Real-time Mobile Food Recognition System

Yoshiyuki Kawano and Keiji Yanai
Department of Informatics,
The University of Electro-Communications,
Tokyo, JAPAN

ⓒ 2013 UEC Tokyo.

Background

- Recording of food habits has become popular.
  - by using PC or mobile phone
  - quite troublesome task.

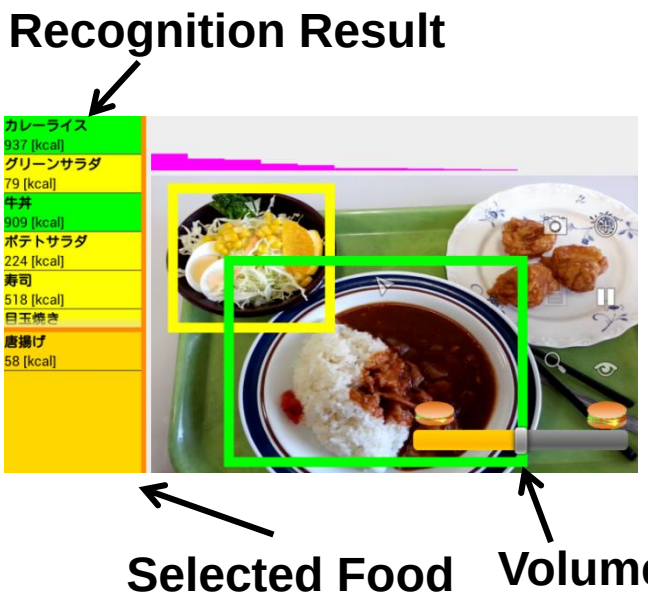
- Rapid progress of smartphone
  - obtain enough computational power for real-time image recognition

ⓒ 2013 UEC Tokyo.

Objective

Mobile Food Recognition System

- recognize food in real-time
- require no connection with serve
  - Recognition on a smartphone
- interactive system
  - bounding box is given by user

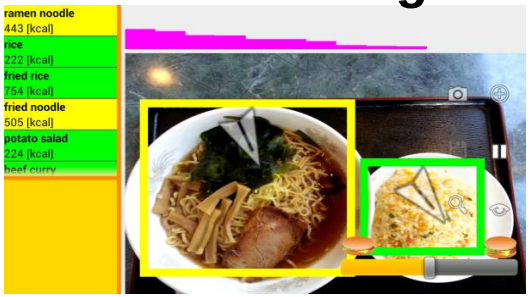


ⓒ 2013 UEC Tokyo.

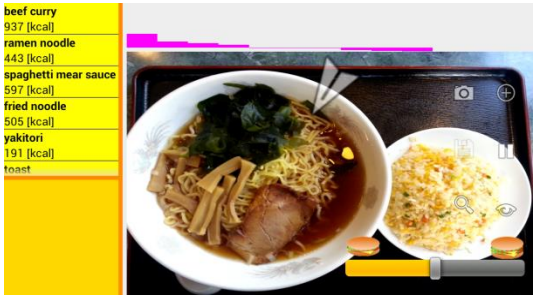
Flow of proposed system

Point a smartphone toward food items

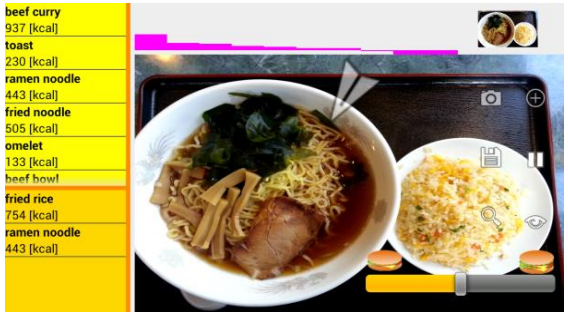
Draw bounding box



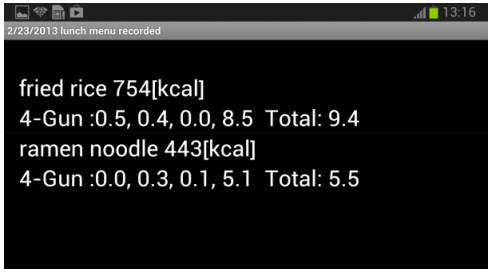
Real-time recognition



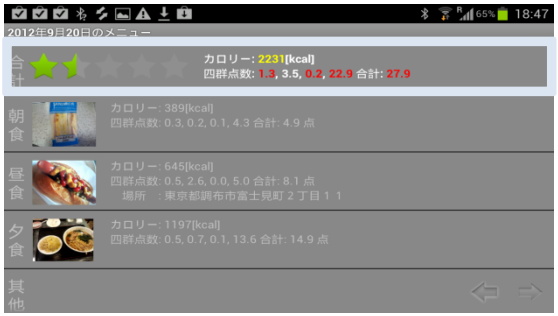
Select food items



Nutrition information Register food record



Check balance



ⓒ 2013 UEC Tokyo.

Demonstration

ⓒ 2013 UEC Tokyo.

Related Work (1)

- [Matsuda et al. ICME'12]
  - 100 kinds of food categories
  - carry out on server side

Food Image

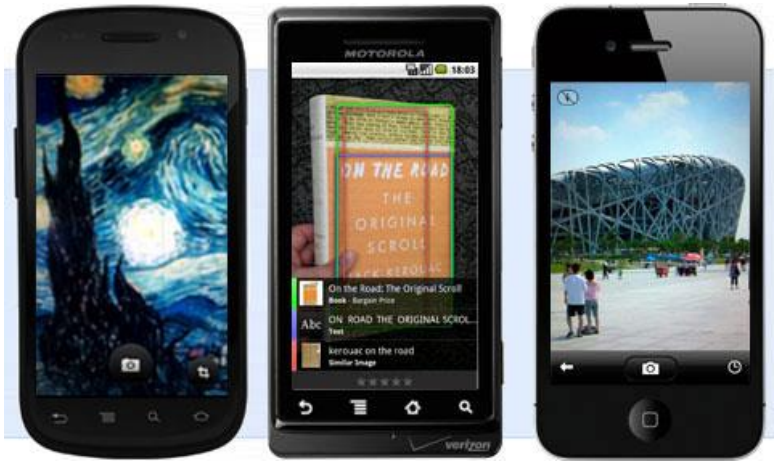
Result

1. Rice
2. miso soup
3. Egg sunny-side up
4. Sirloin cutlet
5. Salmom meuniere
6. Fried fish
7. Boiled fish
8. Sausage
9. Sandwiches
10. Roll bread

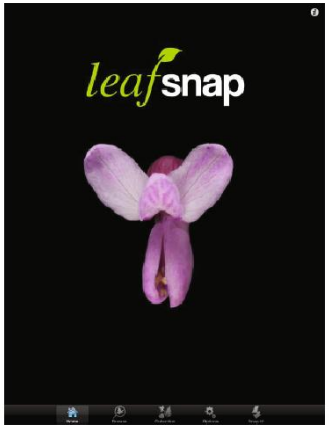
ⓒ 2013 UEC Tokyo.

Related Work (2)

- Google Goggles
  - Specific object recognition
  - Similar image search
  - OCR



- Leaf snap(Kumar et al,ECCV'12)
  - Identifying plant species



All the systems adopt server-side recognition

ⓒ 2013 UEC Tokyo.

3 key features of the system

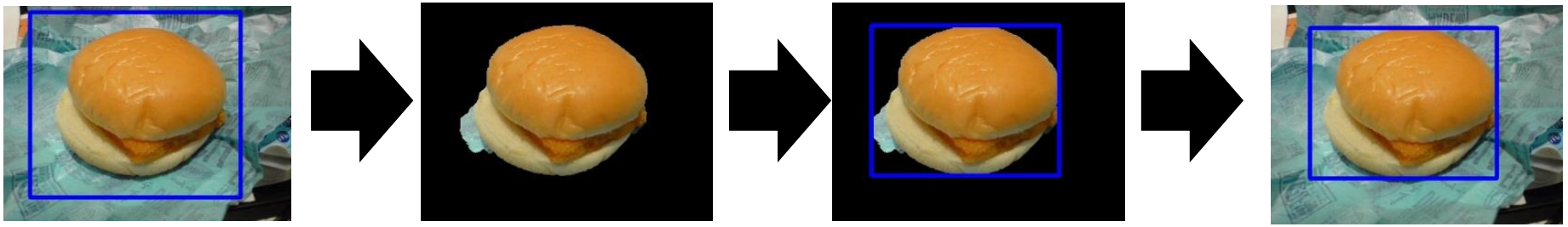
- Bounding box adjustment
- Real-time image recognition
- Estimation of the more reliable direction

ⓒ 2013 UEC Tokyo.

Bounding Box Adjustment

- Restriction : A user draws bounding boxes covering food items by touching

- Performed only once after bounding box was drawn

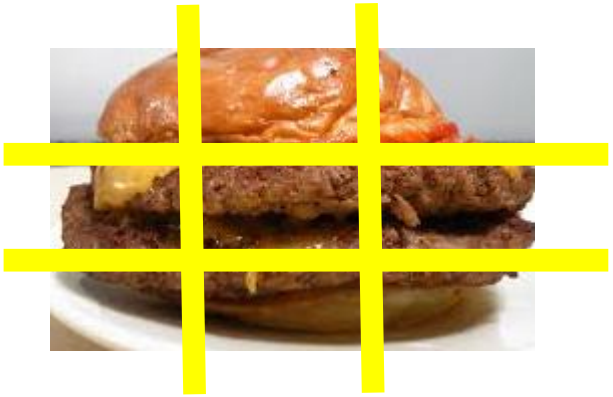


GrabCut

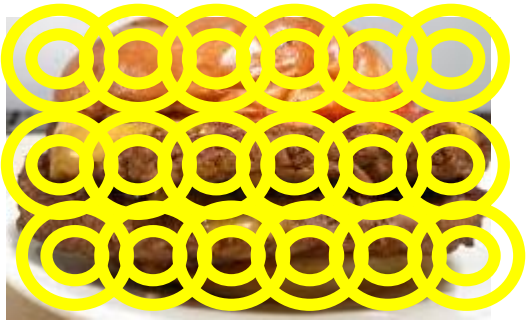
ⓒ 2013 UEC Tokyo.

Food Recognition

- Image Features:
  - Color Histogram
    - $64 \times (3 \times 3) = 576$dim



  - SURF-BoF
    - code book size : 500
    - soft assignment



ⓒ 2013 UEC Tokyo.

Food Recognition

- Classifier:
  - Linear SVM

$$f(x) = \sum_{i=1}^{M} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b$$

$$= \sum_{i=1}^{M} y_i \alpha_i \langle \mathbf{x}, \mathbf{x}_i \rangle + b$$

$$= \langle \sum_{i=1}^{M} y_i \alpha_i \, \mathbf{x}_i, \mathbf{x} \rangle + b$$

$$= \langle \mathbf{w}, \mathbf{x} \rangle + b$$

Independent of the number of samples

computation: $O(N)$

memory : $O(N)$

ⓒ 2013 UEC Tokyo.

Food Recognition

- fast χ^2 feature map(Vedaldi et al, PAMI'12)
  - approximate the implicit feature mapping
  - the same accuracy as non linear SVM
  - speed comparable with linear SVM

$$\emptyset(x) = \sqrt{x}\begin{bmatrix} 0.8 \\ 0.6\cos(0.6\log x) \\ 0.6\sin(0.6\log x) \end{bmatrix}$$

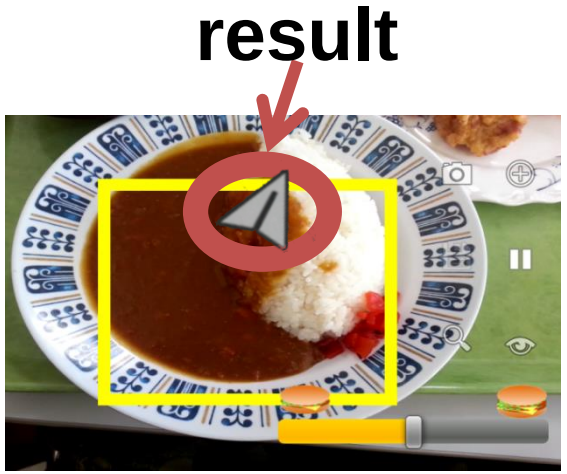
ⓒ 2013 UEC Tokyo.

Estimation of the more reliable direction

- Effective window search
- No transposed SURF-BoF + linear SVM
  - Linear SVM

    $$\mathbf{w} = \mathbf{w}^{+} + \mathbf{w}^{-}$$

  - SVM score
    - Create integral images about $\mathbf{w}^{+}$ and $\mathbf{w}^{-}$
    - $O(1)$operation (refer to ESS (Lampert et al, CVPR '08))



- Show the result
  - The window with the maximum score

ⓒ 2013 UEC Tokyo.

Demonstration of estimation the more reliable direction

Experiment

- Experiment description
  - Recognition accuracy
  - Processing time
  - User study

ⓒ 2013 UEC Tokyo.

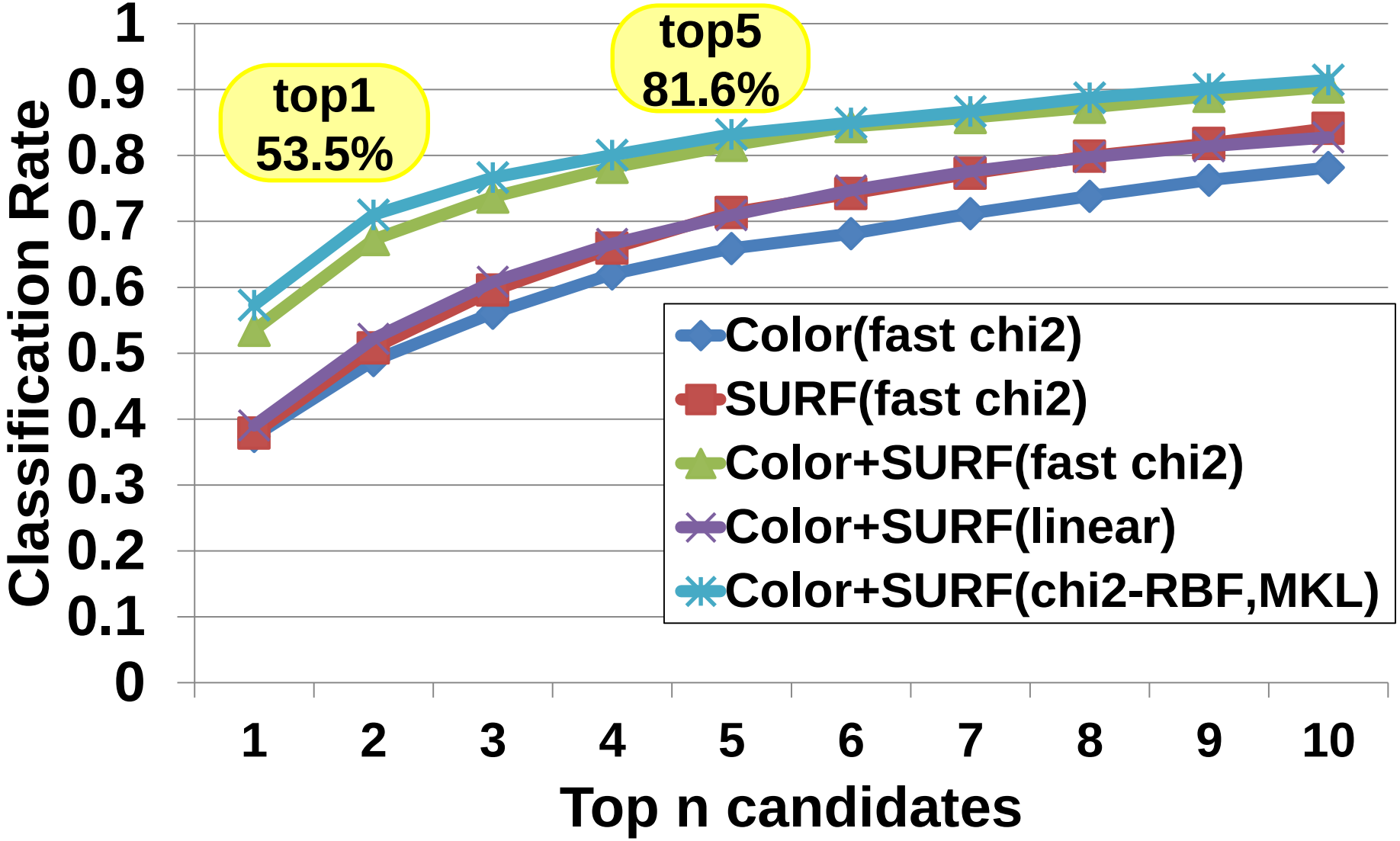
Set up

- Recognition accuracy
  - Dataset
    - includes 50 kinds of food categories
    - has more than 100 images for each category

- processing time
  - device
    - Galaxy Note2 (1.6 GHz,4 core, 4 thread)
  - multi thread for Quad core

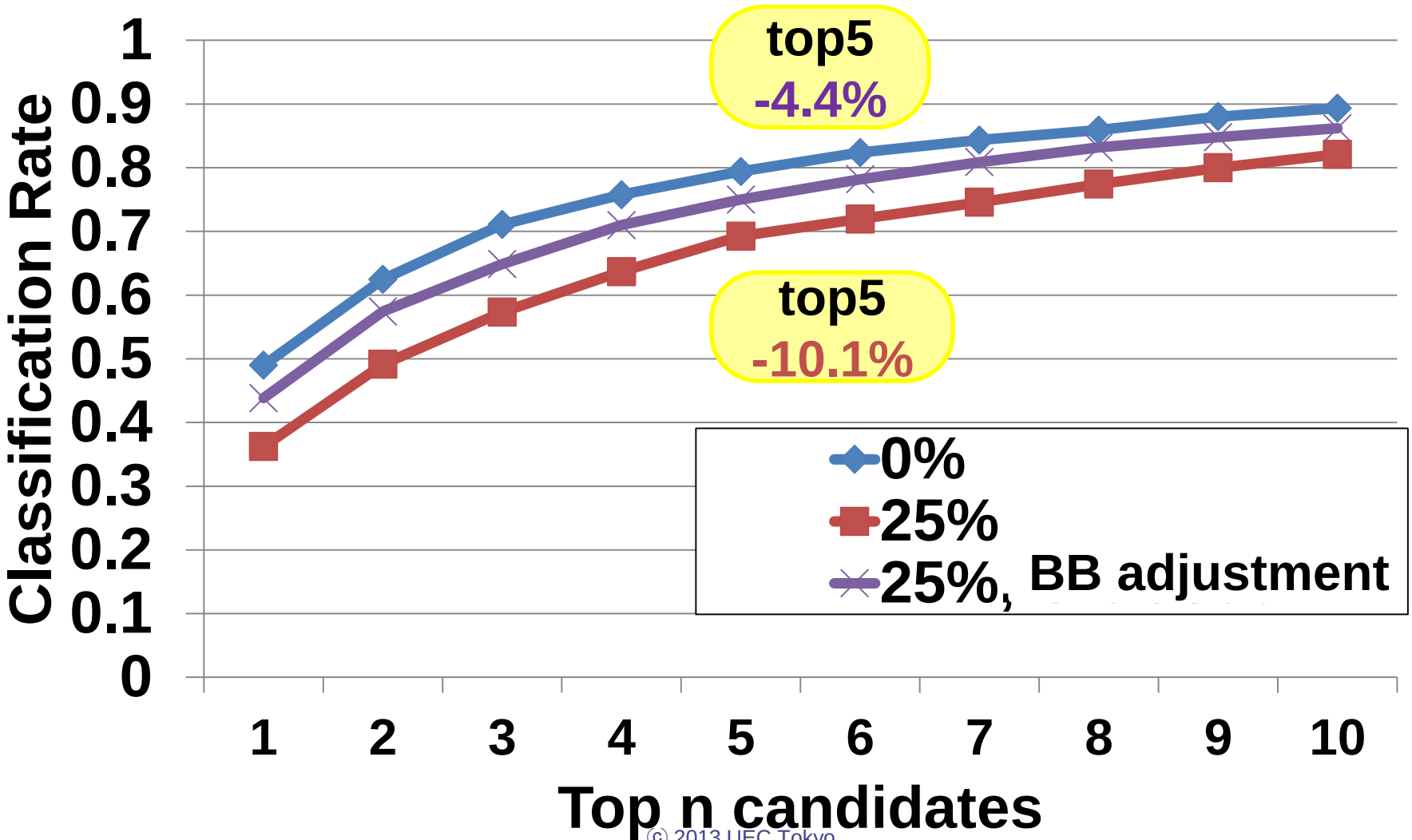
ⓒ 2013 UEC Tokyo.

Evaluation of Recognition Accuracy



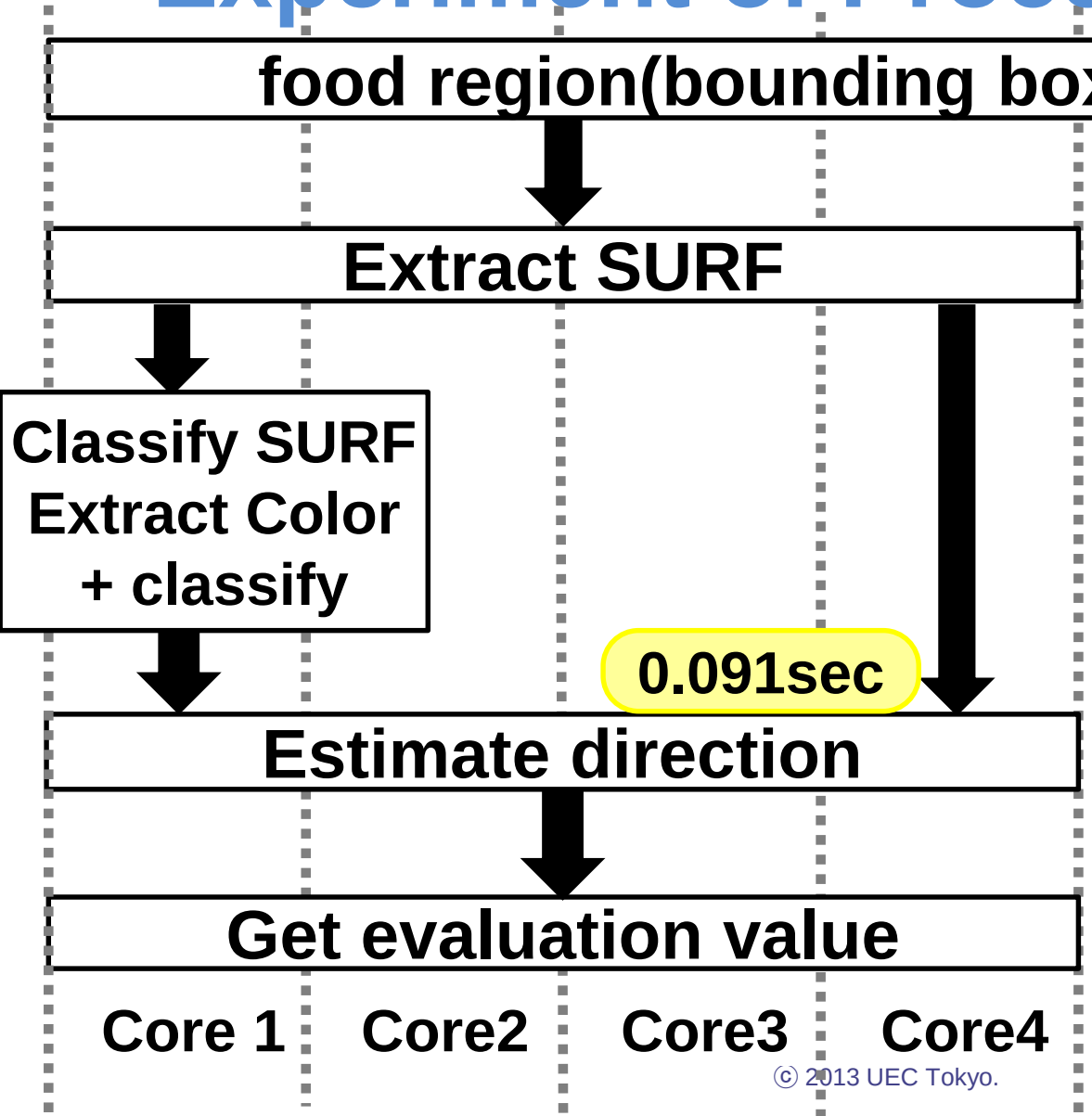
ⓒ 2013 UEC Tokyo.

Evaluation of Bounding box adjustment



© 2013 UEC Tokyo.

Experiment of Processing Time

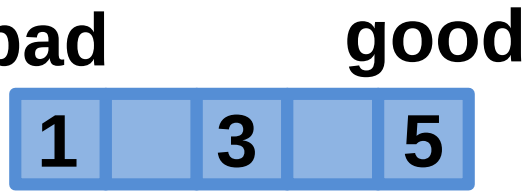


Recognition	+ Estimate direction
0.26 sec	0.34sec

ⓒ 2013 UEC Tokyo.

User Study

- Subject
  - Five students
- Evaluation
  - 3 food items in a meal, 2trial, 3 or4 meal
  - Outcome Measure (5step)
    - Recognition Accuracy
    - Usability
    - Estimation direction
    - Which is better, proposed or manual system?



ⓒ 2013 UEC Tokyo.

User Evaluation

- Result(average value)

Recognition Accuracy	3.4
Usability	4.2
Estimation direction	2.4
Which is better proposed or manual	3.8

ⓒ 2013 UEC Tokyo.

Comments from subjects

- comment
  - If accuracy is improved, I want to try it
  - exclude food items of incorrect result from list
  - increase the number of food categories

- summary
  - unable to recognize some food items
  - Proposed system better than manual system

ⓒ 2013 UEC Tokyo.

Conclusions

- recognize 50 kinds food categories
- run on common smartphone
- adjust bounding box in the background
- estimate of the more reliable direction

ⓒ 2013 UEC Tokyo.

Future Work

- draw bounding box
  - Make less complex such as touch just a point
- use additional information
  - User's meal history
  - GPS and time information

ⓒ 2013 UEC Tokyo.

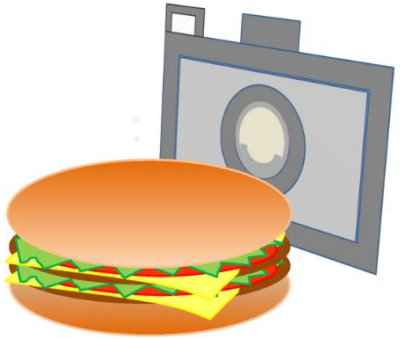
You can try it

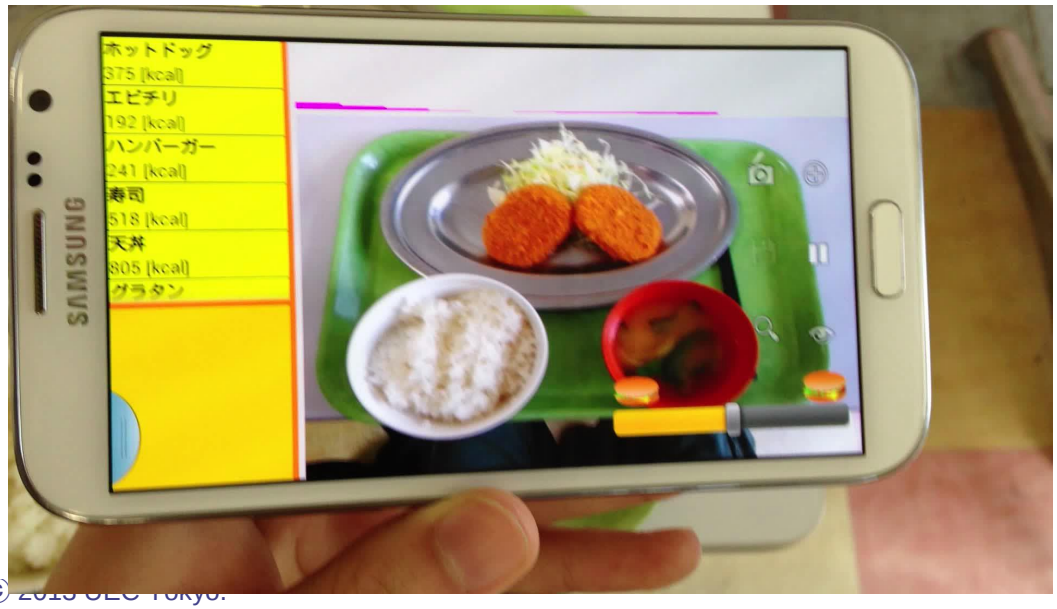
- **http://foodcam.mobi/**

Requirements : Android 4.0 and up

: Quad core CPU

(No iPhone version)





© 2013 UEC Tokyo.

ⓒ 2013 UEC Tokyo.

Meal Records

on the smartphone

on the Web

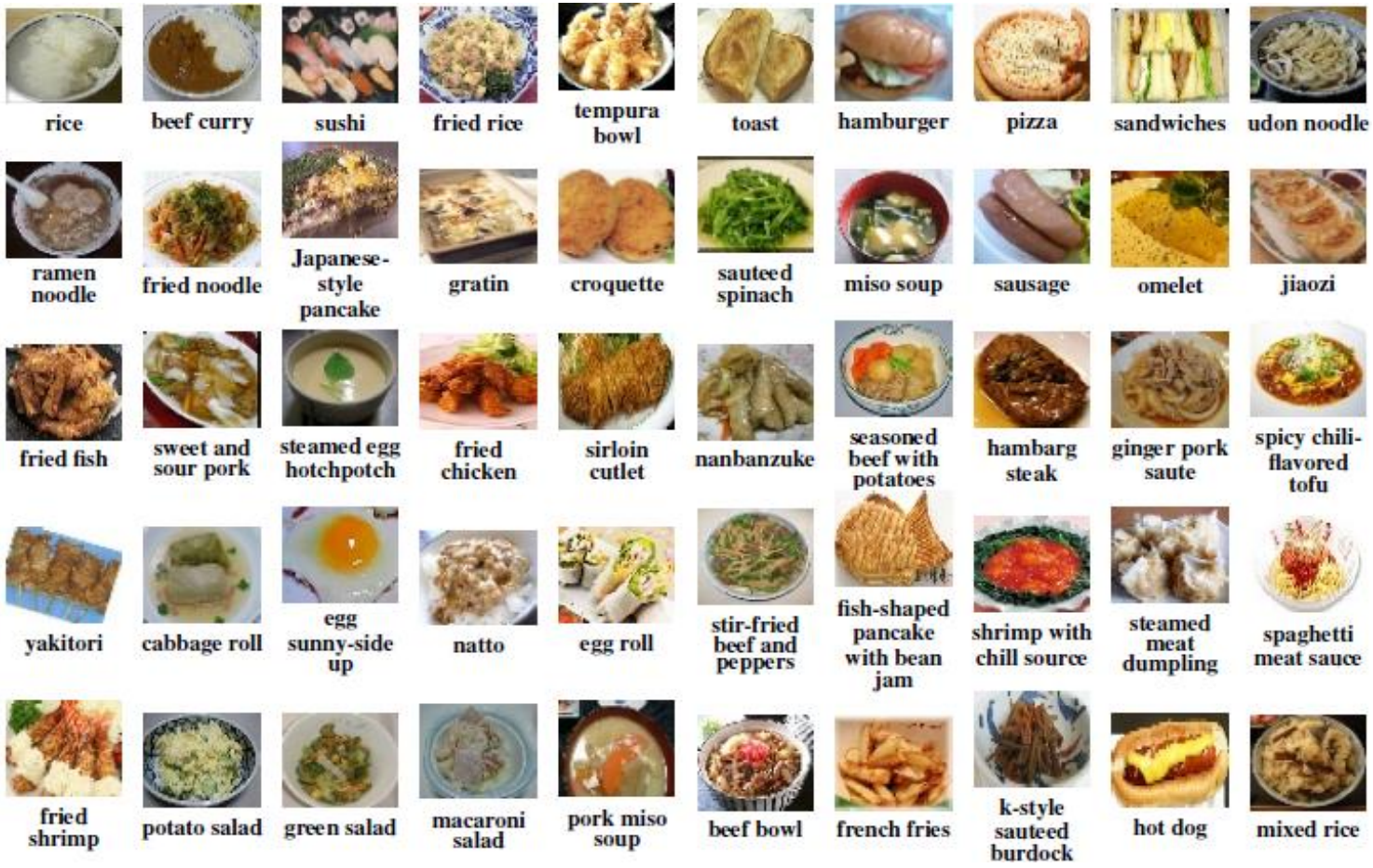
ⓒ 2013 UEC Tokyo.

4-Gun

ⓒ 2013 UEC Tokyo.

Data Set

- 50 kinds food items



ⓒ 2013 UEC Tokyo.

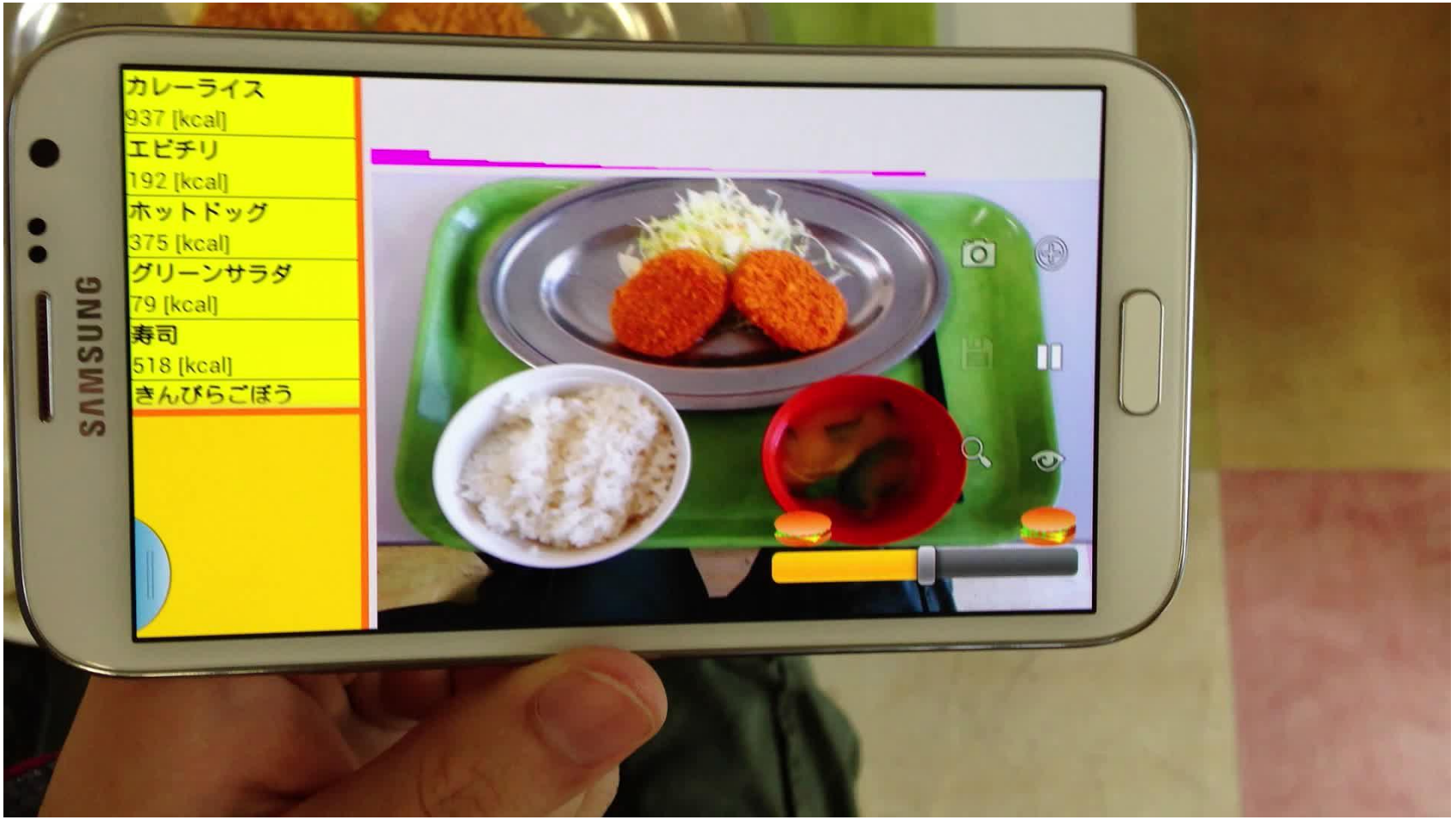
Demonstration

ⓒ 2013 UEC Tokyo.

Demonstration of estimaton

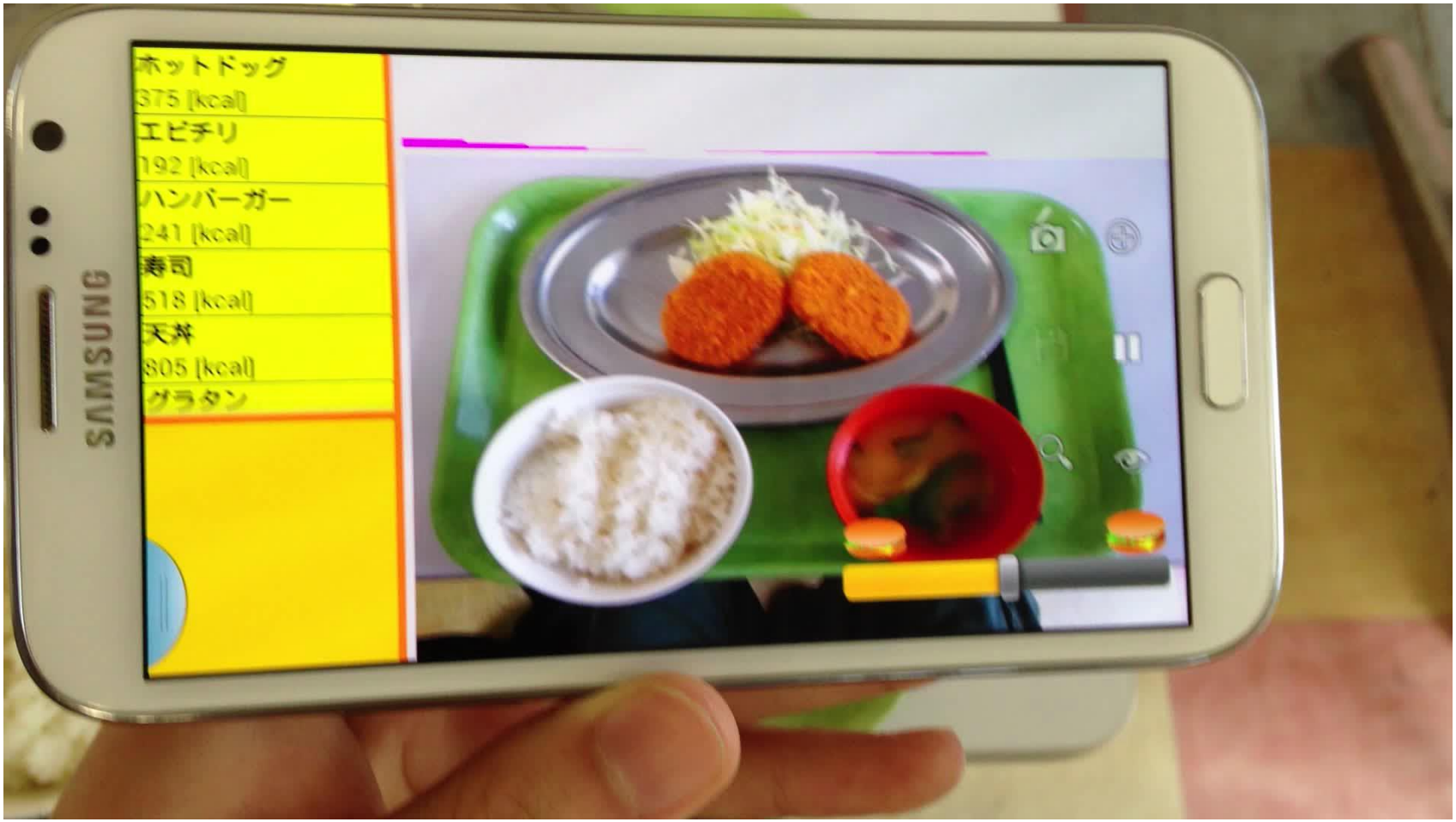
ⓒ 2013 UEC Tokyo.

Demonstration

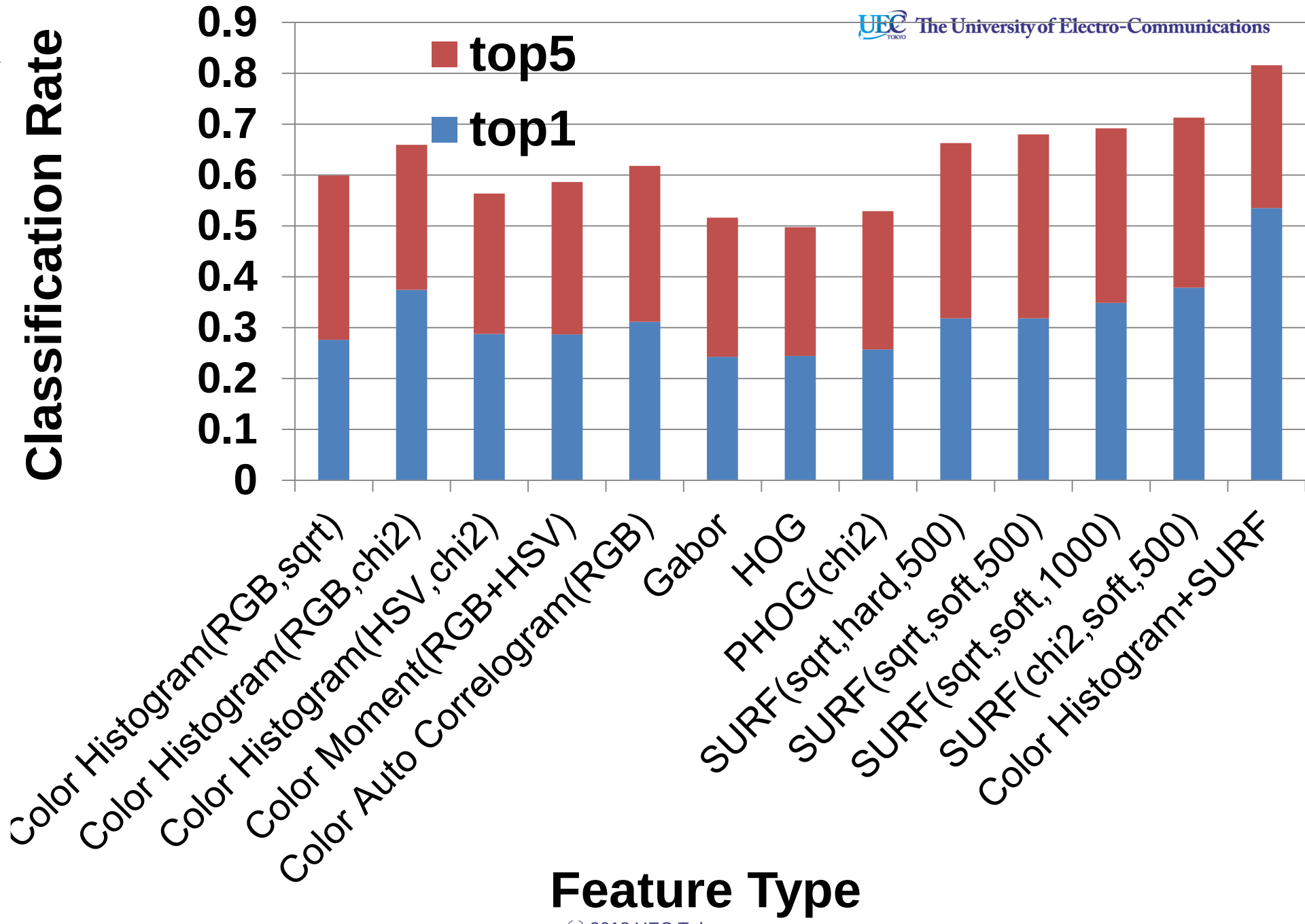


ⓒ 2013 UEC Tokyo.

Demonstration



ⓒ 2013 UEC Tokyo.



ⓒ 2013 UEC Tokyo.

Set up for Experiment

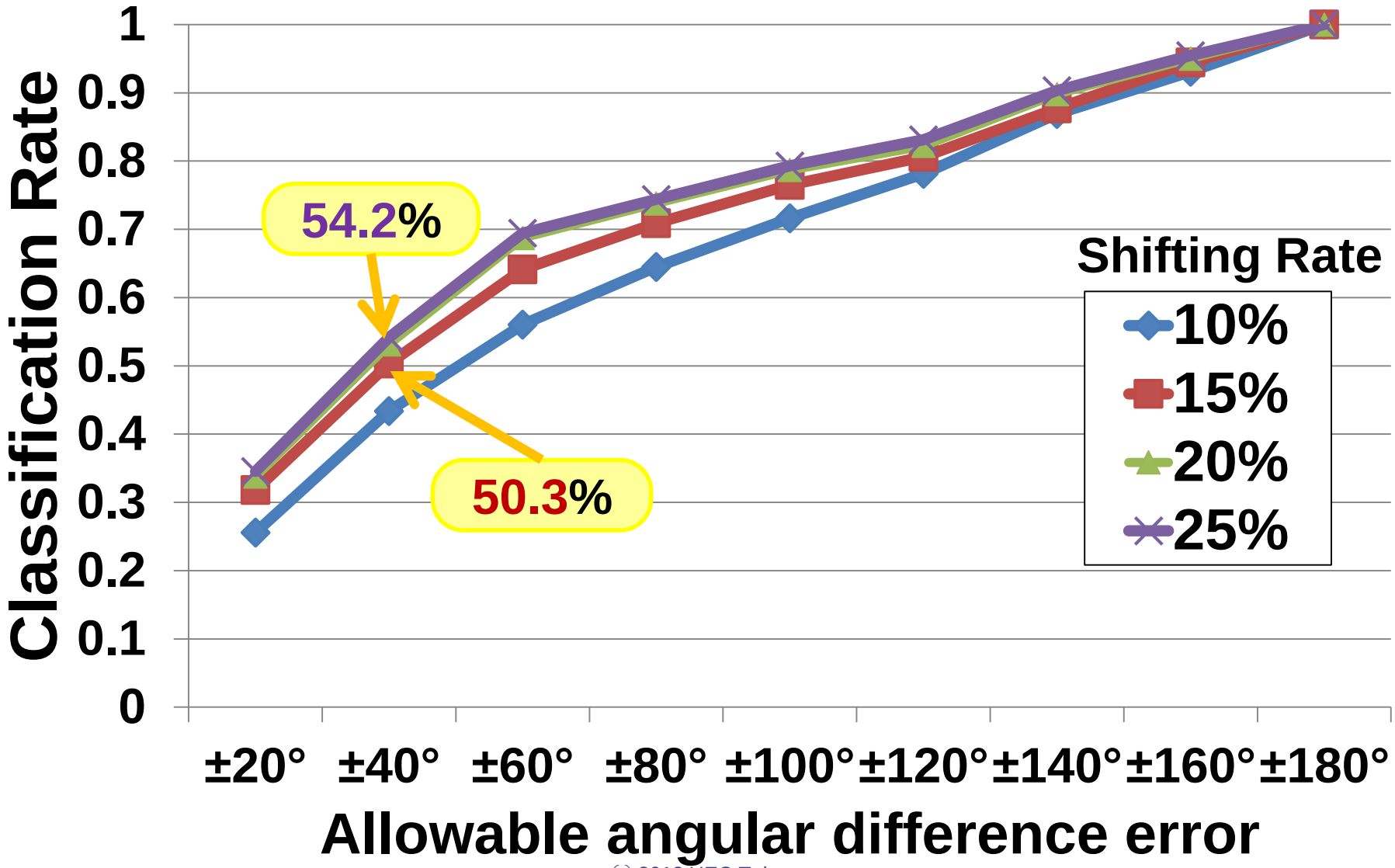
- Dataset (test)
  - Shifting the ground-truth bounding boxes – x%
  - To each of eight directions

- Evaluation
  - Error in the direction estimation
  - Cumulative classification rate

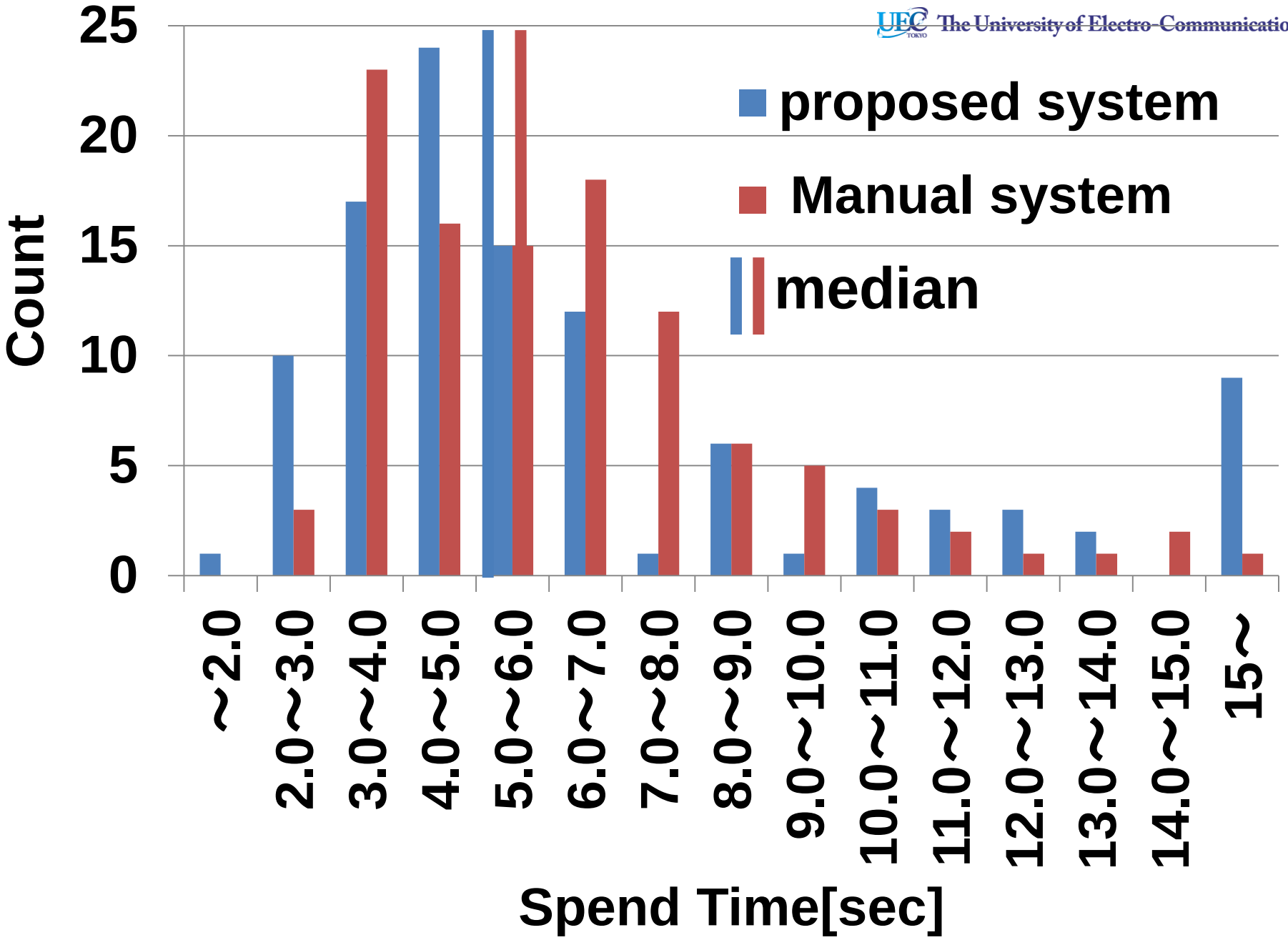
$$\frac{number\ of\ less\ than\ y^0}{number\ of\ all\ the\ test\ images}$$

ⓒ 2013 UEC Tokyo.

Experiment of estimation direction



ⓒ 2013 UEC Tokyo.



ⓒ 2013 UEC Tokyo.

Set up for Experiment

- Dataset
  - Includes 50 kinds of food categories
  - Has more 100 images for each category

- Performance:
  - Classification Rate =

$$\frac{number\ of\ correctly\ detected\ images\ in\ Top\ n}{number\ of\ all\ the\ test\ images}$$

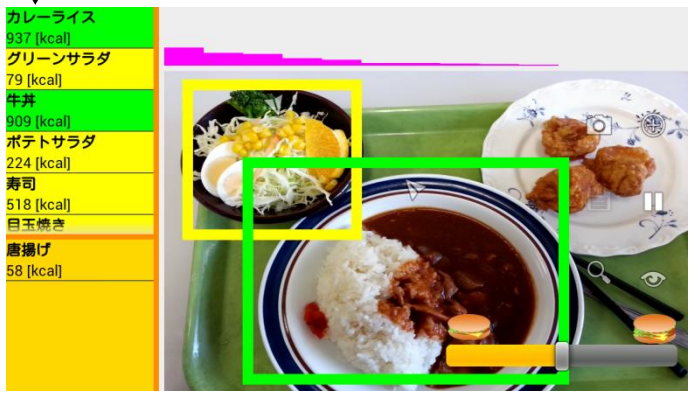
ⓒ 2013 UEC Tokyo.

Objective

- To record food items in every meal
  - quite troublesome task.

More Quickly and Easier

- Proposed System
  - Recognize food in real-time
  - require no connection with se

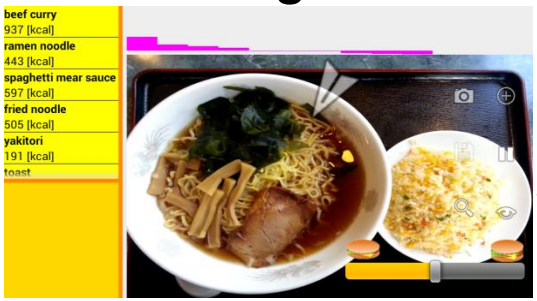


ⓒ 2013 UEC Tokyo.

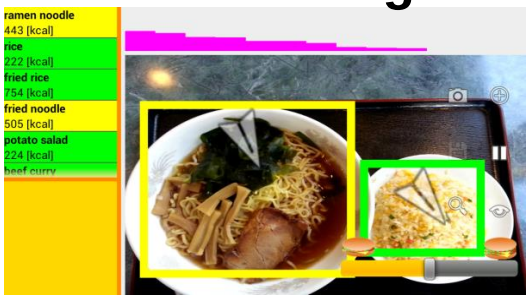
Flow of recognition and record

Point a smartphone toward food items

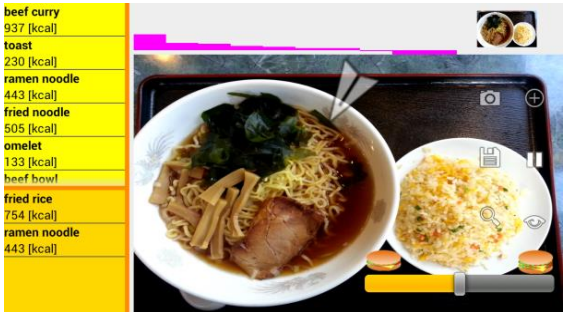
Start recognition



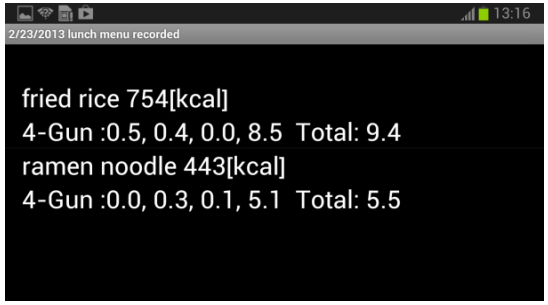
Draw bounding box



Select food items



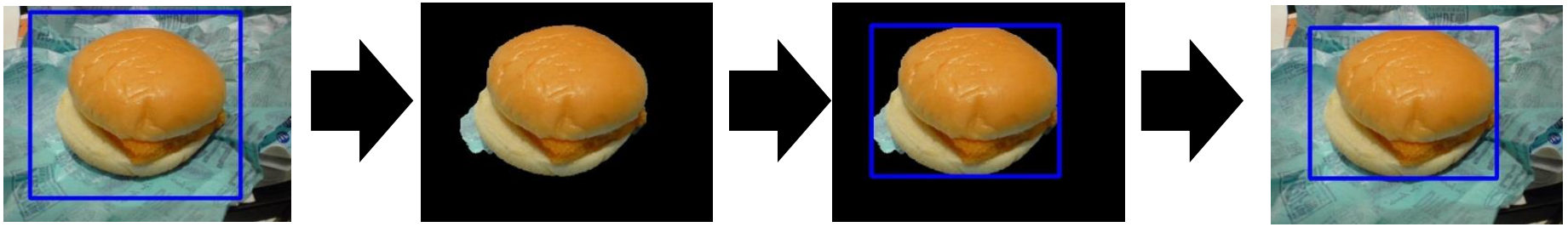
Register food record



ⓒ 2013 UEC Tokyo.

Bounding Box Adjustment

- Why use Bounding Box?
  - recognizes repeatedly in real-time
  - a smartphone is not fixed.



GrabCut

ⓒ 2013 UEC Tokyo.

Set up for Experiment

- Evaluation of processing time
  - device
    - Galaxy Note2(1.6GHz,4core,4thread, Android 4.1)
  - multi thread for Quad core

- Evaluation
  - processing time
  - average

ⓒ 2013 UEC Tokyo.

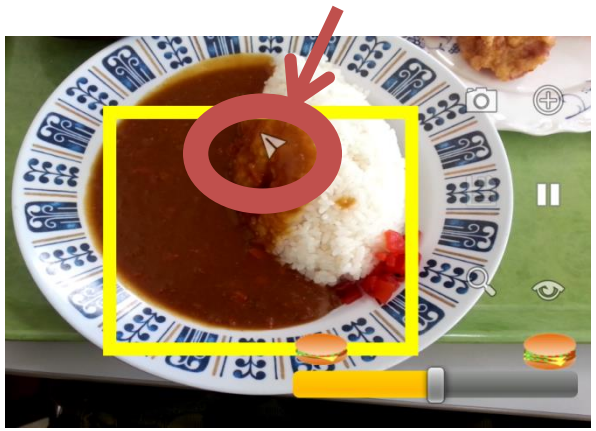
Estimation of the more reliable direction

- Effective window search
- No transposed SURF-BoF + linear SVM
  - Linear SVM

    $$\mathbf{w} = \mathbf{w}^{+} + \mathbf{w}^{-}$$

  - SVM score
    - Create integral images about $\mathbf{w}^{+}$ and $\mathbf{w}^{-}$
    - $O(1)$ operation (ref ESS)

result



- Show the result
  - The window with the maximum score

ⓒ 2013 UEC Tokyo.

Future Work

- Estimate direction
  - Consider other information such as shape
  - Move the bounding box automatically
- Draw bounding box
  - Make less complex such as touch just a point
- Use additional information
  - User's meal history
  - GPS and time information

ⓒ 2013 UEC Tokyo.

You can try it

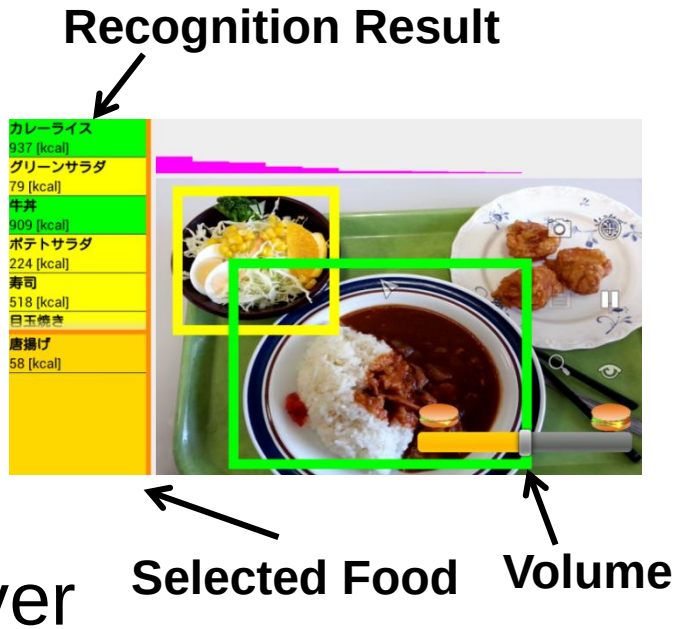
- The Android application can be downloaded.
  - We do not provide iOS version.
- Download site
  - http://foodcam.mobi/
  - Android 3.1 and up
  - Quad cores and up

ⓒ 2013 UEC Tokyo.

Objective

- To record food items in every meal
  - quite troublesome task.

More Quickly and Easier



- Proposed system
  - recognize food in real-time
  - require no connection with server
  - return results with 81.6% in the top5.

ⓒ 2013 UEC Tokyo.

ⓒ 2013 UEC Tokyo.