

Real-Time Computerized Annotation of Pictures

Jia Li

James Z. Wang

The Pennsylvania State University

Email: jjali@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

Overview

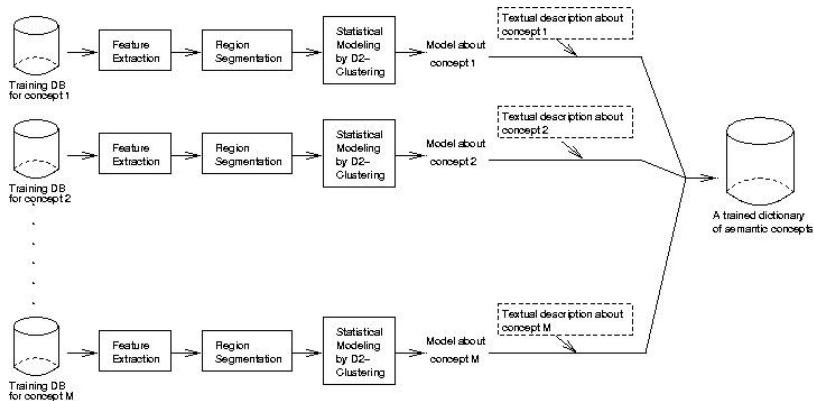
System

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

Learning methodology

- ▶ *D2-clustering*
 - ▶ Bags of weighted vectors— discrete distributions with variable supports
- ▶ Mixture modeling via mapping to conjectural space

Architecture for Training



Architecture for Annotation

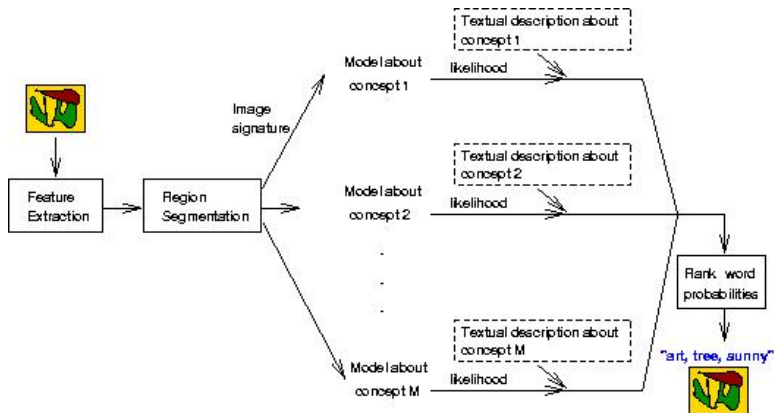
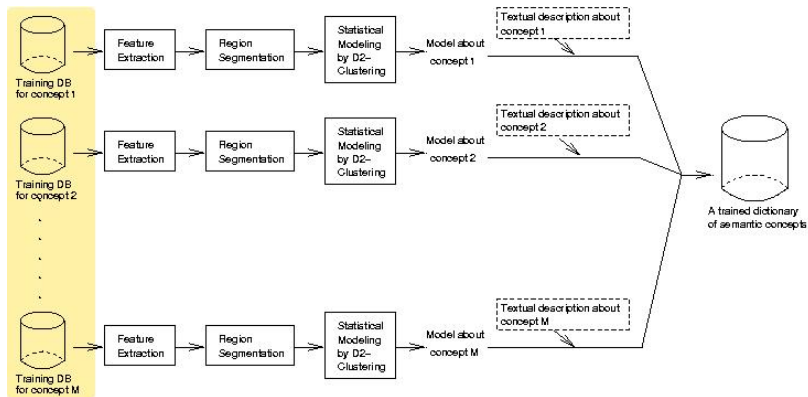
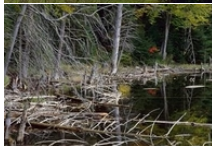


Image “Knowledge Base”

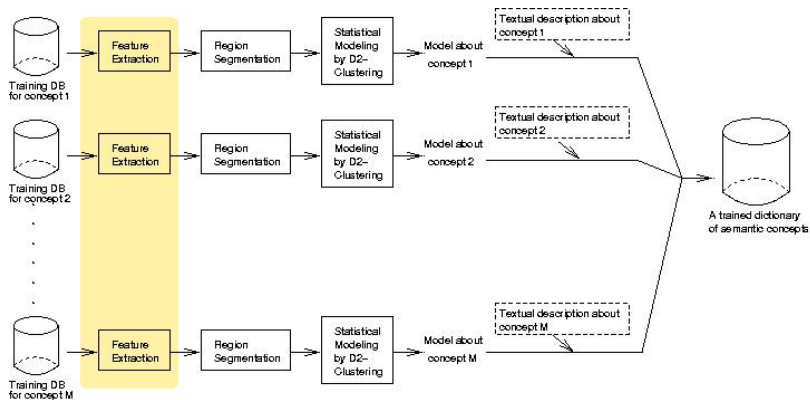


Six Hundred Semantic Categories

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

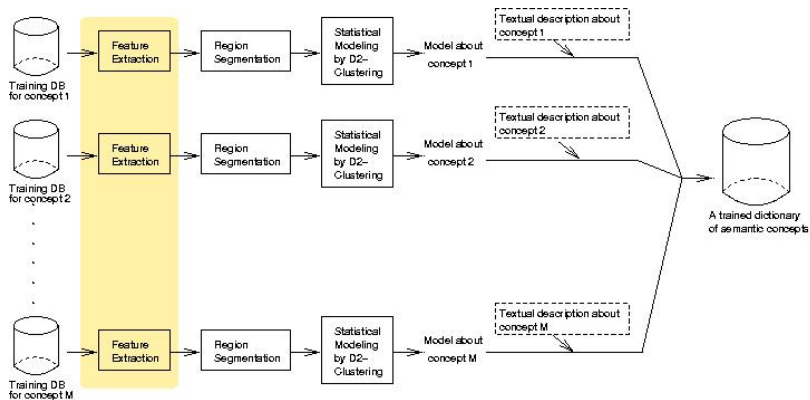


Feature Extraction



- ▶ Color components: LUV
- ▶ Texture features: wavelet coefficients

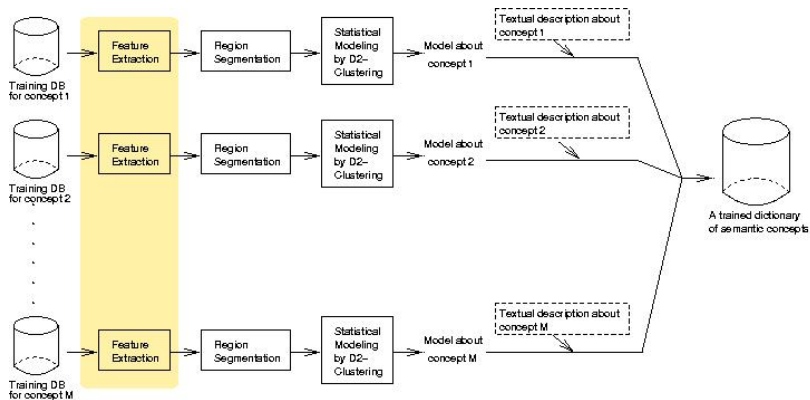
Feature Extraction



► Color components: LUV

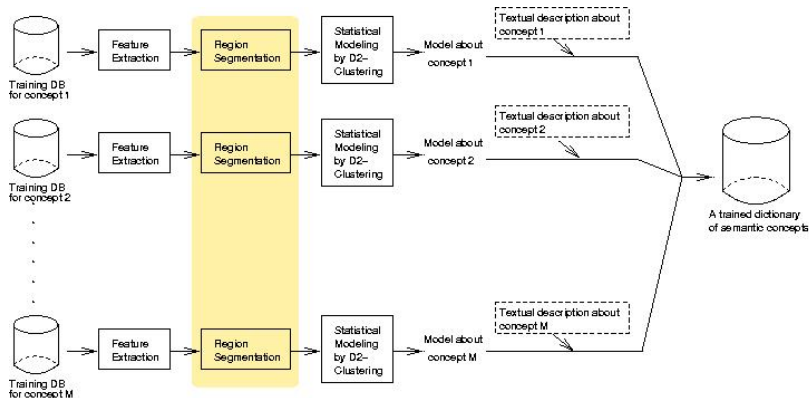
► Texture features: wavelet coefficients

Feature Extraction



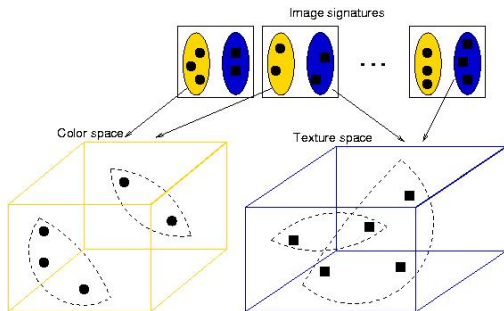
- ▶ Color components: LUV
- ▶ Texture features: wavelet coefficients

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: $\beta_{i,2} \in \Omega_2$.
- ▶ $\beta_{i,j} = \{(v_{i,j}^{(1)}, p_{i,j}^{(1)}), \dots, (v_{i,j}^{(m_{i,j})}, p_{i,j}^{(m_{i,j})})\}$.



Mallows Distance between Distributions

- ▶ Let the two discrete distributions be

$$\gamma_i = \{(z_i^{(1)}, q_i^{(1)}), (z_i^{(2)}, q_i^{(2)}), \dots, (z_i^{(m_i)}, q_i^{(m_i)})\}, i = 1, 2$$

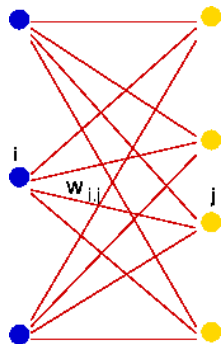
- ▶ The Mallows distance $D(\gamma_1, \gamma_2)$ is defined by

$$D^2(\gamma_1, \gamma_2) = \min_{\{w_{i,j}\}} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} w_{i,j} \|z_1^{(i)} - z_2^{(j)}\|^2$$

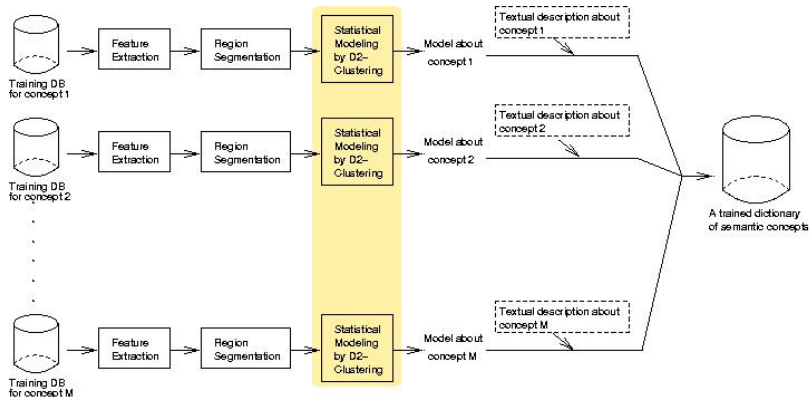
subject to $\sum_{j=1}^{m_2} w_{i,j} = q_1^{(i)}, i = 1, \dots, m_1,$

$\sum_{i=1}^{m_1} w_{i,j} = q_2^{(j)}, j = 1, \dots, m_2, w_{i,j} \geq 0, i = 1, \dots, m_1,$
 $j = 1, \dots, m_2.$

- ▶ Solved by linear programming.



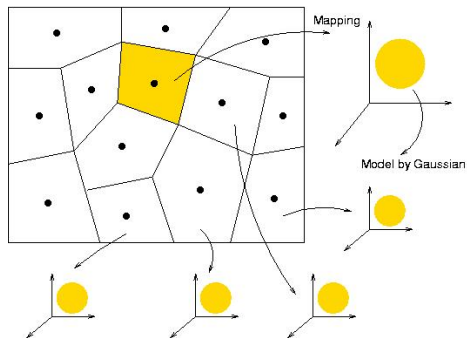
Profiling Image Concepts via Mixture Models



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



D2-Clustering

- ▶ An image set
 $B = \{\beta_i : \beta_i \in \Omega, i = 1, \dots, n\}$,
 $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_d$
- ▶ Distance between arrays of discrete distributions:

$$\tilde{D}(\beta_i, \beta_j) \triangleq \sum_{l=1}^d D^2(\beta_{i,l}, \beta_{j,l})$$

Optimize

- ▶ A set of prototypes
 $A = \{\alpha_i : \alpha_i \in \Omega, i = 1, \dots, \bar{m}\}$.
- ▶ Prototype assignment function
 $c(i) \in \{1, 2, \dots, \bar{m}\}, i = 1, \dots, n$.

D2-Clustering

Optimization Criterion

$$L(B, A^*, c^*) = \min_A \min_c \sum_{i=1}^n \tilde{D}(\beta_i, \alpha_{c(i)})$$

- ▶ K-means \longrightarrow D2-Clustering
(vectors) \longrightarrow (bags of weighted vectors)

D2-Clustering

Algorithm

1. For every image i , set

$$c(i) = \arg \min_{j=1, \dots, \bar{m}} \tilde{D}(\beta_i, \alpha_j).$$

2. Let $\mathcal{C}_j = \{i : c(i) = j\}$, $j = 1, \dots, \bar{m}$. That is, \mathcal{C}_j contains indices of images assigned to prototype j .

For each prototype j , set

$$\alpha_j = \arg \min_{\alpha \in \Omega} \sum_{i \in \mathcal{C}_j} \tilde{D}(\beta_i, \alpha) \quad \leftarrow \textit{challenging}$$

Update of Prototypes

1. For every image i , set $c(i) = \arg \min_{j=1, \dots, \bar{m}} \tilde{D}(\beta_i, \alpha_j)$.
2. Let $\mathcal{C}_\eta = \{i : c(i) = \eta\}$, $\eta = 1, \dots, \bar{m}$. Update each $\alpha_{\eta,l}$, $\eta = 1, \dots, \bar{m}$, $l = 1, \dots, d$, individually by the following steps. Denote $\alpha_{\eta,l} = \{(z_{\eta,l}^{(1)}, q_{\eta,l}^{(1)}), (z_{\eta,l}^{(2)}, q_{\eta,l}^{(2)}), \dots, (z_{\eta,l}^{(m'_{\eta,l})}, q_{\eta,l}^{(m'_{\eta,l})})\}$.

- 2.1 Fix $z_{\eta,l}^{(k)}$, $k = 1, \dots, m'_{\eta,l}$. Update $q_{\eta,l}^{(k)}$, $w_{k,j}^{(i)}$, $i \in \mathcal{C}_\eta$, $k = 1, \dots, m'_{\eta,l}$, $j = 1, \dots, m_{i,l}$ by solving the linear programming problem:

$$\min_{q_{\eta,l}^{(k)}} \sum_{i \in \mathcal{C}_\eta} \min_{w_{k,j}^{(i)}} \sum_{k=1}^{m'_{\eta,l}} \sum_{j=1}^{m_{i,l}} w_{k,j}^{(i)} \|z_{\eta,l}^{(k)} - v_{i,l}^{(j)}\|^2, \text{ subject to } \sum_{k=1}^{m'_{\eta,l}} q_{\eta,l}^{(k)} = 1;$$

$$q_{\eta,l}^{(k)} \geq 0, k = 1, \dots, m'_{\eta,l}; \sum_{j=1}^{m_{i,l}} w_{k,j}^{(i)} = q_{\eta,l}^{(k)}, i \in \mathcal{C}_\eta, k = 1, \dots, m'_{\eta,l}; \sum_{k=1}^{m'_{\eta,l}} w_{k,j}^{(i)} = p_{i,l}^{(j)}, i \in \mathcal{C}_\eta, j = 1, \dots, m_{i,l}; w_{k,j}^{(i)} \geq 0, i \in \mathcal{C}_\eta, k = 1, \dots, m'_{\eta,l}, j = 1, \dots, m_{i,l}.$$

- 2.2 Fix $q_{\eta,l}^{(k)}$, $w_{k,j}^{(i)}$, $i \in \mathcal{C}_\eta$, $1 \leq k \leq m'_{\eta,l}$, $1 \leq j \leq m_{i,l}$. Update $z_{\eta,l}^{(k)}$, $k = 1, \dots, m'_{\eta,l}$ by

$$z_{\eta,l}^{(k)} = \frac{\sum_{i \in \mathcal{C}_\eta} \sum_{j=1}^{m_{i,l}} w_{k,j}^{(i)} v_{i,l}^{(j)}}{\sum_{i \in \mathcal{C}_\eta} \sum_{j=1}^{m_{i,l}} w_{k,j}^{(i)}}.$$

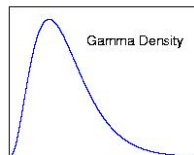
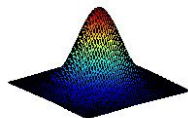
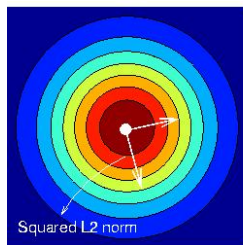
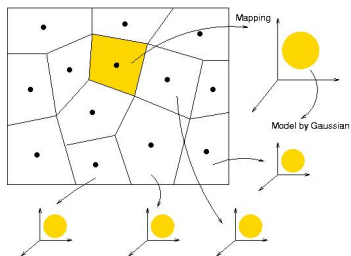
- 2.3 Compute $\sum_{k=1}^{m'_{\eta,l}} \sum_{j=1}^{m_{i,l}} w_{k,j}^{(i)} \|z_{\eta,l}^{(k)} - v_{i,l}^{(j)}\|^2$. If the rate of decrease from the previous iteration is below a threshold, go to Step 3; otherwise, go to Step 1.

3. Compute $L(B, A, c)$. If the rate of decrease from the previous iteration is below a threshold, stop; otherwise, go back to Step 1.

Mixture Modeling

Modeling

- ▶ Model by Gaussian: spherical covariance, prototype being the mean.
- ▶ $X \sim N(\mu, \sigma^2 I), X \in \mathcal{R}^k \rightarrow \|X - \mu\|^2 \sim (\gamma : 2\sigma^2, \frac{k}{2})$.



Fitting Gamma Density

The pdf of $(\gamma : b, s)$

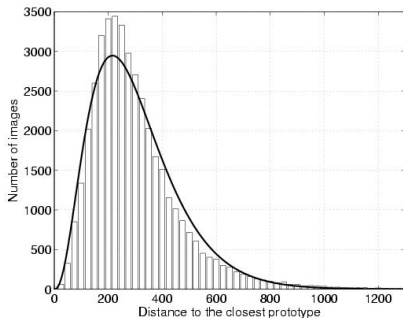
$$f(u) = \frac{\left(\frac{u}{b}\right)^{s-1} e^{-u/b}}{b\Gamma(s)}, \quad u \geq 0$$

ML estimation of $(\gamma : b, s)$

$$\begin{cases} \log \hat{s} - \psi(\hat{s}) = \log [\bar{u} / (\prod_{i=1}^n u_i)^{1/n}] \\ \hat{b} = \bar{u} / \hat{s} \end{cases}$$

di-gamma function:

$$\psi(s) = \frac{d \log \Gamma(s)}{ds}, \quad s > 0.$$



$$(\gamma : b = 86.34, s = 3.5)$$

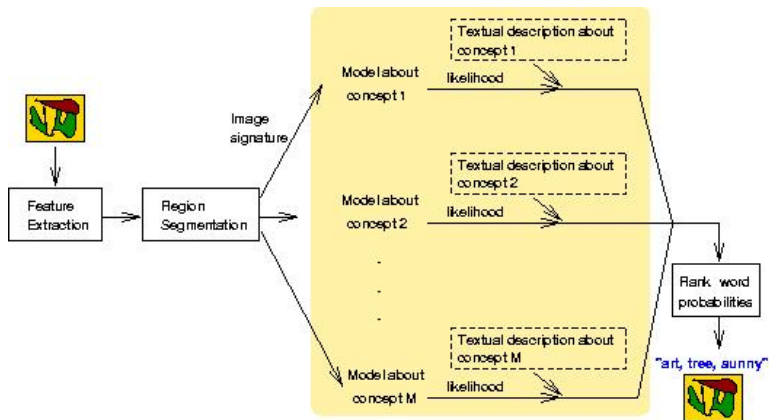
Mixture Model

The mixture model

- ▶ M prototypes: $\{\alpha_1, \alpha_2, \dots, \alpha_M\} \subset \Omega$.
- ▶ The mixture density:

$$\phi(\beta) = \sum_{\eta=1}^M \omega_{\eta} \left(\frac{1}{\sqrt{\pi b}} \right)^{2s} e^{-\frac{\tilde{D}(\beta, \alpha_{\eta})}{b}}$$

Annotation



Word Probabilities

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

Category prob. given signature β

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

Word probability

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

Word Probabilities

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

Category prob. given signature β

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

Word probability

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

Word Probabilities

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

Category prob. given signature β

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

Word probability

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

ALIPR Online Demo: alipr.com

ALIPR

Automatic Linguistic Indexing of Pictures - Real Time

URL of your image to be annotated

Example: http://wang.ist.psu.edu/~jwang/photography/2005/9_3/icons/RW_5725.jpg
<http://wang.ist.psu.edu/~jwang/photography/2004/6/icons/alaska0050.jpg>

[Computerized annotation results for over 54,000 images](#)

Computer annotation for example images



landscape building
historical
mountain
man-made indoor
people lake animal



flower plant
rose cactus
flora grass
landscape
water perennial



grass animal
wild-life sport
people rock
tree horse polo



texture indoor
food natural
people animal
landscape rock
man-made



man-made
indoor painting
people food
fruit mural old
poster



grass
landscape tree
lake autumn
people rural
texture natural

ALIPR Online Demo

flickr.com images



Computer annotation:
 ocean boat water people sky ship cruise
 beach wind work landscape sport surf wave
 building

User annotation: (10767)

pointreyes lighthouse pacific california marin pointreyesnationaleashore



Computer annotation:
 snow people ocean ice mountain
 winter water landscape blue sport ski
 dogsled plant wind surf

User annotation: (50023)



Computer annotation:
 texture natural micro_image plant orange
 food candie rock flower yellow fractal
 landscape sky sunset sun

User annotation: (20074)

bonn germany



Computer annotation:
 people man-made city modern building
 indoor stamp life historical festival
 cloth face child youth group

User annotation: (30958)

harleydavidson losangeles california littletokyo friends



Computer annotation:
 building historical landscape people landmark
 old works architecture world mountain grass house
 rock water ocean

User annotation: (16285)

munich germany europe summer2003



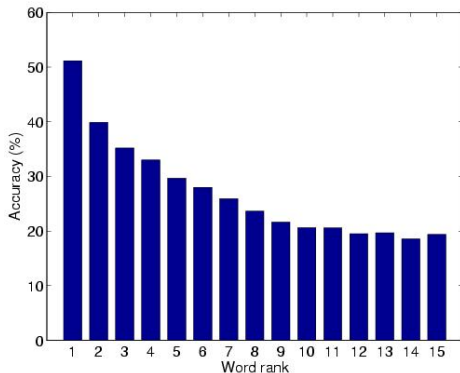
Computer annotation:
 animal wild_life grass tree landscape
 antelope rock elephant lake indoor
 man-made old dish natural sea

User annotation: (542)

orie birthday haas mba

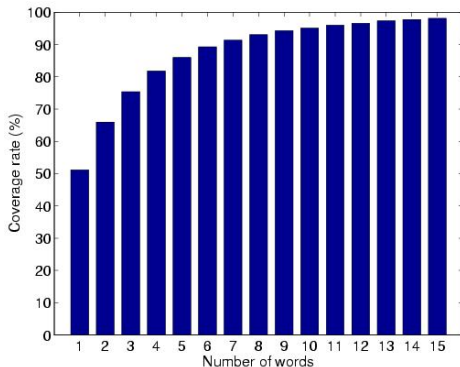
Human Evaluation on flickr.com Images

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



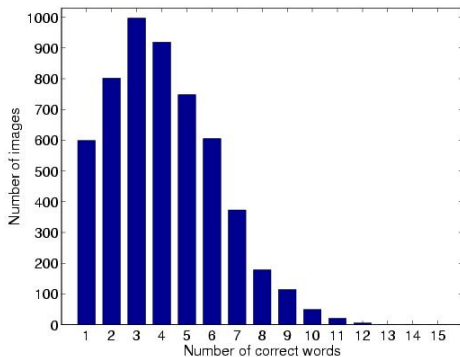
Human Evaluation on flickr.com Images

- ▶ 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



ALIPR Interactive

ALIPR

Automatic Linguistic Indexing of Pictures - Real Time

ALIPR Computerized Annotation



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

- people man-made royal_guard guard parade
 decoration cloth festival grass car
 sky steam engine building landscape

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

submit

ALIPR Interactive

ALIPR

Automatic Linguistic Indexing of Pictures - Real Time



ALIPR Computerized Annotation

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

- ocean
 water
 people
 sport
 boat
 wild_life
 wind
 lake
 cruise
 wave
 wet
 bird
 grass
 surf
 sky

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

submit

ALIPR Interactive

ALIPR

Automatic Linguistic Indexing of Pictures – Real Time



ALIPR Computerized Annotation

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

- | | | | | |
|---------------------------------------|-------------------------------------|------------------------------------|----------------------------------|---------------------------------|
| <input type="checkbox"/> building | <input type="checkbox"/> historical | <input type="checkbox"/> landscape | <input type="checkbox"/> water | <input type="checkbox"/> flower |
| <input type="checkbox"/> people | <input type="checkbox"/> church | <input type="checkbox"/> worship | <input type="checkbox"/> house | <input type="checkbox"/> plant |
| <input type="checkbox"/> rural_France | <input type="checkbox"/> lake | <input type="checkbox"/> modern | <input type="checkbox"/> reflect | <input type="checkbox"/> desert |

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

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