Image-Based Estimation of Real Food Size for Accurate Food Calorie Estimation

Takumi Ege, Yoshikazu Ando, Ryosuke Tanno, Wataru Shimoda and Keiji Yanai Department of Informatics, The University of Electro-Communications, Tokyo
Introduction

- **Food calorie estimation**
- Existing works
  - Associated with the estimated food categories
  - Require users to enter information such as size or volume
- Our works
  - Fully-automatic food calorie estimation
Review of our works

• (1) “CalorieCam”
  – reference-object-based food calorie estimation system.

• (2) Weakly-supervised segmentation based food calorie estimation

• (3) “AR DeepCalorieCam V2”
  – A real food size and calorie estimation system based on iOS ARKit.

• (4) “DepthCalorieCam” (new system)
  – food calorie estimation system exploiting iPhone stereo cameras.

• (5) Rice grain based size estimation (new method)
  – uses rice grained the size of which are usually almost the same as a reference object.
CalorieCam

• Image-based calorie estimation system
• User needs to register a size-known reference object
  – Wallet
  – Creditcard-size card
CalorieCam

• Segmentation
  – Food items
  – Pre-registered reference object.

• Real size of each detected food items
  – Comparing the number of pixels

• Process
  – Estimate rough position based on edge detection
  – Apply color-pixel-based k-means clustering
  – Apply GrabCut with the detected bounding box
Weakly-supervised Segmentation Based Calorie Estimation

• Calorie Cam
  – One meal photo contains only one dish
• This work
  – Estimate calories from segmentation results.
  – Without multiple-view photos
  – Without specific reference objects such as wallets and cards.
Weakly-supervised Segmentation Based Calorie Estimation

(1) Selective Search
(2) BB clustering
(3) Back propagation
(4) Saliency maps
(5) Segmentation by GrabCut
(6) NMS

Final result

Ⓒ 2017 UEC Tokyo.
Weakly-supervised Segmentation Based Calorie Estimation

• Choose base food region
• We decide priorities based on a tendency of unchanging food volumes.
• Some food volumes change frequently, while some foods volume rarely change.
• In “Teishoku”
  – Japanese traditional food combo menu
  – we can often change the volume of “rice”
  – we cannot change “miso-soup” volume in general.
<table>
<thead>
<tr>
<th>Meal</th>
<th>Components</th>
<th>Total kcal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Meal</td>
<td>rice, miso soup, potage, steamed egg, green salad</td>
<td>159.9 kcal, 21.8 kcal, 54.5 kcal, 111.4 kcal, 45.6 kcal, 131.0 kcal, 524.2 kcal</td>
</tr>
<tr>
<td>2nd Meal</td>
<td>rice, miso soup, potage, seasoned beef</td>
<td>290.9 kcal, 30.0 kcal, 130.3 kcal, 418.2 kcal, 869.4 kcal</td>
</tr>
<tr>
<td>3rd Meal</td>
<td>pork cutlet, pork miso</td>
<td>817.2 kcal, 167.0 kcal, 984.2 kcal</td>
</tr>
<tr>
<td>4th Meal</td>
<td>hamburger, french fries</td>
<td>369.7 kcal, 93.0 kcal, 462.7 kcal</td>
</tr>
<tr>
<td>5th Meal</td>
<td>hamburger, french fries</td>
<td>716.7 kcal, 93.0 kcal, 809.7 kcal</td>
</tr>
</tbody>
</table>
AR DeepCalorieCam

• Uses an inertial sensor built in a standard smartphone in addition to photos

• Recognize a category of each food item.
• Directly calculate the size using AR
• Calculate food calories based on their
AR DeepCalorieCam
AR DeepCalorieCam
with iOS ARKit
ARKit estimates real size of the bounding box.
DepthCalorieCam

An obtained RGB image.  An obtained depth image.  Estimated result with calorie intake and segmented food region.
Croquette calorie estimation demo:

1 croquette = 160 kcal

2 croquettes = 320 kcal
DepthCalorieCam

• Food regions are extracted by U-NET trained with UECFood-100 and segmentation mask sets.
• Estimate the depth of each of the pixels and their actual volumes.
• Calculate food calories from the actual volume and the regression curves
DepthCalorieCam

• CNN-based food region segmentation
  – Network
    • U-Net
  – Training
    • 5,301 masks
    • UECFOOD-100
Comparative Experiments with CalorieCam and AR DeepCalorieCam V2

- Target

<table>
<thead>
<tr>
<th>category</th>
<th>calorie [kcal]</th>
</tr>
</thead>
<tbody>
<tr>
<td>pork with sweet and sour source</td>
<td>500</td>
</tr>
<tr>
<td>fried chicken</td>
<td>655</td>
</tr>
<tr>
<td>croquette</td>
<td>246</td>
</tr>
</tbody>
</table>
Comparative Experiments with CalorieCam and AR DeepCalorieCam V2

Table 2. Comparison on calorie estimation error (Avg. ± SD [kcal]) among CalorieCam [1], AR CalorieCam V2 [3] and DepthCalorieCam.

<table>
<thead>
<tr>
<th>category</th>
<th>CalorieCam</th>
<th>AR CalorieCam</th>
<th>DepthCalorieCam</th>
</tr>
</thead>
<tbody>
<tr>
<td>pork with source</td>
<td>364±552</td>
<td>-112±163</td>
<td>2±52</td>
</tr>
<tr>
<td>fried chicken</td>
<td>-123±171</td>
<td>343±51</td>
<td>-5±64</td>
</tr>
<tr>
<td>croquette</td>
<td>-48±16</td>
<td>-104±12</td>
<td>-35±22</td>
</tr>
</tbody>
</table>
Rice grain based size estimation

• Proposed network

• Examples of dataset
Dataset

- 360 images

<table>
<thead>
<tr>
<th>Amount of water</th>
<th>Small amount</th>
<th>Medium amount</th>
<th>Large amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera</td>
<td>180ml/150g</td>
<td>200ml/150g</td>
<td>220ml/150g</td>
</tr>
<tr>
<td>COOLPIX AW120</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>3264x2448</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>iPhone8 Plus</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>4032x3024</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
</tbody>
</table>
Experiments

• Training

• Evaluation
Experiments

<table>
<thead>
<tr>
<th>Evaluation data.</th>
<th>abs. err. (cm/224pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera: COOLPIX, Small amount of water</td>
<td>0.212</td>
</tr>
<tr>
<td>Camera: COOLPIX, Medium amount of water</td>
<td>0.178</td>
</tr>
<tr>
<td>Camera: COOLPIX, Large amount of water</td>
<td>0.197</td>
</tr>
<tr>
<td>Camera: iPhone8 Plus, Small amount of water</td>
<td>0.127</td>
</tr>
<tr>
<td>Camera: iPhone8 Plus, Medium amount of water</td>
<td>0.170</td>
</tr>
<tr>
<td>Camera: iPhone8 Plus, Large amount of water</td>
<td>0.105</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rel. err. (%)</th>
<th>corr.</th>
<th>≤ 5% err. (%)</th>
<th>≤ 10% err. (%)</th>
<th>≤ 20% err. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.182</td>
<td>0.958</td>
<td>41.667</td>
<td>75.000</td>
<td>91.667</td>
</tr>
<tr>
<td>6.550</td>
<td>0.973</td>
<td>43.333</td>
<td>76.667</td>
<td>93.333</td>
</tr>
<tr>
<td>6.668</td>
<td>0.962</td>
<td>48.333</td>
<td>78.333</td>
<td>90.000</td>
</tr>
<tr>
<td>5.652</td>
<td>0.945</td>
<td>50.000</td>
<td>75.000</td>
<td>98.333</td>
</tr>
<tr>
<td>7.512</td>
<td>0.903</td>
<td>43.333</td>
<td>68.333</td>
<td>88.333</td>
</tr>
<tr>
<td>4.800</td>
<td>0.967</td>
<td>58.333</td>
<td>88.333</td>
<td>98.333</td>
</tr>
</tbody>
</table>
Conclusion

• DepthCalorieCam is the most promising approach.

• Large-scale calorie annotated 3D food volume data is needed to extend the system
  — very costly and time-consuming.

• The rice grain based method is also promising
  — Appropriate for Japanese foods