Food Recognition Using Statistics of Pairwise Local Features

Shulin Yang
University of Washington

Mei Chen, Dean Pomerleau, Rahul Sukthankar
Intel Labs Pittsburgh
Motivation

1. A challenging computer vision problem compared to other forms of object recognition:
   • Large variation in appearance
   • Deformable internal structure

2. Useful for health and wellness applications such as diet monitoring, nutrition research, obesity control.
Challenges for Food Recognition

Main challenges of food recognition

• How to identify food ingredients in images

• How to represent and quantify the highly varied spatial relationships of the ingredients
Related Work

Bolle, et al. (1996): VeggieVision- an automatic produce ID to help with the produce checkout process

Shroff, et al. (2008): A wearable computing system to recognize food for calorie monitoring

Yang, et al. (2009): PFID- a dataset for food recognition

Wu and Yang (2009): Analyzed eating videos to recognize food 160 food items and estimate caloric intake
Our Framework

Food Images → Label Image Pixels → Represent Geometric Relationship → Classify Images
Our Framework

1. Label Image Pixels
2. Represent Geometric Relationship
3. Classify Images

- Food Images
- Label Image Pixels
- Represent Geometric Relationship
- Classify Images
Our Framework

Label Image Pixels → Represent Geometric Relationship → Classify Images

Food Images
Labeling Image Pixels

- Nine food ingredients: bread, beef, chicken, pork, vegetable, tomato, cheese, egg, others
- Soft Labels: a vector of likelihoods that a pixel belongs to each food ingredient type

- Soft labels generated using Semantic Texton Forest (STF) algorithm (Shotton, et al. CVPR2008)
Our Framework

1. Food Images
2. Label Images
3. Represent Geometric Relationship
4. Classify Images
Our Framework

1. Food Images
2. Represent Geometric Relationship
3. Classify Images
Image Representation – Pairwise Features

- Using pairwise feature distribution to capture the spatial relationship among pixels of food ingredients
Image Representation – Pairwise Features

- Using pairwise feature distribution to capture the spatial relationship among pixels of food ingredients
Image Representation – Pairwise Features

- Using pairwise feature distribution to capture the spatial relationship among pixels of food ingredients
Image Representation – Pairwise Features

- Using pairwise feature distribution to capture the spatial relationship among pixels of food ingredients
Distance

Orientation

Midpoint category

Between-pair categories

Distance:
P2 (cheese pixel)
P1 (bread pixel)

Orientation:
P2 (cheese pixel)
P1 (bread pixel)

Midpoint category:
P2 (cheese pixel)
P1 (bread pixel)

Between-pair categories:
P2 (bread pixel)
P1 (bread pixel)

Image Representation – Pairwise Features
Geometric Relationship Representation

Tomato-Bread Histogram

Pair Counts

Log Distance

0 1 2 3 4 5 6 7
Geometric Relationship Representation

Tomato-Bread Histogram

Pair Counts

Log Distance
Geometric Relationship Representation
Geometric Relationship Representation

Tomato-Bread Histogram

Beef-Cheese Histogram

Pair Counts

Log Distance

0 1 2 3 4 5 6 7

7
Geometric Relationship Representation

Tomato-Bread Histogram

Beef-Cheese Histogram

Beef-Bread Histogram
Geometric Relationship Representation

- Tomato-Bread Histogram
- Beef-Cheese Histogram
- Beef-Bread Histogram
- Tomato-Cheese Histogram

Pair Counts

Log Distance

Pair Counts

0 1 2 3 4 5 6 7
Our Framework

Food Images

Images Represent Geometric Relationship

Classify Images
Classification

• SVM using pre-computed $\chi^2$ kernel

\[ d_{\chi^2}(x, y) = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i} \]

• Image dissimilarity is the sum of $\chi^2$ distances of all their histograms
The Complete Framework

1. **Food Images**
   - Images of food items like a burger, chicken, vegetable, and bread.

2. **Image Pixel Soft Labels**
   - Soft labels for the image content:
     - Original
     - Chicken
     - Vegetable
     - Bread

3. **Pairwise Local Feature Distribution**
   - Graph showing pairwise feature distribution for:
     - Tomato-Bread
     - Beef-Cheese
     - Beef-Bread
     - Tomato-Cheese

4. **Image Classification**
   - Result: **Chicken Club Salad**
Experiment – Dataset

Dataset: PFID (Pittsburgh Fast-Food Image Dataset)

• Available at [http://pfid.intel-research.net/baseline.php](http://pfid.intel-research.net/baseline.php)
• Fast food from 13 chain restaurants, such as Spicy Chicken Sandwich from Wendy’s, and Italian B.M.T. Sandwich from Subway
• 61 categories × 3 instances × 6 views = 1098 in total
• Background segmented
Experiment – Soft Labels

Soft label all images using STF

• Train on 16 manually labeled images

• Generate soft labels for the other 1098 images
Experiment – Classification

Training and testing data
• 732 (61×2×6) images for training
• 366 (61×1×6) images for testing
• Three-fold cross-validation

Two classification tasks
• On 61 food categories
• On 7 major groups of food

<table>
<thead>
<tr>
<th>Sandwich</th>
<th>Salad</th>
<th>Taco</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicken</td>
<td>Bread</td>
<td>Donut</td>
</tr>
<tr>
<td>Bagel</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Classification Results on 61 Categories

Accuracy

- Color Hist: 11.3%
- SIFT: 9.2%
- Ingre Hist: 18.9%
- Pair Dist: 19.2%
- Pair Orien: 20.8%
- Pair Mid: 22.6%
- Pair In-betw: 21.3%
- Dis+Ori: 28.2%
- Ori+Mid: 28.4%
- Dis+Ori+Mid: 28.4%

Baselines
Classification Results on 61 Categories

Accuracy

<table>
<thead>
<tr>
<th>Baselines</th>
<th>Color Hist</th>
<th>SIFT</th>
<th>Ingre Hist</th>
<th>Pair Dist</th>
<th>Pair Orien</th>
<th>Pair Mid</th>
<th>Pair In-btw</th>
<th>Dis+Ori</th>
<th>Ori+Mid</th>
<th>Dis+Ori+Mid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11.3%</td>
<td>9.2%</td>
<td>18.9%</td>
<td>19.2%</td>
<td>20.8%</td>
<td>22.6%</td>
<td>21.3%</td>
<td>21.2%</td>
<td>28.2%</td>
<td>28.4%</td>
</tr>
</tbody>
</table>

Chance 1.6%
Classification Results on 61 Categories

Accuracy

<table>
<thead>
<tr>
<th>Baselines</th>
<th>Color Hist</th>
<th>SIFT</th>
<th>Ingre Hist</th>
<th>Pair Dist</th>
<th>Pair Orien</th>
<th>Pair Mid</th>
<th>Pair In-btw</th>
<th>Dis+Ori</th>
<th>Ori+Mid</th>
<th>Dis+Ori+Mid</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.3%</td>
<td>9.2%</td>
<td>18.9%</td>
<td>19.2%</td>
<td>20.8%</td>
<td>22.6%</td>
<td>21.3%</td>
<td>21.2%</td>
<td>28.2%</td>
<td>28.4%</td>
<td></td>
</tr>
</tbody>
</table>

Chance: 1.6%
Classification Results on 7 Major Groups

<table>
<thead>
<tr>
<th>Baselines</th>
<th>Color Hist</th>
<th>SIFT</th>
<th>Ingre Hist</th>
<th>Pair Dist</th>
<th>Pair Orien</th>
<th>Pair Mid</th>
<th>Pair In-betw</th>
<th>Dis+Ori</th>
<th>Ori+Mid</th>
<th>Dis+Ori+Mid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>49.7%</td>
<td>55.3%</td>
<td>69.0%</td>
<td>69.7%</td>
<td>71.0%</td>
<td>74.3%</td>
<td>73.8%</td>
<td>72.2%</td>
<td>78.0%</td>
<td>78.7%</td>
</tr>
<tr>
<td>Chance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Pairwise Features for Scene Recognition

Scene images are similar to the food images:
- Composed of some basic scene “ingredients”
- Geometric relationships between ingredients important but highly variable

Test on the MIT outdoor scene dataset
8 categories including coast, mountain, highways, etc.
Pairwise Feature for Scene Recognition

• 10 basic ingredients for outdoor scenes: sky, water, rocks, trees, road, building, etc. Manually labeled 24 images for soft labeling training

• Combine the Pairwise features (Distance and Orientation) with GIST features (Oliva and Torralba, IJCV2001).

• Preliminary result shows that the accuracy is boosted by 3% from 85% to 88%
Conclusion

1. Developed new technique to capture shape regularities using geometric statistics computed over pairwise features

2. Applied the technique to food recognition problem

3. Outperformed baseline approaches significantly on food recognition

Future work:
- Design and test new pairwise feature
- Generalize the approach for other domains
- Use it for real food recognition applications
Acknowledgements

We would like to thank:

Professor Linda Shapiro and the Multimedia Group in UW for suggestions and advice,

Lei Yang for providing SIFT+SVM baseline code,

The PFID project for help with the food recognition dataset.
Q&A?