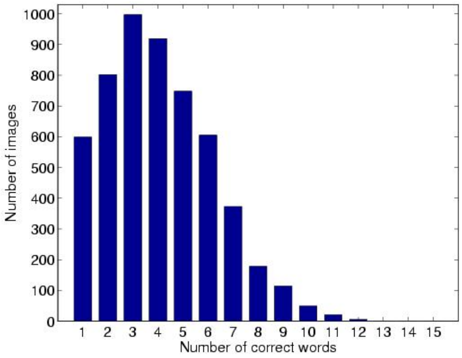
Human Evaluation on flickr.com Images

- Coverage rate: percentage of images correctly annotated by at least one word.
- Top 4 words: $> 80\%$.
- Top 7 words: 91.37%.
- Top 15 words: 98.13%.



Human Evaluation on flickr.com Images

- Annotate using top 15 words.
- # correct: 4.1 on average



Speed

- Training:
  - 109 seconds on ave.
  - 80 images per category
  - 2.4 GHz AMD processor
- Annotation:
  - 1.4 seconds on ave. for example images
  - 3.0 GHz Intel processor
  - Convert from JPEG to raw format; extract image signature; find annotation words.

Conclusions

System

- The ALIPR system: real-time automatic annotation of pictures
- Human evaluation on web images

Learning methodology

- D2-clustering
  - Generalized k-means to bags of weighted vectors
- Mixture modeling via mapping to conjectural space

- Human evaluation on $5,400+$ Web images has demonstrated promising results.
- Future work: bridge with retrieval, incremental learning, improve modeling, Web applications ...
- ALIPR your pictures: http://alipr.com

ALIPR Computerized Annotation

Please help us to train ALIPR. Check those correctly annotated words.

- [ ] indoor
- [ ] animal
- [ ] food
- [ ] drawing
- [ ] fruit
- [ ] art
- [ ] man-made
- [ ] flower
- [ ] dog
- [ ] pet
- [ ] ancestor
- [ ] drink
- [ ] antique
- [ ] dinosaur
- [ ] poster

Any tags missing by ALIPR? (Add tags here – separate with ',')

submit