Real-time Computerized Annotation of Pictures

James Z. Wang
Joint work with Jia Li
The Pennsylvania State University
also with Carnegie Mellon University

ALIPR.com
The Needs

- Photographs created: 75 billions/year, most of them “invisible”
- Other applications: art, cultural, commercial, biomedicine, defense, law-enforcement, space exploration, education, entertainment, Web, ……

Data: UCB SIMS 2003
Challenges: High³

Satellite imaging
Up to 20,000x20,000
at 0.3-20 m/pixel
daily image acquisition
multi-spectra

- High throughput
- High resolution
- High dimensionality

Screening of zebrafish to study drugs
Up to 50,000x50,000; robotic stage
Can a Computer Tag Pictures?

water ocean people surf
boat sport up wind
wet fun summer

landscape animal art
mountain wild-life grass

people snow pioneer
man-made mountain
ice winter landscape
sky

cloth animal indoor
sport face male rural
horse compete
man-made
Potential Biomedical Applications

- Filtering images for pathologists and radiologists
- Generating medical knowledge based on millions of past cases
Outline

- Introduction
  - Our prior related work
    - SIMPLIcity visual similarity engine
  - ALIPR real-time image tagging
    - ALIP: Automatic Linguistic Indexing of Pictures
    - ALIPR: ALIP - real time
- Other work
  - Computational aesthetics
  - Story picturing engine
  - Art and cultural imaging
  - Image-based security
- Conclusions
Content-based Image Retrieval

- The retrieval of relevant images from an image database on the basis of automatically-derived image features

- Our approach:
  - Wavelets
  - Integrated Region Matching
  - Statistical methods
  - Supervised/unsupervised learning...

- History
  - 1995 Stanford University Art Library
  - 1996-2000 SIMPLIcity/WIPE/RIME
    - NSF, IBM QBIC, NEC, SRI
  - 2000-now ALIP/ALIPR
    - NSF, SUN

- Surveys
The Field of CBIR

- Highly interdisciplinary
- Highly challenging
- Diverse approaches

From a user perspective:

From a system perspective:

ALIPR
Penn State
Automatic Tagging

- Love
- Petals
- Flower
- Rose
- Corolla

- Symmetry
- Tower
- Eiffel Tower
- France
- Paris

- Australian Floribunda Rose

- Elegance
Basic Approach

Image Database

Query

Extract Features

Color
Texture
Shape

……

Visual Similarity
Computation/Ranking

Knowledge Base

Statistical Modeling

Tagging

Feature Database

Extract Features

flower plant landscape rural sky…….
Outline

- Introduction

- Our prior related work
  - SIMPLIcity visual similarity engine

- ALIPR real-time image tagging
  - ALIP: Automatic Linguistic Indexing of Pictures
  - ALIPR: ALIP - real time

- Other work
  - Computational aesthetics
  - Story picturing engine
  - Art and cultural imaging
  - Image-based security

- Conclusions
SIMPLIcity (1999)

- **Semantics-sensitive Integrated Matching for Picture Libraries**

- **Major features**
  - Sensitive to semantics: combine statistical semantic classification with image retrieval
  - Efficient processing: wavelet-based feature extraction
  - Reduced sensitivity to inaccurate segmentation and simple user interface: Integrated Region Matching (IRM)
Wavelets

- Wavelet: decomposes a 2-D image
  - trend (low-frequency) bands
  - fluctuation (high-frequency) bands
  - different scales
- Image applications:
  - processing (denoising, enhancement)
  - analysis, classification, compression
- Lossless inverse transform
- Haar wavelets
- Daubechies’ wavelets

\[
\phi_{r,j,k}(x) = 2^{j/2} \phi_r(2^j x - k), \ j, k \in \mathbb{Z}
\]
\[
< u, v > = \int_0^1 u(x)v(x)dx
\]
\[
f_j(x) = \sum_k < f, \phi_{r,j,k} > \phi_{r,j,k}(x)
\]
\[
d_j(x) = f_{j+1}(x) - f_j(x)
\]
Fast Segmentation

Original Image

Wavelet, RGB → LUV

K-means

Segmentation Result

(Centroid, Inertia) region features

Feature 1

Feature 2

Feature 1

Feature 2
Fast Segmentation

- **K-means clustering**
  - Minimize the mean squared error between the samples and their representative prototypes
  
  \[ D(k) = \sum_{i=1}^{L} \min_{1 \leq j \leq k} (x_i - \hat{x}_j)^2 \]

- **Normalized inertia for shape**

![Images of segmentation results]
Integrated Region Matching

- IRM defines an image-to-image distance as a weighted sum of region-to-region distances

\[ d(R_1, R_2) = \sum_{i,j} s_{i,j}d_{i,j} \]

- Integrate point-wise distance by linear combination
- Weighting matrix is determined based on significance constraints and a ‘MSHP’ greedy algorithm
SIMPLIcity in the Real World

airliners.net

mindat.org

terragalleria.com
Clustering by Graph Partitioning

Graph: Vertices → Images, Edges → Similarity
Clustering by Graph Partitioning

Beijing Cluster 1 (61 images)  
Beijing Cluster 2 (59 images)  
Beijing Cluster 3 (43 images)  
Beijing Cluster 4 (31 images)
Outline

- Introduction
- Our prior related work
  - SIMPLIcity visual similarity engine
  - ALIPR real-time image tagging
    - ALIP: Automatic Linguistic Indexing of Pictures
    - ALIPR: ALIP - real time
- Other work
  - Computational aesthetics
  - Story picturing engine
  - Art and cultural imaging
  - Image-based security
- Conclusions
Why Image Tagging

- DB size is less of a challenge
- Understandability & Vision
  - “meaning” depend on the point-of-view
  - Can we translate contents and structure into linguistic terms
- Query formulation
  - SIMILARITY
  - OBJECT: contains a dog
  - OBJECT RELATIONSHIP: contains a dog and a person
  - MOOD: a sad picture
  - TIME/PLACE: sunset near the Capital
- Kyoto
- dogs

ALIPR™ PENNSTATE
Limitless Diversity and Creativity

Telephoto lens
Macro lens
Small aperture
Large aperture
Long exposure while zooming in
Multiple exposures
Sky shot, minutes long exposure
Fast exposure
Seconds long exposure
Lens artifacts
ALIP (2002)

- Automatic Linguistic Indexing of Pictures
- Observations
  - Human beings can build models about objects or concepts based on visual information
  - The learned models are stored in the brain and used in the recognition process
- Question: Can a computer build and use models about a large collection of concepts using images?
ALIP Overview

- **Train** statistical models of a dictionary of concepts using sets of training images
  - 2D images are currently used
  - 3D-image training can be much better
- **Model based comparisons**
- Select the most statistical significant concept(s)
- **600** concepts, each trained with 40 images
  - 15 minutes Pentium CPU time per concept, train only once
  - highly parallelizable algorithm
- **Concepts**: Basic building blocks in determining the semantic meanings of images
  - **Basic Object**: flower, beach
  - **Object composition**: building+grass+sky+tree
  - **Location**: Asia, Venice
  - **Time**: night sky, winter frost
  - **Abstract**: sports, sadness
- Model: inter-scale and intra-scale dependence
- Hierarchical Markov mesh
- Color and texture features in SIMPLIcity: multivariate Gaussian distributed given states
- Use EM algorithm to estimate parameters
- J. Li et al. [IEEE Trans. Info. Theory, 1999]
Annotation Process

Suppose a word appears $j$ and $m$ times in the annotation of top $k$ and all $n$ categories, respectively. The prob. the word appeared by chance:

$$P(j, k) = \sum_{i=j}^{k} I(i \leq m) \frac{\binom{n}{i}}{\binom{n}{k}} = \sum_{i=j}^{k} I(i \leq m) \frac{m!}{(m-i)! (m-i)! (n-m-k+i)! n!}$$

When $n, m >> k$, we have the approximation

$$P(j, k) = \sum_{i=j}^{k} \binom{k}{i} p^i (1-p)^{k-i} = \sum_{i=j}^{k} \frac{k!}{i!(k-i)!} p^i (1-p)^{k-i}$$

- Rank the categories by the likelihoods of an image to be annotated under the profiling 2-D MHMMs
- Select annotation words from those used to describe the top-ranked categories
  - Statistical significance is computed for each candidate word
  - Words that are unlikely to have appeared by chance are selected
  - Favor the selection of rare words
## Concepts

<table>
<thead>
<tr>
<th>ID</th>
<th>Category Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Africa, people, landscape, animal</td>
</tr>
<tr>
<td>10</td>
<td>England, landscape, mountain, lake, European, people, historical building</td>
</tr>
<tr>
<td>20</td>
<td>Monaco, ocean, historical building, food, European, people</td>
</tr>
<tr>
<td>30</td>
<td>royal guard, England, European, people</td>
</tr>
<tr>
<td>40</td>
<td>vegetable</td>
</tr>
<tr>
<td>50</td>
<td>wild life, young animal, animal, grass</td>
</tr>
<tr>
<td>60</td>
<td>European, historical building, church</td>
</tr>
<tr>
<td>70</td>
<td>animal, wild life, grass, snow, rock</td>
</tr>
<tr>
<td>80</td>
<td>plant, landscape, flower, ocean</td>
</tr>
<tr>
<td>90</td>
<td>European, historical building, grass, people</td>
</tr>
<tr>
<td>100</td>
<td>painting, European</td>
</tr>
<tr>
<td>110</td>
<td>flower</td>
</tr>
<tr>
<td>120</td>
<td>decoration, man-made</td>
</tr>
<tr>
<td>130</td>
<td>Alaska, landscape, house, snow, mountain, lake</td>
</tr>
<tr>
<td>140</td>
<td>Berlin, historical building, European, landscape</td>
</tr>
<tr>
<td>150</td>
<td>Canada, game, sport, people, snow, ice</td>
</tr>
<tr>
<td>160</td>
<td>castle, historical building, sky</td>
</tr>
<tr>
<td>170</td>
<td>cuisine, food, indoor</td>
</tr>
<tr>
<td>180</td>
<td>England, landscape, mountain, lake, tree</td>
</tr>
<tr>
<td>190</td>
<td>fitness, sport, indoor, people, cloth</td>
</tr>
<tr>
<td>200</td>
<td>fractal, man-made, texture</td>
</tr>
<tr>
<td>210</td>
<td>holiday, poster, drawing, man-made, indoor</td>
</tr>
<tr>
<td>220</td>
<td>Japan, historical building, garden, tree</td>
</tr>
<tr>
<td>230</td>
<td>man, male, people, cloth, face</td>
</tr>
<tr>
<td>240</td>
<td>wild, landscape, north, lake, mountain, sky</td>
</tr>
<tr>
<td>250</td>
<td>old, poster, man-made, indoor</td>
</tr>
<tr>
<td>260</td>
<td>plant, art, flower, indoor</td>
</tr>
<tr>
<td>270</td>
<td>recreation, sport, water, ocean, people</td>
</tr>
<tr>
<td>280</td>
<td>ruin, historical building, landmark</td>
</tr>
<tr>
<td>290</td>
<td>sculpture, man-made</td>
</tr>
<tr>
<td>300</td>
<td>St Moritz, ski, snow, ice, people</td>
</tr>
<tr>
<td>310</td>
<td>texture, man-made, painting</td>
</tr>
<tr>
<td>320</td>
<td>texture, natural</td>
</tr>
<tr>
<td>330</td>
<td>train, landscape, man-made</td>
</tr>
<tr>
<td>340</td>
<td>Virginia, historical building, landscape, rural</td>
</tr>
<tr>
<td>350</td>
<td>wild life, art, animal</td>
</tr>
<tr>
<td>360</td>
<td>work, people, cloth</td>
</tr>
<tr>
<td>370</td>
<td>architecture, building, historical building</td>
</tr>
<tr>
<td>380</td>
<td>Canada, British Columbia, landscape, mountain</td>
</tr>
<tr>
<td>390</td>
<td>blue</td>
</tr>
<tr>
<td>400</td>
<td>Canada, landscape, historical building</td>
</tr>
<tr>
<td>410</td>
<td>city, life, people, modern</td>
</tr>
<tr>
<td>420</td>
<td>Czech Republic, landscape, historical building</td>
</tr>
<tr>
<td>430</td>
<td>Easter egg, decoration, indoor, man-made</td>
</tr>
<tr>
<td>440</td>
<td>fashion, people, cloth, female</td>
</tr>
<tr>
<td>450</td>
<td>food, man-made, indoor</td>
</tr>
<tr>
<td>460</td>
<td>green</td>
</tr>
<tr>
<td>470</td>
<td>interior, indoor, man-made</td>
</tr>
<tr>
<td>480</td>
<td>marine time, water, ocean, building</td>
</tr>
<tr>
<td>490</td>
<td>museum, old, building</td>
</tr>
<tr>
<td>500</td>
<td>owl, wild life, bird</td>
</tr>
<tr>
<td>510</td>
<td>plant, flower</td>
</tr>
<tr>
<td>520</td>
<td>reptile, animal, rock</td>
</tr>
<tr>
<td>530</td>
<td>sail, boat, ocean</td>
</tr>
<tr>
<td>540</td>
<td>Asia, historical building, people</td>
</tr>
<tr>
<td>550</td>
<td>skin, texture, natural</td>
</tr>
<tr>
<td>560</td>
<td>summer, people, water, sport</td>
</tr>
<tr>
<td>570</td>
<td>car, man-made, landscape, plane, transportation</td>
</tr>
<tr>
<td>580</td>
<td>US, landmark, historical building, landscape</td>
</tr>
<tr>
<td>590</td>
<td>women, face, female, people</td>
</tr>
</tbody>
</table>
Training images used to train the concept “Paris/France” with description “Paris, European, historical_building, beach, landscape, water”
Training images used to train the concept “male” with description “man, male, people, cloth, face”
ALIP: Results and Advantages

- Accumulative learning
- Highly scalable (compare to CART, SVM, ANN)
- Flexible amount of training for each concept
- Statistical likelihood rather than image similarity
- Spatial relations among pixels

Snow, animal, wildlife, sky, cloth, ice, people

Building, sky, lake, landscape, Europe, tree

Food, indoor, cuisine, dessert
ALIPR: Real-time Challenge

- **ALIP**: 10 minutes/image
- **ALIPR**: within one second
  - Download and scale down the picture
  - Color space: RGB to LUV
  - Daubechies wavelet transform
  - Region segmentation
  - Extraction of color and texture distributions
  - Sequential comparison to 600 trained classes
  - Rank word probabilities

people man-made building historical landscape life face indoor food occupation cloth child youth decoration male Buddha furniture laptop
The Training Process

- Corel images
  - 80 images per category
  - Each category is described with a few words
  - 332 distinct words
indoor food fruit man-made poster old people grass animal rural dessert flower yellow barbecue texture fish fingers plate
Region Signatures

- An image signature resides in \( \Omega = \Omega_1 \times \Omega_2 \).

- Color distribution: \( \beta_{i,1} \in \Omega_1 \).

- Texture distribution:
  \( \beta_{i,2} \in \Omega_2 \).

- \( \beta_{i,j} = \{ (v^{(1)}_{i,j}, p^{(1)}_{i,j}), \ldots, (v^{(m_{i,j})}_{i,j}, p^{(m_{i,j})}_{i,j}) \} \).

landscapel buildingpeoplehistoricalruralbeachindoorparadisewateranimalhousefarmmanmadeoceantrainboat
Mallows Distance between Distributions

Let the two discrete distributions be

\[ \gamma_i = \{(z_i^{(1)}, q_i^{(1)}), (z_i^{(2)}, q_i^{(2)}), \ldots, (z_i^{(m_i)}, q_i^{(m_i)})\}, i = 1, 2 \]

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 \( \sum_{j=1}^{m_2} w_{i,j} = q_1^{(i)}, i = 1, \ldots, m_1, \)
\[ \sum_{i=1}^{m_1} w_{i,j} = q_2^{(j)}, j = 1, \ldots, m_2, w_{i,j} \geq 0, i = 1, \ldots, m_1, j = 1, \ldots, m_2. \]

Solved by linear programming.
Mixture Modeling for Space $\Omega$

- Carve $\Omega$ into cells by clustering.
- Map each cell to an Euclidean space, preserving pairwise distances.
- Model the mapped points by Gaussian.
**D2 (Discrete Distribution) - Clustering**

- An image set
  \[ B = \{ \beta_i : \beta_i \in \Omega, i = 1, ..., n \}, \]
  \[ \Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_d \]

- 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
  \[ A = \{ \alpha_i : \alpha_i \in \Omega, i = 1, ..., \bar{m} \}. \]

- Prototype assignment function
  \[ c(i) \in \{1, 2, ..., \bar{m}\}, i = 1, ..., n. \]
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 \rightarrow D2-Clustering
- (vectors) \rightarrow (bags of weighted vectors)

building historical landscape
mountain snow people cloud
water ice glacier sky
technology communication
texture natural peaks
Algorithm

1. For every image \( i \), set
   \[ c(i) = \arg \min_{j=1, \ldots, m} \bar{D}(\beta_i, \alpha_j). \]

2. Let \( C_j = \{ i : c(i) = j \} \), \( j = 1, \ldots, m \). That is, \( C_j \) contains indices of images assigned to prototype \( j \).
   For each prototype \( j \), set
   \[ \alpha_j = \arg \min_{\alpha \in \Omega} \sum_{i \in C_j} \bar{D}(\beta_i, \alpha) \quad \text{— challenging} \]

Update of Prototypes

1. For every image \( i \), set \( c(i) = \arg \min_{j=1, \ldots, m} \bar{D}(\beta_i, \alpha_j) \).
2. Let \( C_\eta = \{ i : c(i) = \eta \}, \eta = 1, \ldots, m \). Update each \( \alpha_{\eta, l} \), \( \eta = 1, \ldots, m \), \( l = 1, \ldots, d \), individually by the following steps. Denote \( \alpha_{\eta, l} = \{(z^{(1)}_{\eta, l}, q^{(1)}_{\eta, l}), (z^{(2)}_{\eta, l}, q^{(2)}_{\eta, l}), \ldots, (z^{(m_{\eta, l})}_{\eta, l}, q^{(m_{\eta, l})}_{\eta, l})\} \).
   2.1 Fix \( z^{(k)}_{\eta, l}, k = 1, \ldots, m_{\eta, l} \). Update \( q^{(k)}_{\eta, l}, w^{(l)}_{k,j} \), \( i \in C_\eta, k = 1, \ldots, m_{\eta, l}, j = 1, \ldots, m_{l,j} \) by solving the linear programming problem:
      \[
      \begin{align*}
      &\min q^{(k)}_{\eta, l} \sum_{i \in C_\eta} \min_{w^{(l)}_{k,j}} \sum_{k=1}^{m_{\eta, l}} \sum_{j=1}^{m_{l,j}} w^{(l)}_{k,j} \| z^{(k)}_{\eta, l} - v^{(l)}_{i,j} \|^2, \\
      &\text{subject to } \sum_{k=1}^{m_{\eta, l}} q^{(k)}_{\eta, l} = 1; \\
      &q^{(k)}_{\eta, l} \geq 0, k = 1, \ldots, m_{\eta, l}; \\
      &\sum_{j=1}^{m_{l,j}} w^{(l)}_{k,j} = q^{(k)}_{\eta, l}, i \in C_\eta, k = 1, \ldots, m_{\eta, l}; \\
      &\sum_{k=1}^{m_{\eta, l}} w^{(l)}_{k,j} = p^{(l)}_{j,i}, i \in C_\eta, j = 1, \ldots, m_{l,j}; \\
      &w^{(l)}_{k,j} \geq 0, i \in C_\eta, k = 1, \ldots, m_{\eta, l}, j = 1, \ldots, m_{l,j}. \\
      \end{align*}
      \]

2.2 Fix \( q^{(k)}_{\eta, l}, w^{(l)}_{k,j} \), \( i \in C_\eta, 1 \leq k \leq m_{\eta, l}, 1 \leq j \leq m_{l,j} \). Update \( z^{(k)}_{\eta, l}, k = 1, \ldots, m_{\eta, l} \) by
      \[
      z^{(k)}_{\eta, l} = \frac{\sum_{i \in C_\eta} \sum_{j=1}^{m_{l,j}} w^{(l)}_{k,j} v^{(l)}_{i,j}}{\sum_{i \in C_\eta} \sum_{j=1}^{m_{l,j}} w^{(l)}_{k,j}}.
      \]

2.3 Compute \( \sum_{k=1}^{m_{\eta, l}} \sum_{j=1}^{m_{l,j}} w^{(l)}_{k,j} \| z^{(k)}_{\eta, l} - v^{(l)}_{i,j} \|^2 \). 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.
Fitting Gamma Density

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

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

\gamma : b = 86.34, s = 3.5
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: \( \rho_m \) (set uniform).

**Category prob. given signature \( \beta \)**

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

**Word probability**

\[
q(\beta, w_i) = \sum_{m: m \in \mathcal{C}(w_i)} p_m(\beta)
\]
ALIPR Online Demo

ALIPR™
Automatic Linguistic Indexing of Pictures - Real Time

Image Upload: Browse... alipr it!

Or Image URL: alipr it!

http://... Try drag and drop from another website.

search + most-voted + emotions + pictures

or random: seed - key word - bell - lady bird - force addiction - sunflower - banff - pets - passage - fish - human...

boat ocean water sky ship work people
sport landscape lake cloth snow ski winter
man-made man fish

ALIPR.com
ALIPR Online Demo

- Training: 109 seconds per category, 80 images
- 2.4 GHz AMD
- Annotation: one second per image
Human Subject Evaluation

- Manual evaluation on 5400+ flickr.com images
- Accuracy of the first word: 51.17%

flower people night group studio fruit still_life building plant indoor man-made mask sky firework city Hong Kong New Year celebration
Coverage Rate

- Percentage of images correctly annotated by at least one word
- Top 4: > 80%
- Top 7: 91.37%
- Top 15: 98.13%

people face cloth female male fashion child youth man-made texture fabric hair_style sky woman indoor James Bond 007 movie Casino Royale
On average, 4.1 words out of top 15 are correct.
Alipred Images in News or Blog

man-made texture color people indoor food painting royal guard fruit feast holiday mural cloth abstract guard thirsty kitty

Gizmodo blog

building historical landscape people man-made train mountain lake landmark ruin city modern architecture sport snow

GeEK blog

flower natural pattern landscape texture man-made rural pastoral plant tree green rock color animal grass China

Gizmodo blog

building historical landscape animal landmark ruin grass snow wild_life sky people photo rock fox castle forest cloud lake

ZDNet News

indoor animal drawing ancestor art man-made thing food dinosaur antique people cloth sport fitness Firearm USB memory Make Differences blog

lanscape indoor animal art mountain wild_life people flower plant food fruit building historical lake grass horse horses sky wind

Discovery Channel

animal grass wild_life people sport dog horse polo lion tree water rare_animal wild_cat pet landscape

Scientific American.com
SIMPLIcity Integration
Related Image Search

Based on verified tags
Database: 1 million images
Computer vs. Human

- **Computer training**
  - 2-D images only
  - No size info in images
  - Cannot imagine
  - Massive scale possible

- **Human training**
  - 3-D stereo, interpretation
  - Has size info
  - Imagination power
  - Memory limitations
Unsuccessful Examples

building people water modern city work historical cloth horse
User: photo unfound molly dog animal
Suspected problem: unusual shot

texture indoor food natural cuisine man-made fruit vegetable dessert
User: phonecamera car
Suspected problem: fuzzy shot, unusual white balance

texture natural flower sea micro_image fruit food vegetable indoor
User: me selfportrait orange mirror
Suspected problem: unusual shot, unusual white balance

texture painting flower landscape rural pastoral plant grass natural
User: 911 records money green n2o
Suspected problem: unlearned object or concept
Outline

● Introduction
● Our prior related work
  ● SIMPLIcity visual similarity engine
● ALIPR real-time image tagging
  ● ALIP: Automatic Linguistic Indexing of Pictures
  ● ALIPR: ALIP - real time
● Other work
  ● Computational aesthetics
  ● Story picturing engine
  ● Art and cultural imaging
  ● Image-based security
● Conclusions
Computational Aesthetics

- “Train” a computer to infer visual aesthetics
- Rating data from: photo.net
- Highly subjective, noisy data, evolving concept
- Steps: (1) hypothesize visual features that may correlate with aesthetics; (2) statistically select features; (3) learn to classify
- Features: intensity, saturation, aspect ratio, colorfulness, rule of thirds, familiarity, texture, low DOF, shape, ……

Aesthetics: 4.85/7
Aesthetics: 6.40/7
Story Picturing Engine

Vermont is mostly a rural state. Many of the towns and villages, churches, covered bridges, weathered farms and barns date back to the 17th century, where the ancestors of many Americans lived before moving west. The countryside has the cozy feeling of a place which has been lived in for a long time, before the age of the machines. Each autumn, the landscape is transformed in vibrant palette of colors that few places can rival, for the Green Mountains have a great variety of trees, many of which turned red or orange.

Paris. The cradle of Paris is Île de la Cité, a small island in the middle of the Seine river. Paris then first developed along the river, and from the Louvre to the Eiffel tower or the Place de la Concorde to Notre Dame, its evolution and its history can be seen from the river. The central portion of the banks form a very homogeneous ensemble and makes for a fine walk. The banks of the Seine are UNESCO World heritage site. The right bank of the Seine is dominated by the large perspectives due to the avenues designed by Haussman in the 19th century. The most prominent of them is the one extending from the Louvre to the Arc de Triomphe, through the Champs Elysees, France’s most famous avenue. (by: Q.-T. Luong)
Zhang, Daqian  Shen, Zhou

Heavy and thick wash

Straight, light wash with some vertical lines

Smooth region

Swift, thin strokes

Flat controlled strokes

Small dark strokes

Sharp lines and straight wash

Diluted and washed
Image-based Security

CAPTCHA - Completely Automated Public Test to Tell Computers and Humans Apart

Our wok: IMAGINATION: Image Generation for Internet Authentication

Exploit the semantic gap

Click Phase – Select center of an image

Annotate Phase – Select best label from list
Miscellaneous Projects

- Satellite image databases
- High-throughput biomedical imaging
- 3-D hidden Markov models
Miscellaneous Projects

Bridging the annotation-retrieval gap
Outline

- Introduction
- Our prior related work
  - SIMPLIcity visual similarity engine
- ALIPR real-time image tagging
  - ALIP: Automatic Linguistic Indexing of Pictures
  - ALIPR: ALIP - real time
- Other work
  - Computational aesthetics
  - Story picturing engine
  - Art and cultural imaging
  - Image-based security
- Conclusions
Conclusions

- Content-based image retrieval is important with digital imaging technologies and the Web
- With massive and efficient training, computers can assist the tagging process
- A paradigm shift in research focus is expected
  - Application-oriented CBIR
    - Art, culture, biomedicine, security, Web, ……
  - Domain-specific system leveraging domain knowledge
  - Profound impact to Web users
  - A wide range of approaches
- semantic gap, aesthetic gap
The Team

Collaborators:
Jia Li, Penn State
Yixin Chen, Univ. of Mississippi
Dhiraj Joshi, Kodak Research

Students:
Ritendra Datta
Diane Flowers
Walter Weiss

Funding:
NSF
PNC Bank
SUN

Data:
Amico
Terragalleria
Airliners.net
Yahoo!Research
Flickr
Photo.net

More Info:
Googling “image search research”
or “automatic image tagging”

ALIPR.com
Difficulties

“The truth of the matter is that these images (on sites like Flickr) are largely all about the same thing -- people mostly take pictures of other people. So just using the word ‘people’ already tags a large percentage of the images correctly.”

- a CMU faculty commenting on the ALIPR work, November 2006
Top 15 Computer-Predicted Tags
ALIPR is like a child trying to learn about the world. Please help us to teach ALIPR. Check those correctly annotated words.

- ☑ indoor
- ☐ animal
- ☑ food
- ☑ drawing
- ☐ fruit

- ☑ art
- ☑ man-made
- ☐ flower
- ☐ dog
- ☐ pet

- ☐ ancestor
- ☐ drink
- ☐ antique
- ☐ dinosaur
- ☐ poster

Thought of other terms missed by ALIPR? Please add here, separated by commas ','; cookies, chocolate, thank you. Teach ALIPR and make the picture searchable.