Hands On: Multimedia Methods for Large Scale Video Analysis
(Lecture)

Dr. Gerald Friedland,
fractor@icsi.berkeley.edu
Today

- Response to Questions
- More on Visual Analysis:
  - More on Features
  - Outlook on Systems
Co-occurrence Matrix: What’s the Intuition?

\[ C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^{n} \sum_{q=1}^{m} \begin{cases} 1, & \text{if } I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{otherwise} \end{cases} \]
Co-Occurrence Matrix: Example
FAQ

• How do you Equalize/Gaussianize?

• Easy Answer: Use Kernel

• Sophisticated Answer: Google Research Papers...
• If you hold the Nintendo 3DS upside down, will the 3D effect still work?

See yourself…
More on Features

• Finishing up Basic Features
• More Advanced Features:
  - Color Signatures
  - SIFT
  - GIST
• **Edge features:**

```matlab
gray_image = rgb2gray(my_image);
double_i = im2double(gray_image);
subplot(221),imshow(gray_image);
edge_image1 =
    edge(double_i,'sobel');
subplot(222),imshow(edge_image1)
edge_image2 =
    edge(double_i,'prewitt');
subplot(223),imshow(edge_image2)
edge_image3 =
    edge(double_i,'canny');
subplot(224),imshow(edge_image3);
```
Advanced Visual Features

• Shape features
  - A good shape representation feature for an object should be invariant to translation, rotation, and scaling
  - The use of shape features for image retrieval has been limited to special applications where objects or regions are readily available
  - Shape description can be categorized into either boundary-based or region-based method
Color Signatures for Retrieval

Image retrieval:

Color Signatures

CIELAB space:
Create Color Signatures

Find representative colors by equal-size clustering (modified KD-Tree clustering [Rubner et. al.]):

Stage 1:  
In each recursion look at one of L, a, and b axis.  
Create two equal-sized subintervals.  
Repeat until interval is smaller than a threshold.  
Calculate the centroids for all clusters.

Stage 2:  
Repeat Step 1 using the cluster centroids.  
Eliminate all clusters that represent less points in the image than a threshold.
Create Color Signatures

Example:
Scale Invariant Feature Transform (SIFT)

Used for Object Recognition:


Mostly invariant to:

- Scale
- Rotation
- Illumination
- Viewpoint
Steps:

- Scaling + blurring
- Finding difference of Gaussians (keypoints)
- Remove “bad keypoints”
- Assign orientation to keypoints
- Rotation and scale normalization
SIFT: Scaling, Blurring
SIFT: Difference of Gaussians
SIFT: Difference of Gaussians
SIFT: Keypoints by Finding Maxima
SIFT: Removing “Bad” Keypoints

Removing low contrast

Removing edge responses
SIFT: Normalizing
Categorization of Natural Scenes

Confusion Matrix (in % using Layout template):
Classification of prototypical scenes (400 / category)

<table>
<thead>
<tr>
<th></th>
<th>Coast</th>
<th>Countryside</th>
<th>Forest</th>
<th>Mountain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coast</td>
<td>88.6</td>
<td>8.9</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>Countryside</td>
<td>9.8</td>
<td>85.2</td>
<td>3.7</td>
<td>1.3</td>
</tr>
<tr>
<td>Forest</td>
<td>0.4</td>
<td>3.6</td>
<td>91.5</td>
<td>4.5</td>
</tr>
<tr>
<td>Mountain</td>
<td>0.4</td>
<td>4.6</td>
<td>3.8</td>
<td>91.2</td>
</tr>
</tbody>
</table>

Local organization: correct for 92% images
(4 similar images on 7 K-NN)

Slide Credit: Olivia
# Categorization of Manmade Scenes

## Confusion Matrix (in % using Layout template)

Classification of prototypical scenes (400 / category)

<table>
<thead>
<tr>
<th></th>
<th>Highway</th>
<th>Street</th>
<th>City centre</th>
<th>Tall building</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highway</td>
<td>91.6</td>
<td>4.8</td>
<td>2.7</td>
<td>0.9</td>
</tr>
<tr>
<td>Street</td>
<td>4.7</td>
<td>89.6</td>
<td>1.8</td>
<td>3.4</td>
</tr>
<tr>
<td>Centre</td>
<td>2.5</td>
<td>2.3</td>
<td>87.8</td>
<td>7.4</td>
</tr>
<tr>
<td>Tall Building</td>
<td>0.1</td>
<td>3.4</td>
<td>8.5</td>
<td>88</td>
</tr>
</tbody>
</table>

## Local organization:
correct for 86 % images

(4 similar images on 7 K-NN)

---

Slide Credit: Olivia
Temporal Video Features

Difference frame:
Other Features

- Histogram of Oriented Gradients (HoG)
- Speeded Up Robust Feature (SURF)
- Local Energy-based Shape Histogram (LESH)
- etc...
Temporal Video Features

Optical flow (compressed domain)
High-Level Visual Features

- Region detection: Skin color, textures
- Categories: indoor and outdoor, play and non-play, etc.
- Face detection: number of faces, location of face, etc.
High-Level Visual Features

- Objects, concepts, events, etc.
- Existence of an entity:
  e.g., trees
- Descriptive meaning
  e.g., sports
Manual Feature Selection:

- Choose a feature subset from the original feature set, which best represents the target semantic concepts.
- Having more features should surely result in more discriminating power, but adding irrelevant or distracting features often confuses system.
Example: TrecVid MED 2010

Making a cake

Batting a run in

Assembling a shelter
Example: TrecVid MED 2010

<table>
<thead>
<tr>
<th>Human Action Concepts</th>
<th>Scene Concepts</th>
<th>Audio Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person walking</td>
<td>Indoor kitchen</td>
<td>Outdoor rural</td>
</tr>
<tr>
<td>Person running</td>
<td>Outdoor with grass/trees visible</td>
<td>Outdoor urban</td>
</tr>
<tr>
<td>Person squatting</td>
<td>Baseball field</td>
<td>Indoor quiet</td>
</tr>
<tr>
<td>Person standing up</td>
<td>Crowd (a group of 3+ people)</td>
<td>Indoor noisy</td>
</tr>
<tr>
<td>Person making/assembling stuffs with hands (hands visible)</td>
<td>Cakes (close-up view)</td>
<td>Original audio</td>
</tr>
<tr>
<td>Person batting baseball</td>
<td></td>
<td>Dubbed audio</td>
</tr>
</tbody>
</table>

Newer Approaches

- Automatic Feature Selection:
  - Let the computer learn by example a subset from the original feature set, which best represents the target semantic concepts
  - Problem: Training set might not be representative enough
Automatic Feature Selection

Audio Signal

Percepts Extraction

Percepts Weighing

Classification

Concept (train)

Concept (test)
Distribution of “Percepts”

Near Zipfian Distribution!
Next Week (Project Meeting)

- More on Amazon
- Tools at ICSI
Next Week (Lecture)

• Multimedia Systems