Ph.D. Thesis Proposal

Machine Tagging for Personal Photos

Mei-Chen Yeh

Committee:
Prof. Tim Cheng (Chair)
Prof. Jerry Gibson
Prof. Kenneth Rose
Prof. Matthew Turk
Dr. Yi-Jen Chiu

12/05/2006
A large amount of photos (>10000)

Personal photos in a future digital home currently are not readily accessible in large sets due to the lack of **meta-data**.

Find me the photos at Mei’s qualifying exam if any…
Objectives

Automatic creation of metadata!

**Off-line Process**
One photo might take seconds to be processed

**Online Process**
Real-time photo search

Meaningful tags!

Mei’s Engine

textual query
party

GUI
Agenda

• Introduction
• Image Categorization (*ACM MM 06*)
• Media Content Management
• Summary
Image categorization

(Classify) each input image into one (or none) of the predefined categories

airplane  building  bus  car  highway  street
Categorization framework

Input image

Categorization based on low-level visual cues

Categorization based on detected text lines

Fusion

Image category
Learning *text concepts* is **not** a trivial task!

- Multiple text lines in one image
- Not every text line is representative
- False alarms from text detector
Multiple instance learning

• A variation of supervised learning with **ambiguous labels** on training examples
  – A bag is a collection of instances
  – Positive bag has at least one positive instance
  – Negative bag has only negative instances

• Analogy:
  Image $\approx$ Bag
  Text line $\approx$ Instance

*To learn a text concept for each target category*
Multiple instance learning (cont.)

Diverse Density (DD) [Maron and Lozano-Perez, 1998]

Positives

Negatives

The text concept for the “highway” category is learned!
Categorization using text concepts

- Each text concept is represented as two 16-d vectors
  - A 16-d feature vector with representative values across all the text lines in this category
  - A 16-d weight vector indicating impacts of each feature

For testing...

\[
\text{Decide } C_i \text{ with minimal } d_i
\]
imET (images with Embedded Text) dataset
http://lbmedia.ece.ucsb.edu/resources/dataset/imET.rar

<table>
<thead>
<tr>
<th>Category name</th>
<th>Number of images</th>
<th>Average number of text lines per image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airplane</td>
<td>260</td>
<td>1.7</td>
</tr>
<tr>
<td>Building</td>
<td>207</td>
<td>1.2</td>
</tr>
<tr>
<td>Bus</td>
<td>294</td>
<td>5.2</td>
</tr>
<tr>
<td>Camera</td>
<td>260</td>
<td>2.1</td>
</tr>
<tr>
<td>Car</td>
<td>220</td>
<td>3.3</td>
</tr>
<tr>
<td>Door</td>
<td>200</td>
<td>1.6</td>
</tr>
<tr>
<td>Highway</td>
<td>205</td>
<td>3.4</td>
</tr>
<tr>
<td>Street</td>
<td>250</td>
<td>1.6</td>
</tr>
</tbody>
</table>
Accuracy – cues from text alone

Average accuracy: 72.3%
or 80.1% if “street” category is excluded.
Categorization framework

Input image

- Categorization based on low-level cues
- Categorization based on detected text lines

Fusion

Image category
Bag-of-words model

Object

Bag of ‘words’

http://people.csail.mit.edu/torralba/iccv2005/
Bag-of-words model (cont.)

SIFT descriptors from training samples

Training images

128-D feature space

codeword 1

codeword 2

codeword 3

http://www.cs.ubc.ca/~lowe/keypoints/
Accuracy – the baseline

Categorization based on low-level cues (Bag-of-words model)

Image category

Average accuracy: 81.3%
Categorization framework

Input image

- Categorization based on low-level cues
- Categorization based on detected text lines

Fusion

Image category
Fusion

• Top-k strategy

  ![Image of a road with signs]

  Bag-of-words model

  \[ P_{\text{airplane}} \]
  \[ P_{\text{buildings}} \]
  \[ P_{\text{highway}} \]
  \[
  \vdots
  \\
  \text{Consider those categories with high } P
  \\
  \\
  \]

• A linear SVM classifier
  – Probabilities from the bag-of-words model
  – Distances to the learned text concepts
  – Number of text lines

Accuracy?
Accuracy – direct comparison

- text concept (average accuracy = 72.3%)
- bag-of-words (average accuracy = 81.3%)
- fusion approach (average accuracy = 90.1%)

8.8% improvement
Agenda

• Introduction
• Image Categorization (*ACM MM 06*)
• Media Content Management
  – Tag Creation: A Tagging Tool for Personal Photos
  – Tag Organization: Faceted Metadata
  – Tag Enhancement: Cues from different sources
• Summary
Mei’s Engine

Screen shots of the Intel Schematic GUI
Mei’s Engine

Metadata
2005/03/21 15:02:43
Canon PowerShot SD550
No flash used
One frontal face
Outdoor scene
Objects: Sky, Snow
Dominant Colors:

Colors
23 standard colors with index 0-22

Objects
Faces
Sky
Grass
Waterside
Trees
Snow
Buildings
Flowers
Food

Properties
Photo name
Date / time
Camera make
Camera model
Flash on/off
Indoor y/n
Night scene y/n
Sunset y/n
System flowchart

Extract Exif data
- Date, Time, Camera make/model, Flash used, …

Find faces

Find Dominant Colors
- (score, sc_index, h_offset, v_offset)

Analyze Image-level properties
- Indoor? Night scene? Sunset/sunrise?

Analyze region-level properties

Classify Challenging objects
- Buildings? Food? Flowers?

XML-like metadata file  [continue]  DEMO
Face detector

• OpenCV implementation of VJ’s face detection algorithm plus a skin-model

Evaluation

9 albums, 1669 photos (820 w/ faces)
Image-level ground truth data

<table>
<thead>
<tr>
<th>Image</th>
<th>dt</th>
<th>fp</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMG_0001.jpg</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>IMG_0002.jpg</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>IMG_0003.jpg</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>IMG_0004.jpg</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

\[
dt: \text{how many percent of face images are detected}
\]
\[
fp: \text{how many percent of non-face images are detected}
\]
Finding dominant colors

- Based on MPEG-7 Dominant Color Descriptor
- Modify the score of each dominant color

\[
\text{Score}(c_i) = w_1 \times \text{proportion}(c_i) \\
+ w_2 \times \text{compactness}(c_i) \\
+ w_3 \times \text{saturation}(c_i)
\]

- Decide 23 “standard colors” based on statistics of >12k photos
Image-level analysis

The design of features is the key.
Image-level analysis

- Indoor/Outdoor classification
  - Amount of blue ↓
  - Amount of green ↓
  - Degree of vertical change in brightness ↓
  - Degree of vertical orientation ↑

- Sunset / Sunrise
  - Amount of blue ↑
  - Amount of green ↑
  - Degree of vertical change in brightness ↑
  - Degree of vertical orientation ↓

Color histograms have specific distributions

Use MPEG-7 Scalable Color Descriptors for classification
Region-level analysis
sky, snow, grass, trees and waves

- Patterns might appear anywhere in the image
- A sliding detector is used to find a particular pattern
- A combination of MPEG-7 color, edge and texture descriptors is used as features
## Training Sample Collections

<table>
<thead>
<tr>
<th>Object name</th>
<th># of pos</th>
<th># of neg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indoor</td>
<td>104</td>
<td>124</td>
</tr>
<tr>
<td>Night scene</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Sunset / sunrise</td>
<td>342</td>
<td>342</td>
</tr>
<tr>
<td>Sky</td>
<td>5343</td>
<td>5342</td>
</tr>
<tr>
<td>Snow</td>
<td>1109</td>
<td>1364</td>
</tr>
<tr>
<td>Trees</td>
<td>1203</td>
<td>1202</td>
</tr>
<tr>
<td>Waves</td>
<td>1532</td>
<td>1723</td>
</tr>
<tr>
<td>Grass</td>
<td>1010</td>
<td>1367</td>
</tr>
<tr>
<td>Flowers</td>
<td>560</td>
<td>560</td>
</tr>
<tr>
<td>Food</td>
<td>290</td>
<td>290</td>
</tr>
<tr>
<td>Buildings</td>
<td>512</td>
<td>512</td>
</tr>
</tbody>
</table>
### Performance Evaluation

<table>
<thead>
<tr>
<th>Detector name</th>
<th>Performance statement</th>
<th>Avg. computation time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td>hit: 82.26%, fp: 10.51%</td>
<td>1.0254 s</td>
</tr>
<tr>
<td>Exif</td>
<td>N/A</td>
<td>0.0010 s</td>
</tr>
<tr>
<td>Dominant colors</td>
<td>N/A</td>
<td>0.7843 s</td>
</tr>
<tr>
<td>Indoor scene</td>
<td>classification: 90.35%</td>
<td>0.1032 s</td>
</tr>
<tr>
<td>Night scene</td>
<td>classification: 88.50%</td>
<td>0.3286 s</td>
</tr>
<tr>
<td>Sunset / sunrise</td>
<td>classification: 68.71%</td>
<td></td>
</tr>
<tr>
<td>Sky</td>
<td>classification: 96.46%</td>
<td></td>
</tr>
<tr>
<td>Snow</td>
<td>classification: 65.43%</td>
<td></td>
</tr>
<tr>
<td>Trees</td>
<td>classification: 66.88%</td>
<td></td>
</tr>
<tr>
<td>Waves</td>
<td>classification: 82.43%</td>
<td></td>
</tr>
<tr>
<td>Grass</td>
<td>classification: 75.52%</td>
<td></td>
</tr>
</tbody>
</table>

- Performance is analyzed by 5-fold cross validation on all training samples
- CPU P4 3.2GHz, 2GB RAM, Debug mode (not optimized code)
  (When the program is running, CPU usage: 60%±9%, memory usage: 350MB±11MB)
Online demo (http://lbmm.ece.ucsb.edu/php/photo/)
Agenda

• Introduction

• Image Categorization (*ACM MM 06*)

• Media Content Management
  – Tag Creation: A Tagging Tool for Personal Photos
  – Tag Organization: Faceted Metadata
  – Tag Enhancement: Cues from different sources

• Summary
Tagging: the goal, or just the means?
The problem of hierarchy

- Inflexible
- Wasteful
- Difficult to modify

2005 >> CANON >> PEOPLE >> SNOW...
CANON >> 2005 >> SKY >> PEOPLE...

2005/12/20 13:11:42
Faceted metadata

Goal: help users move through large information spaces without feeling lost

• More than one ways to describe one item

Keywords for the recipe:
Thai, Stir-fry, Main course, Bell Pepper, Curry

What are the facets and labels for photos?
Album name: ICIP @ Genova

Metadata example:

```
<RES>IMG_0001.jpg</RES>
<item>
<album>ICIP @ Genova</album>
<date>20050912 13:27:38</date>
<camera make>Canon</camera>
<ob conf=0.6>architecture</ob>
<ob conf=0.2>waterside</ob>
<face conf=0.86>Mei</face>

......
</item>
```
Agenda

• Introduction
• Image Categorization (ACM MM 06)
• Media Content Management
  – Tag Creation: A Tagging Tool for Personal Photos
  – Tag Organization: Faceted Metadata
  – Tag Enhancement: Cues from different sources
• Summary
Quality enhancement for tags

Context

Content

Spatial context

Temporal context

Community

Sky
Mountain
Snow

Whistler, Canada

Human-generated tags
Clicking behaviors

Time
Summary

- Embedded text can help image categorization
  - Fusing text and low-level visual cues boosts the categorization accuracy
- We are developing a tagging tool: a starting point for the process of automatic machine tagging
- Faceted metadata can be used to organize tags
- Future work: Fusing cues from content, context and community to enhance tag quality
• Online demo: [http://lbmm.ece.ucsb.edu/php/photo/](http://lbmm.ece.ucsb.edu/php/photo/)
• Check our website: [http://lbmedia.ece.ucsb.edu/](http://lbmedia.ece.ucsb.edu/) for project details, datasets, and more resources