Food Image Analysis: The Big Data Problem You Can Eat!

Edward J. Delp and Fengqing Zhu
Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana

Carol J. Boushey
Cancer Epidemiology Program
University of Hawaii Cancer Center
Honolulu, Hawaii

www.viperlab.org

www.tadaproject.org

Research Team

- **Carol Boushey**
- **Fengqing Zhu**
- **Students**
  - **Yu Wang, Shaobo Fang, Chang Liu, Ziad Ahmad, TusaRebecca Schap, Ye He, Chang Xu, Marc Bosch Ruiz**
- **Acknowledge the National Institutes of Health, Curtin University, CTSI, Purdue Research Foundation**

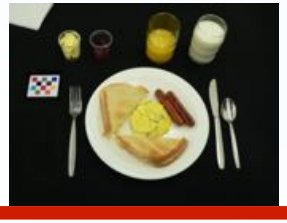
Measuring accurate dietary intake is considered to be an open research problem

TADA System Overview

Image(s) + Metadata

(Geolocation, Time, Barcode, Contextual Info)

Client



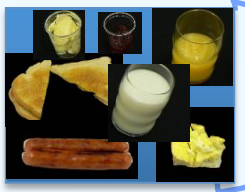
1

Wi-Fi/3G/4G Network

Server

2

- Food Label Type
- Segmented Image
- Machine Learning
- Context Processing
- User Eating Patterns



3

Volume Estimation

4

Output

- Labeled Images With Food Type (e.g. Milk, Toast, Eggs)

+

6

Communication Layer

Web Server

TADA Food Databases

I-TADA

T-FNDDS

E-TADA

5

User Feedback (Confirmation or Correction)

Wi-Fi/3G/4G Network

7

Researchers, Practitioners

Internet

www.tadaproject.org

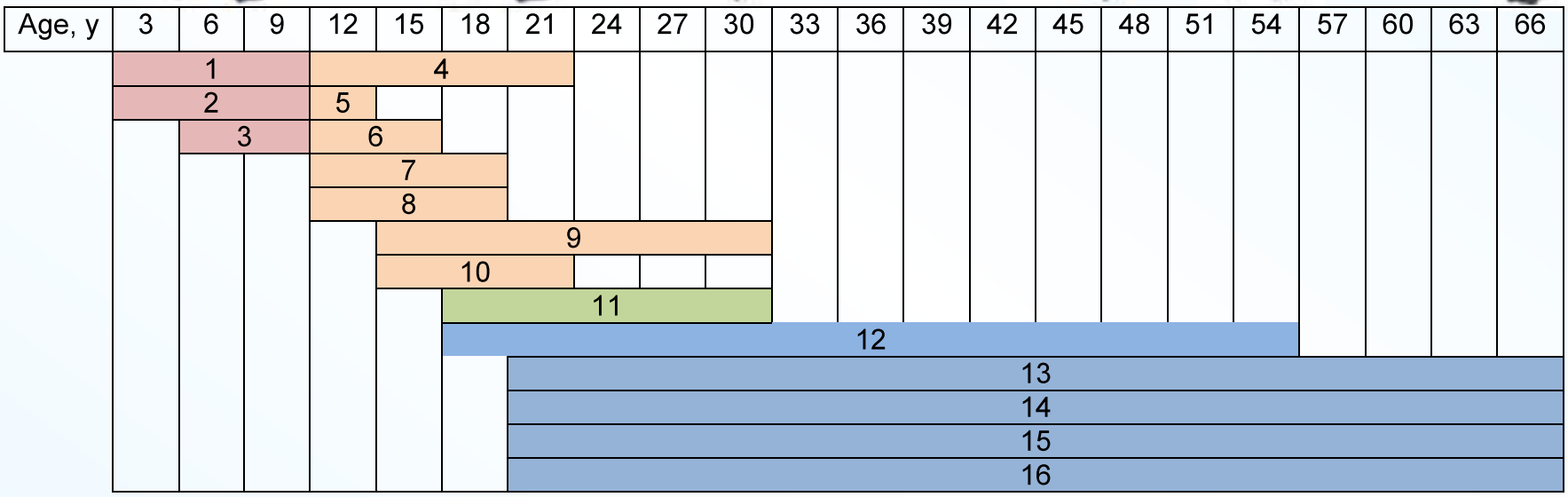
User Studies

- **We have completed a total of 15 user studies**
  - **Free-living environment**
  - **More than 900 participants**
  - **More than 70,000 images acquired**

- **Each food image captures a real eating scene consists of multiple food items**

- **We have published more than 30 conference papers and 18 journal papers published since 2008 (detailed information at: www.tadaproject.org)**

Summary of Studies: Completed or In Process



Study number	Age range, y	Sample size	Supervised, EO	Community, # of days
1	3-10	63	1-4	--
2	3-10	63	--	4
3	7-10	12		2
4	10-21	19	--	4
5	11-13	69	--	3
6	11-15	41	9-11	--
7	11-18	63	2	--
8	11-18	15	24 hr	--
9	11-30	94	--	4

Study number	Age range, y	Sample size	Supervised, EO	Community, # of days
10	14-21	18	--	56-58
11	18-30	247	--	8
12	18-55	20	--	8
13	20-70	77	--	8
14	21-65	57	2	--
15	21-65	45	--	7
16	21-65	22	--	22
ALL	3-70	890	1-11	2-58

KEY: y=years, EO=eating occasions

	Community Dwelling							
Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8
Thu	Fri	Sat	Sun	Mon	Tues	Wed	Thu	Fri
n = 46								n = 45
User training & mobile telephone distribution	Men = 15 (33%) Women = 30 (67%) Mean age = 32 y (range: 21-63 y)							Return mobile telephone

University of Hawaii: Study Eating Behaviors of Children In Guam

Before eating

After eating

How old is this participant?

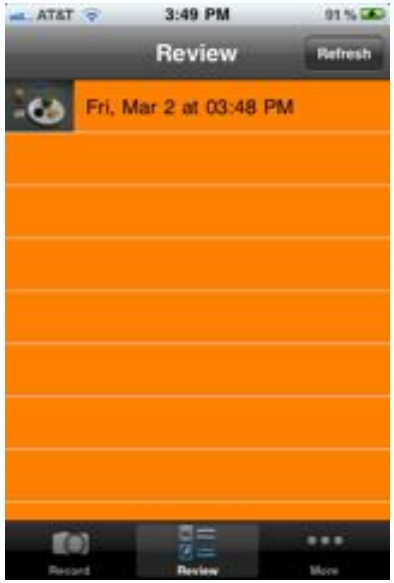
9 years old!

TADA Color Fiducial Marker

- **TADA color fiducial marker is a reference for food classification and portion estimation:**
  - **Geometric reference**
  - **Color reference**
  - **Image quality reference**
- **Real time image quality check is implemented on the mobile phone**

TADA color fiducial marker

TADA App



Apple (iOS8 and iOS9) – iPhone, iPod, iPad

Android (4.3 and above) – phones and tablets

API and SDK are available for various parts of the TADA architecture

Image Acquisition

- **Angle information is obtained from the phone**
- **Colors along with words assist the user in taking an image at the preferred angle**

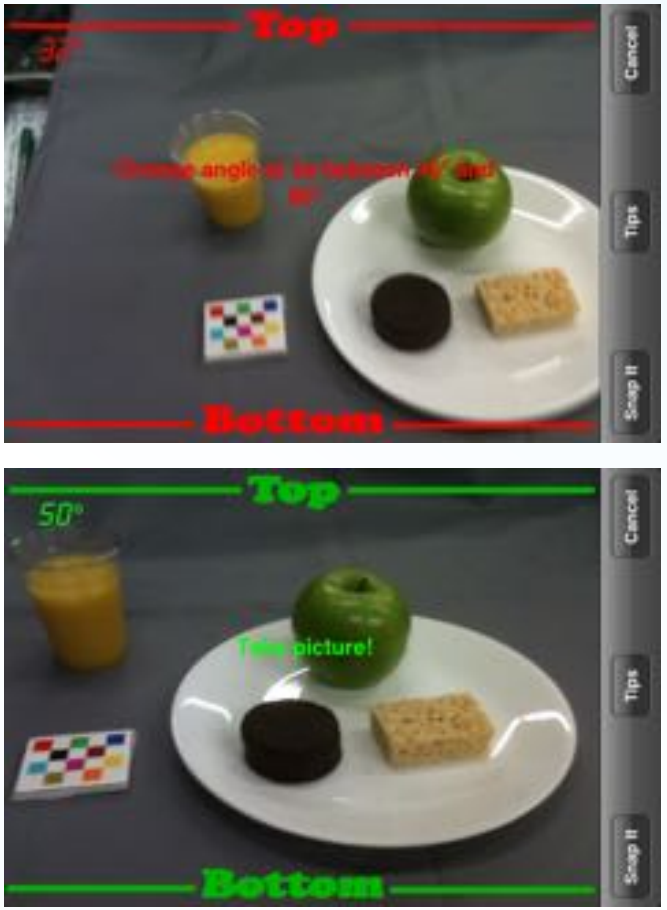
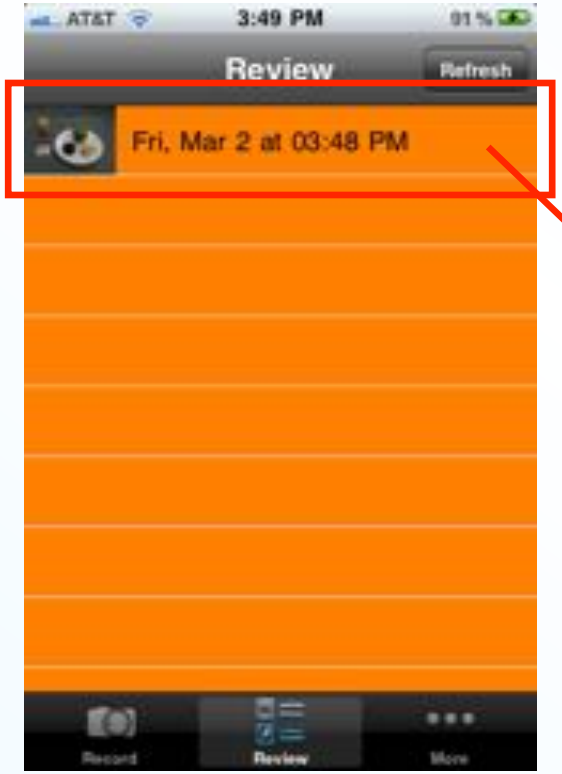


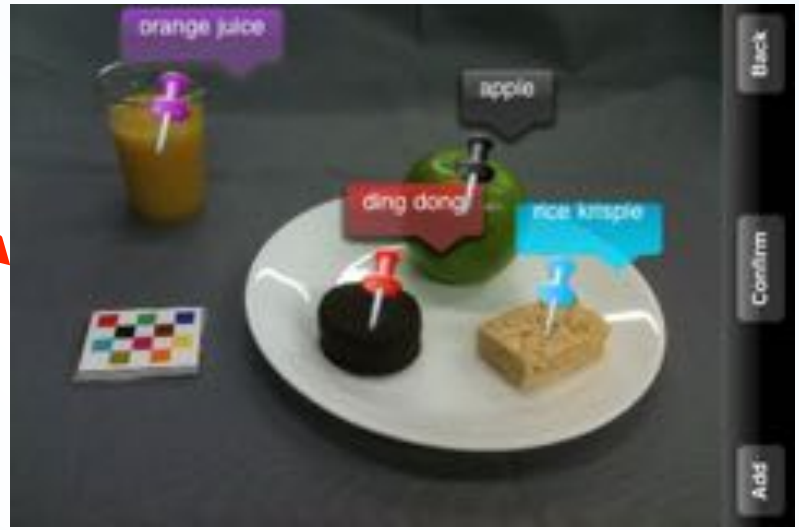
Image Quality Checking

- **Check for:**
  - **Presence of fiducial marker (checkerboard)**
  - **Blurry image**

Checkerboard Missing

Review

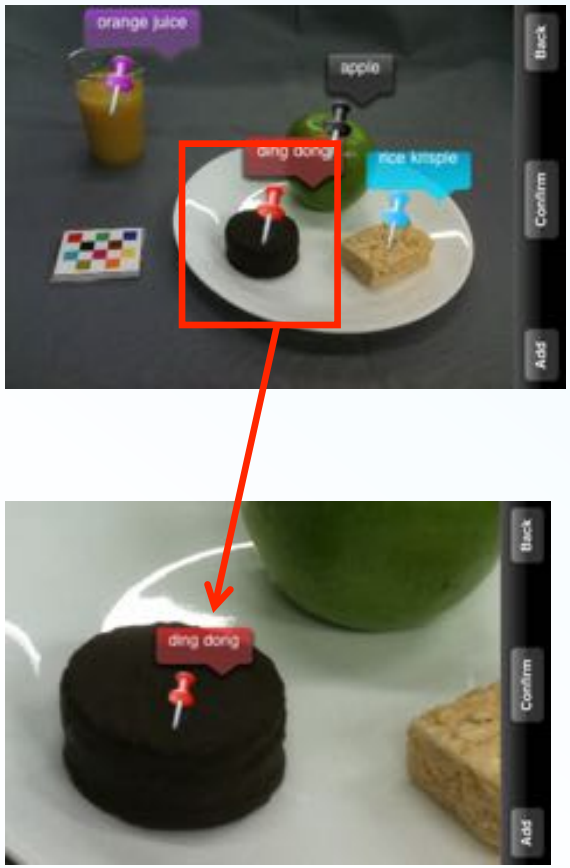




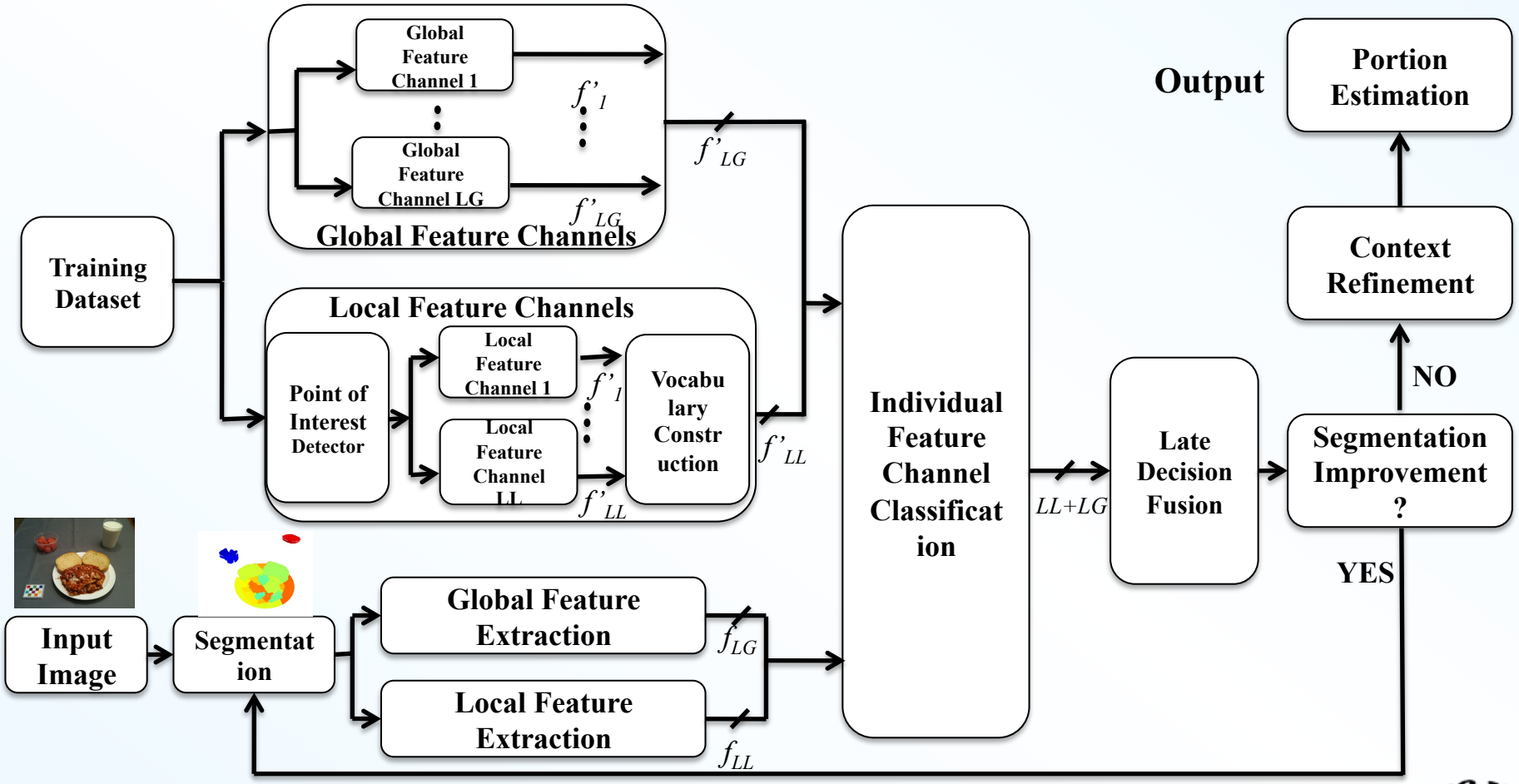
- **The user can select an eating occasion from the list to review**
- **The before eating image is then displayed in landscape view with food labels on it**

Review

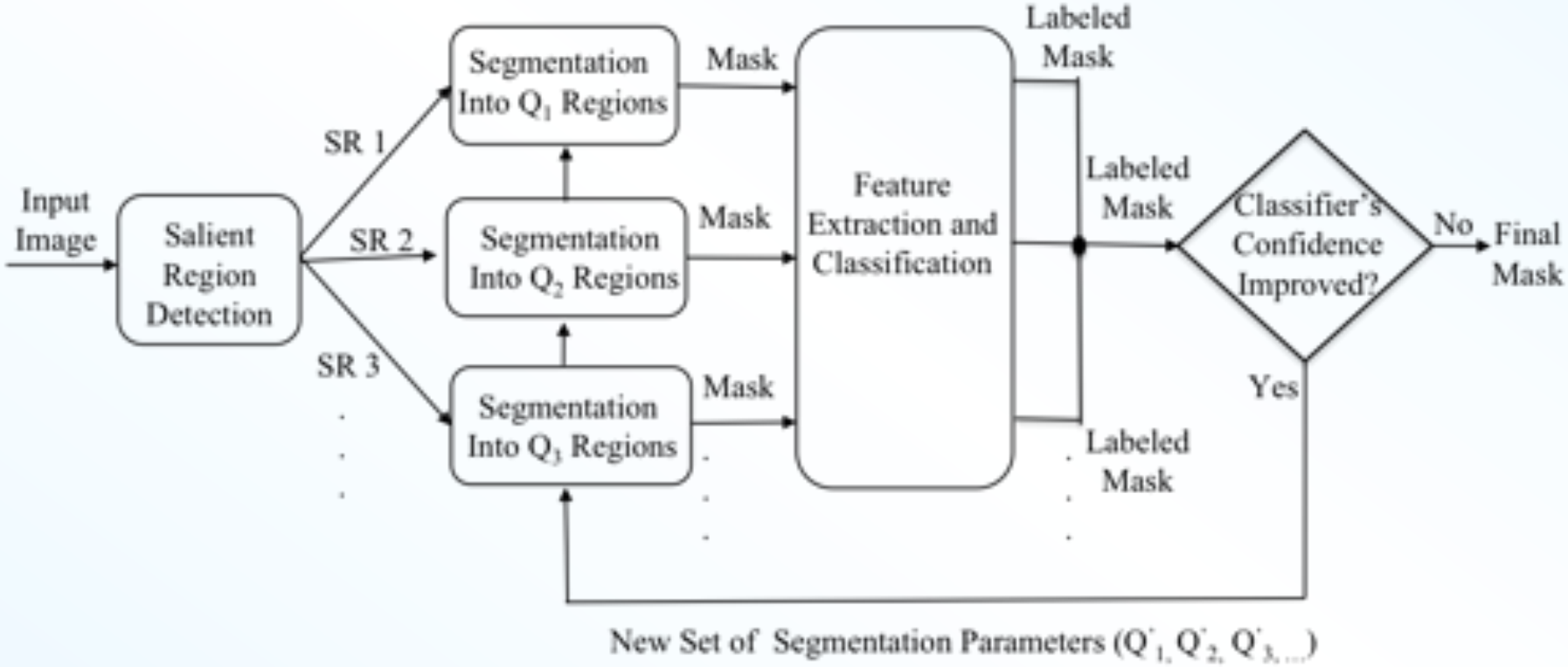
- Different colors help the user in identifying bubble-pin correspondence
- The green color is reserved for confirmed labels
- Zoom-in by pinching the screen to have a better view



TADA Image Analysis System

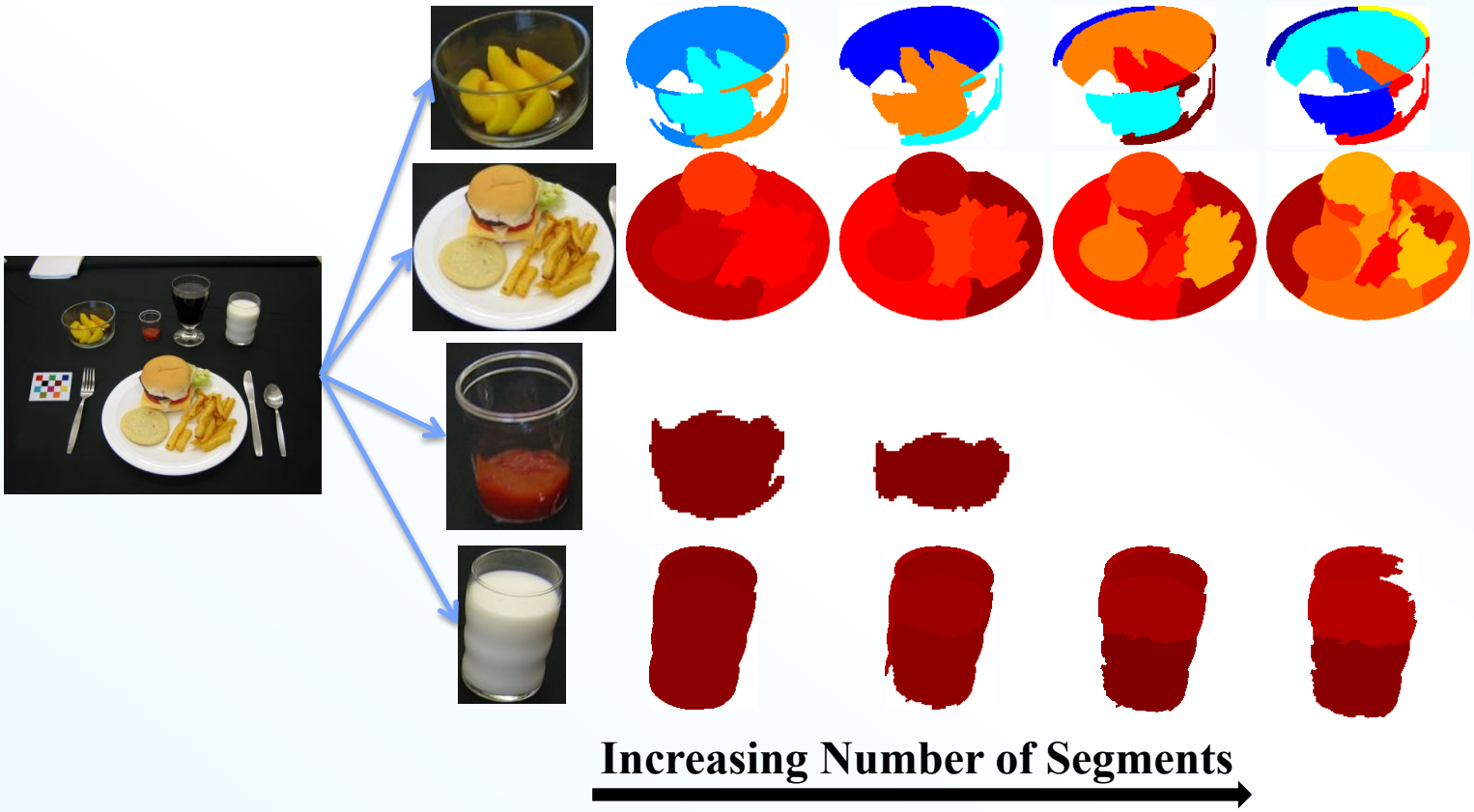


Multiple Hypothesis Segmentation and Classification (MHSC)

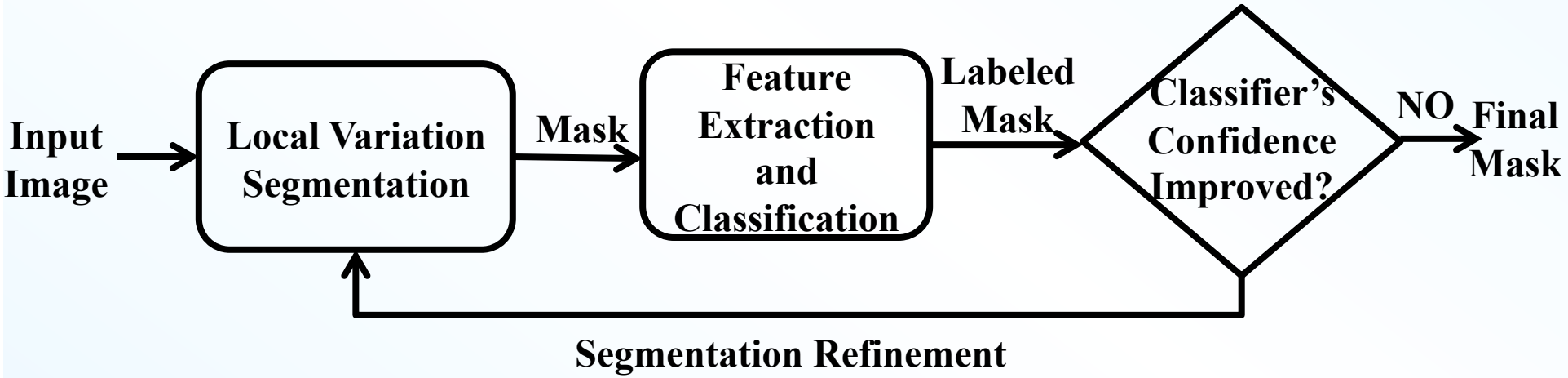


Segmentation

- **Multiscale normalized cut**



Simplified Multiple Hypothesis Segmentation and Classification



Segmentation

- **Local variation**

Original Images **Local Variation**

- **Internal difference of a segmented region**

$$Int(A) = \max_{e \in MST(A,E)} \omega(e)$$

- **Difference between two segmented regions**

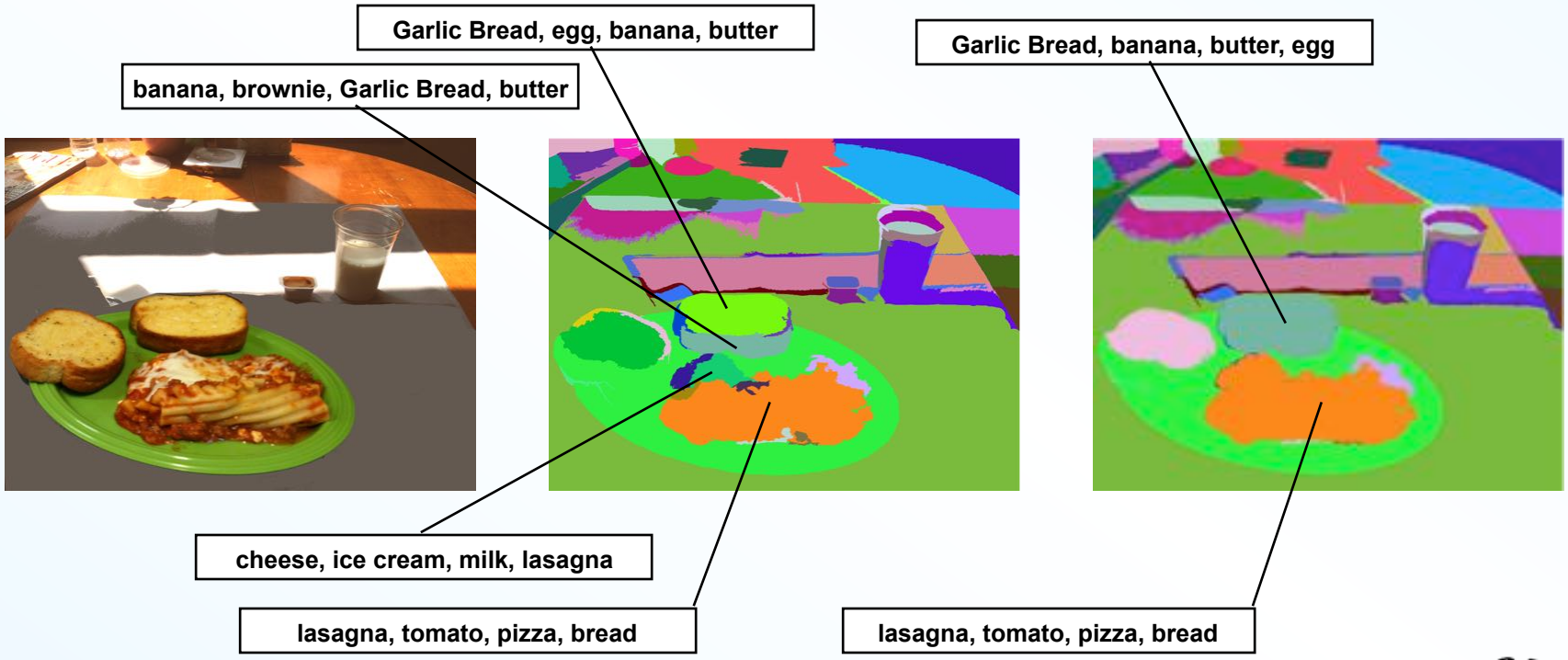
$$Dif(A,B) = \min_{p \in A, q \in B, (p,q) \in E} \omega(p,q)$$

- **Two regions are segmented if**

$$Dif(A,B) > \min(Int(A) + \frac{k}{|A|}, Int(B) + \frac{k}{|B|})$$

Segmentation Refinement

- **Over-Segmentation**

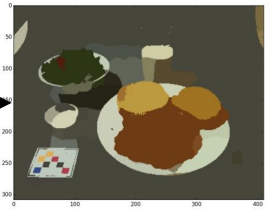


Superpixel Based Segmentation

Simple Linear Iterative Clustering (SLIC)

Build graph model in multiple feature spaces

Normalized Cut (Ncut)



- Use SLIC to get initial segments
- Use graph model to identify adjacent patches
- Create G=(V,E) based on the super pixels
  - Node: superpixel
  - Edge: similarity based on multiple cue (color, proximity and texture)
- Use Ncut on G

Segmentation Comparisons

Original Image

SLIC superpixels

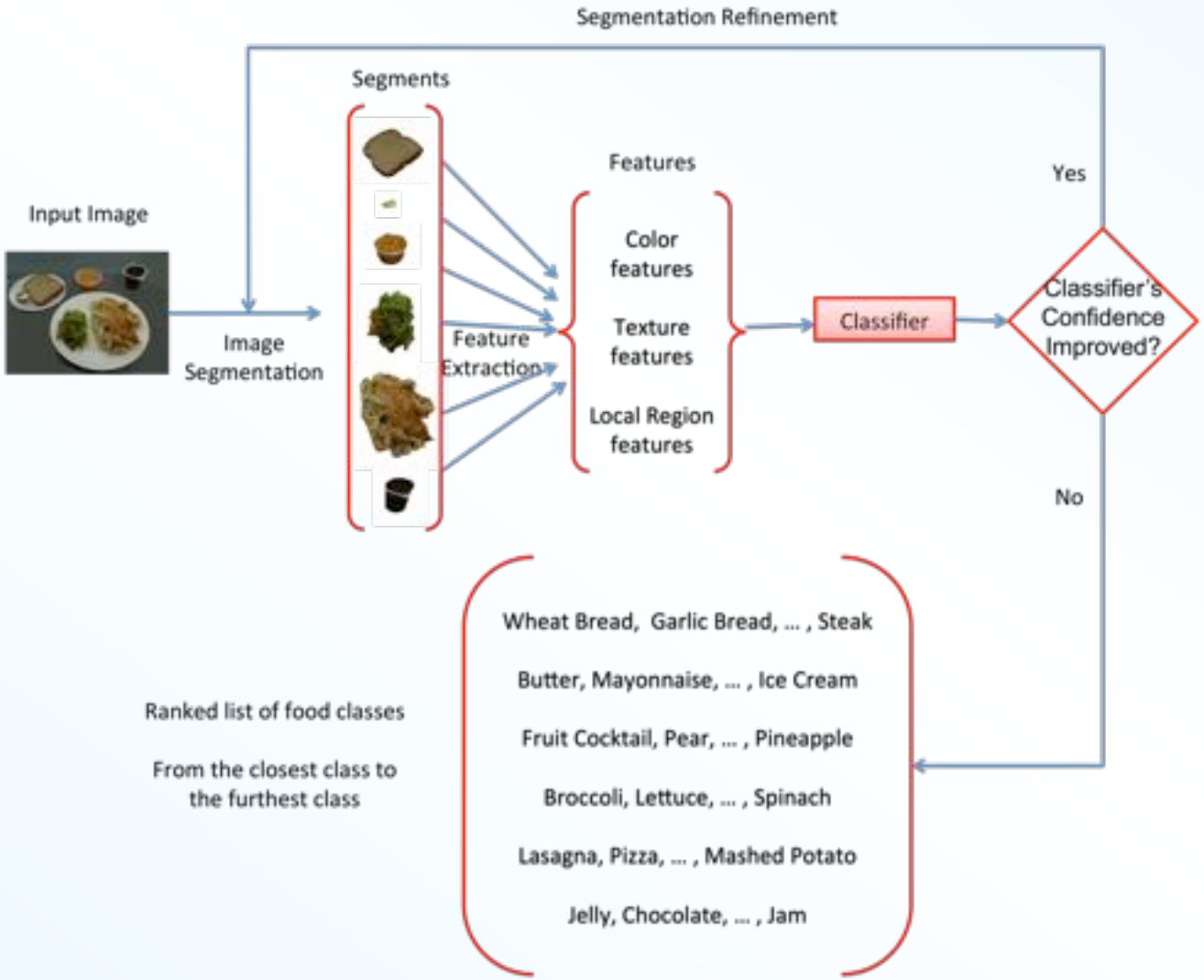
SNcut at ODS

SNcut at OIS

Local Variation

Hierarchical Segmentation

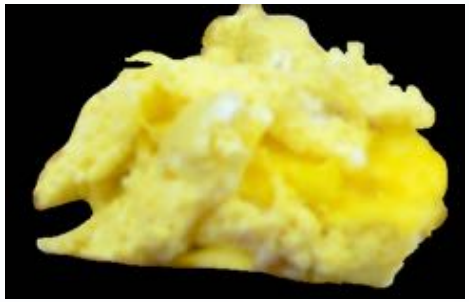
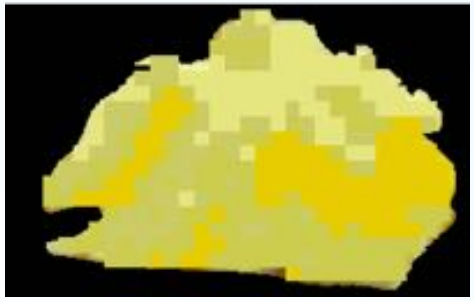
Features


Segmentation Refinement
Segments
Input Image
Image Segmentation
Feature Extraction
Features
Color features
Texture features
Local Region features
Classifier
Classifier's Confidence Improved?
Yes
No
Wheat Bread, Garlic Bread, ... , Steak
Butter, Mayonnaise, ... , Ice Cream
Fruit Cocktail, Pear, ... , Pineapple
Broccoli, Lettuce, ... , Spinach
Lasagna, Pizza, ... , Mashed Potato
Jelly, Chocolate, ... , Jam
Ranked list of food classes
From the closest class to the furthest class

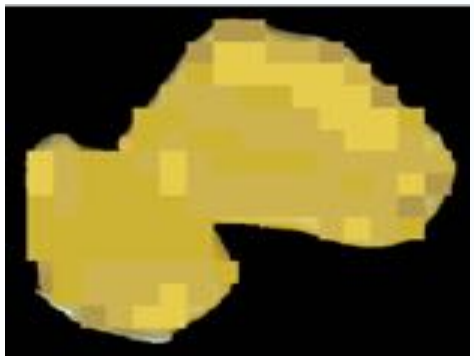
Global Features: Color

- **Color information is the key feature to identify foods like liquids, or to distinct between objects (e.g. mustard and ketchup)**
- **Foods show large variation of color:**
  - **– Homogeneous vs. heterogeneous color distribution**
  - **– Color irregularities (e.g. green vs. ripe fruit)**
- **Investigated 3 color feature channels:**
  - **– Global color statistics**
  - **– Entropy color statistics**
  - **– Predominant color statistics**

Predominant Color Examples

Scrambled Eggs

Canned Pears


Spaghetti

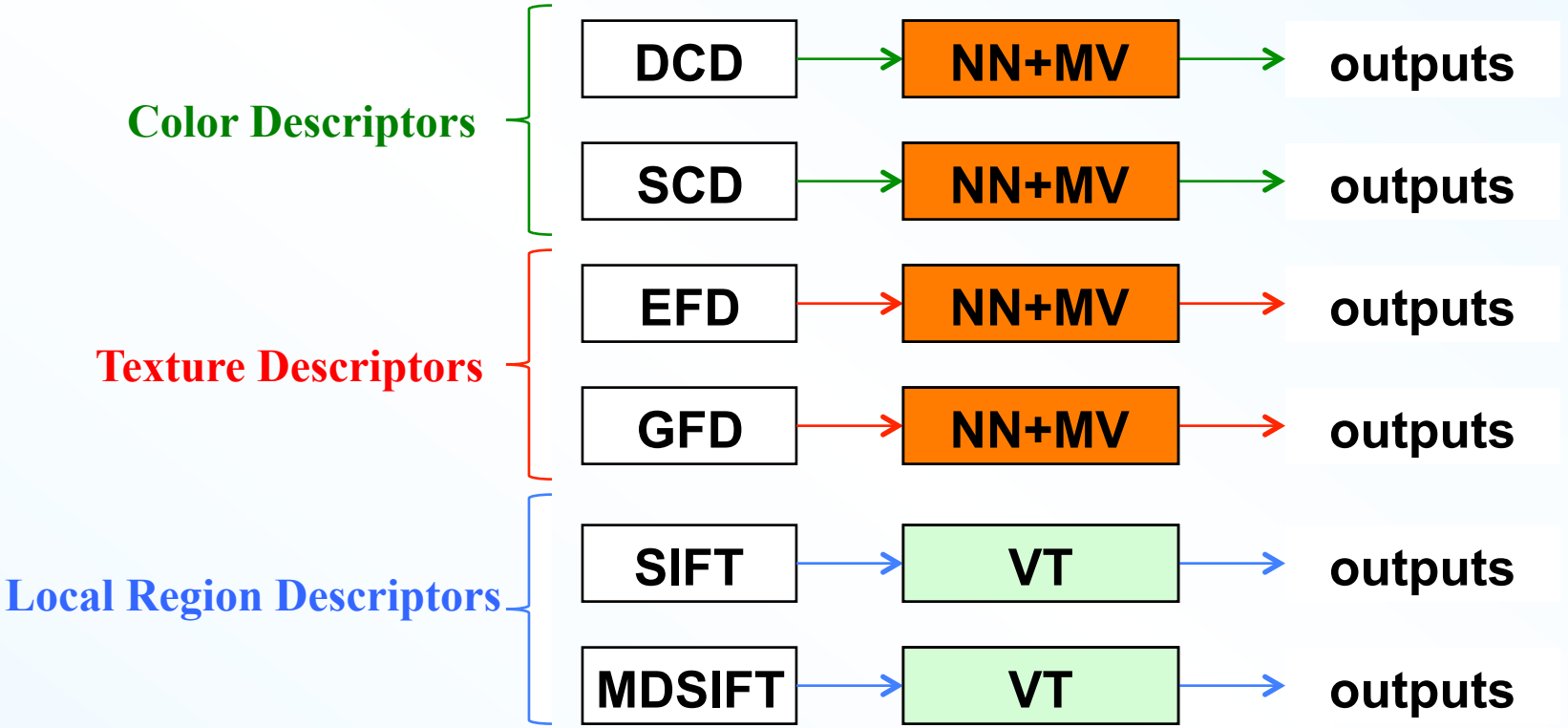
Global Features: Texture

- **Examined texture features for foods**
- **Proposed three texture descriptors:**
  - **Gradient Orientation Spatial-Dependence Matrix (GOSDM)**
  - **Entropy-based categorization and Fractal Dimension (EFD)**
  - **Gabor-based image decomposition and Fractal Dimension (GFD)**
- **Compared with widely used texture features:**
  - **GLCM**
  - **Gabor**
  - **Multifractal Spectrum (MFS)**

Features Used

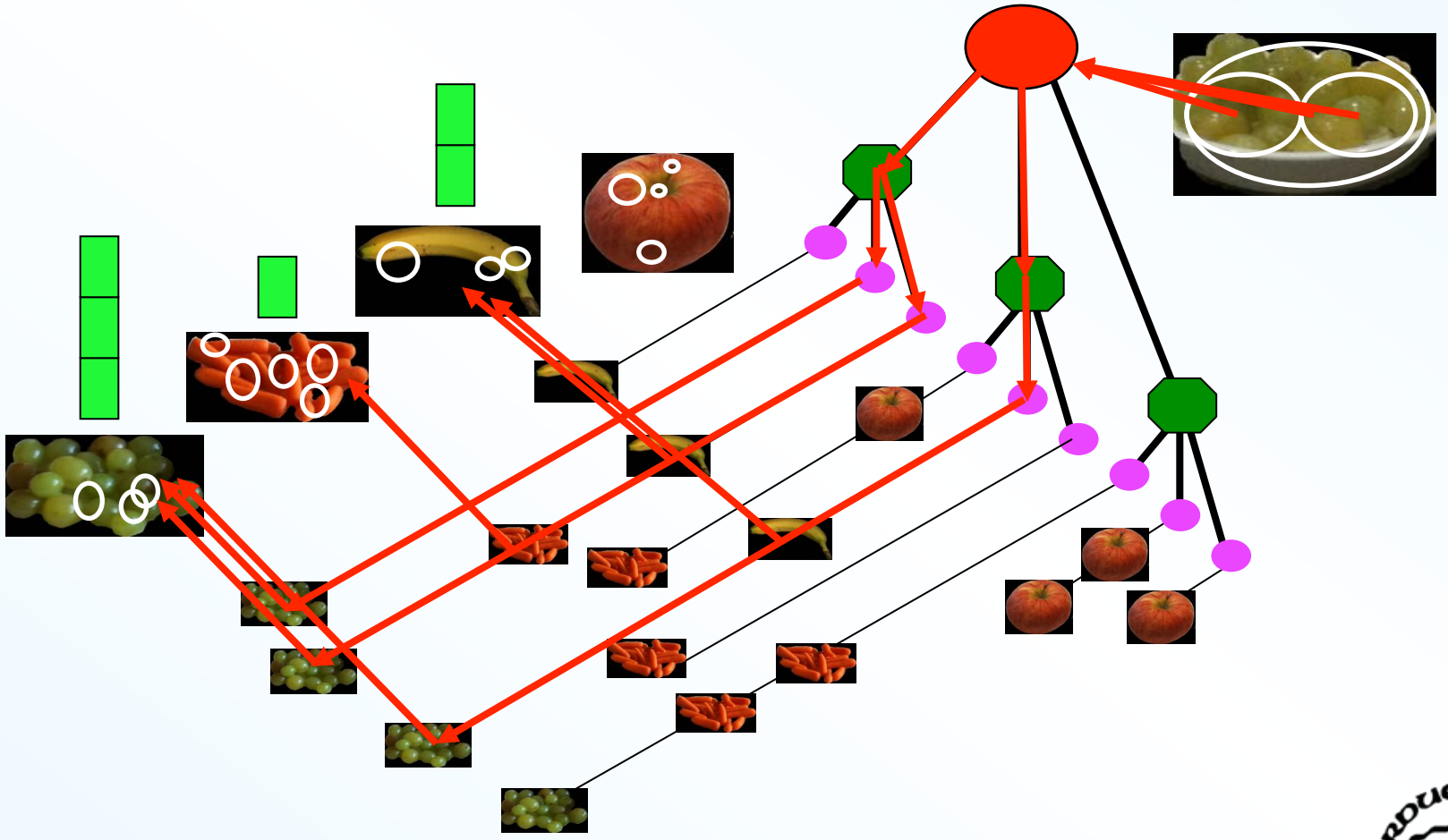
- **Scalable Color Descriptor (SCD)**
- **Dominant Color Descriptor (DCD)**
- **Entropy-Based Categorization and Fractal Dimension Estimation (EFD)**
- **Gabor-Based Image Decomposition and Fractal Dimension Estimation (GFD)**
- **Scale Invariant Feature Transform (SIFT)**
- **Multi-scale Dense SIFT (MDSIFT)**

Feature Classifier



Note: NN: nearest neighbor
MV: majority vote
VT: vocabulary tree

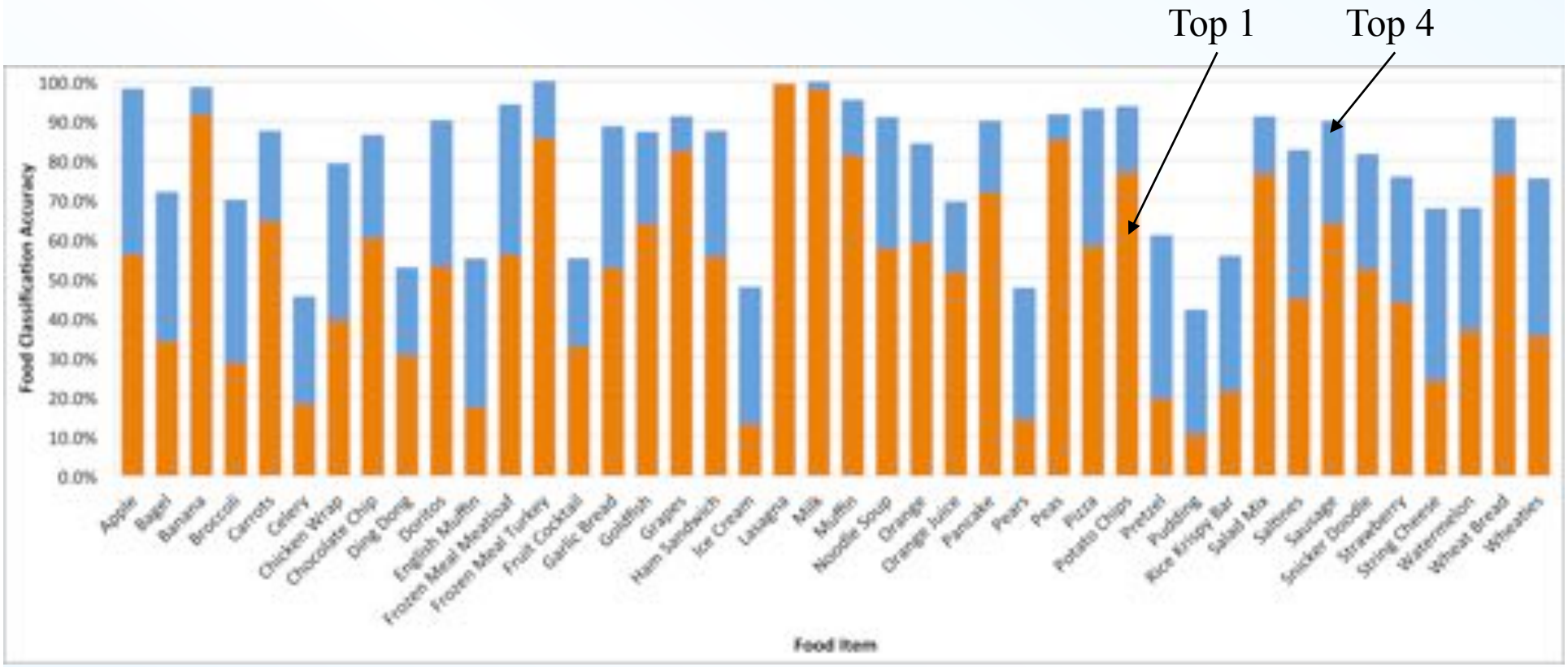
Classification Using A Vocabulary Tree



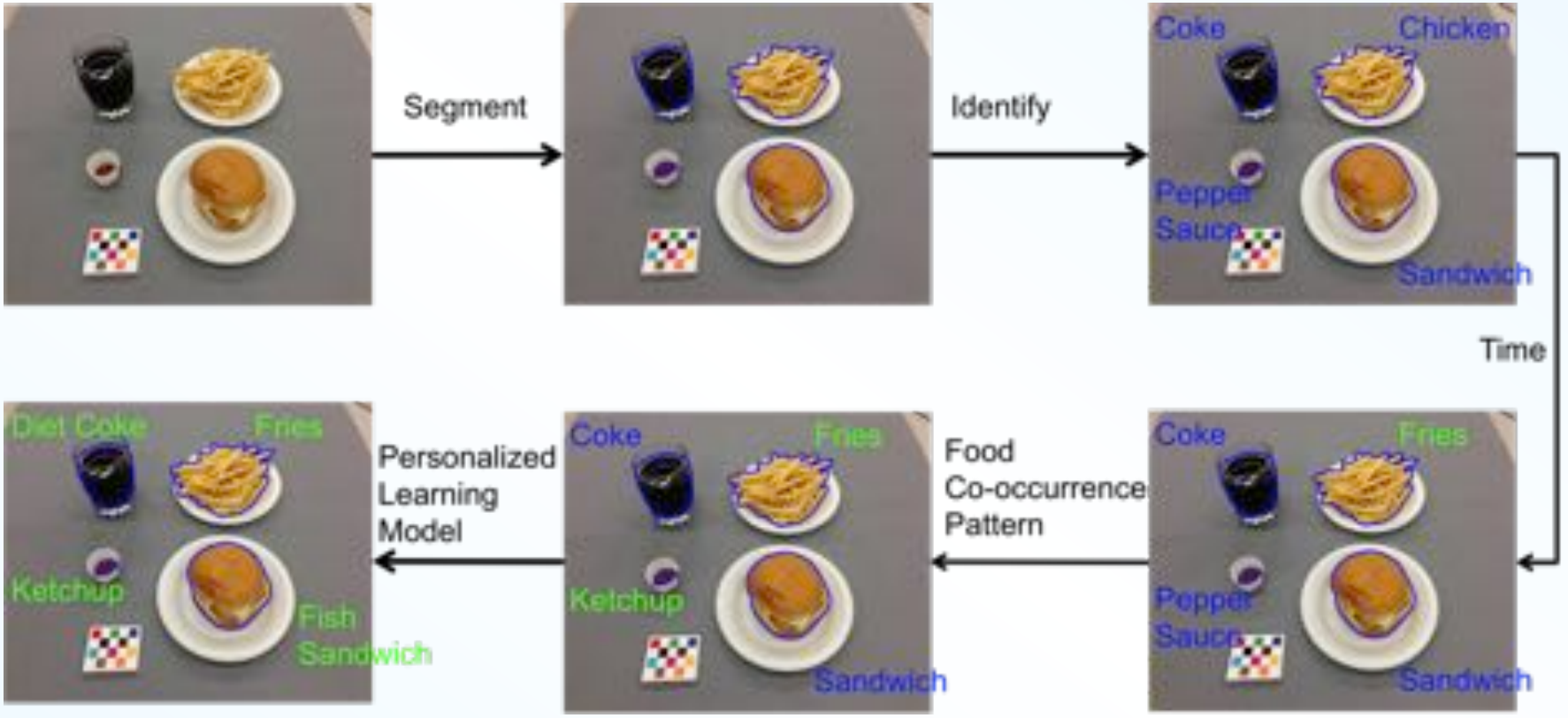
Classification Experiments

- **Test Images: 1453 food images taken by 45 participants in natural eating conditions**
- **Train Images: 20 – 30 images per food class**
- **42 food classes**

Classification Accuracy of Foods



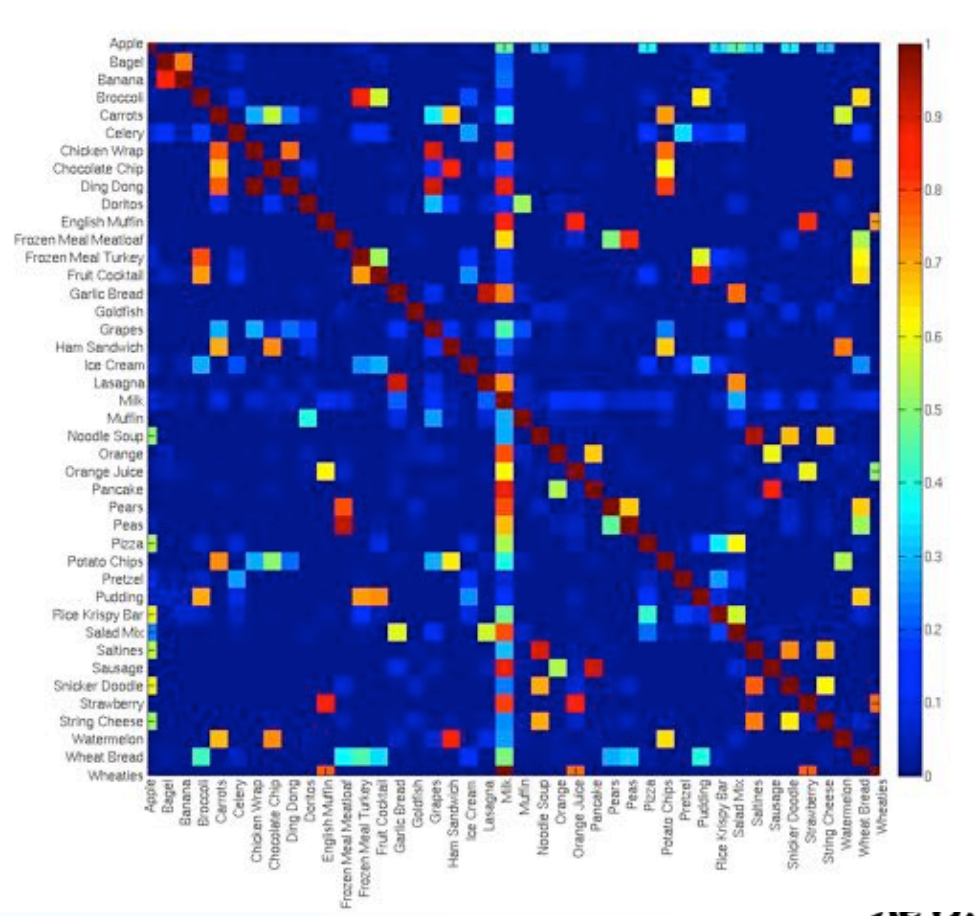
The Use of Contextual Information



- **Time-based food consumption frequency**
- **Food co-occurrence patterns**
- **Date, time, place, dietary habits (patterns), work/ sleep patterns**

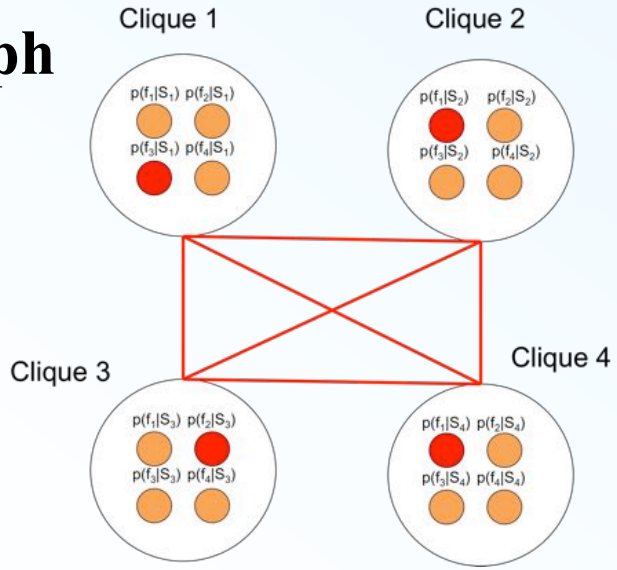
Food Co-Occurrence Patterns

- **The likelihood of food combinations --- their mutual probability of existing together in a single eating occasion**
- **A post-processing stage to promote agreement between the segment labels**



Food Co-Occurrence Patterns

- **Build a fully connected undirected graph between all segments**
- **Adjust the probability of each node by its association with all other nodes**



$$p'(f_k|S_n) = \frac{p(f_k|S_n)A(f,S)}{Z(\phi, S_1, ..., S_N)}$$

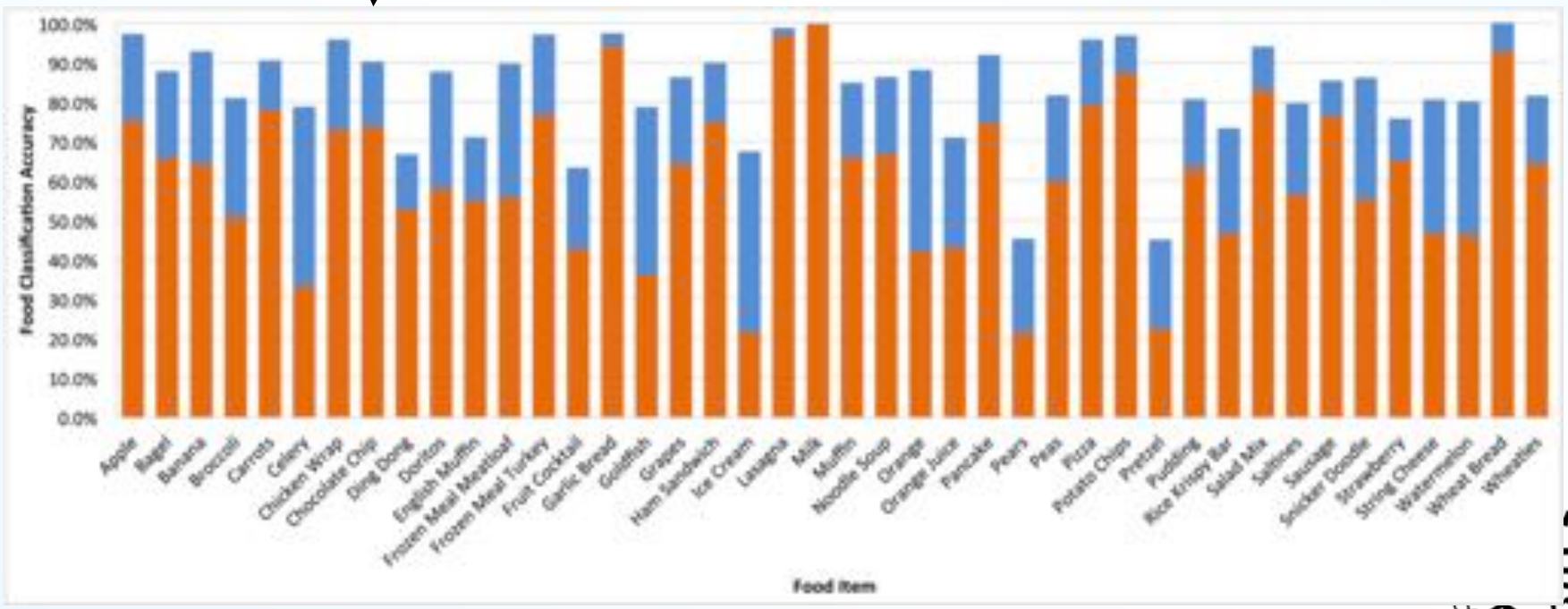
$$A(f,S) = \exp(\sum_{i=0}^{4} \sum_{j=1, j\neq n}^{N} \phi(f_{k,n}, f_{i,j}) p(f_i|S_j))$$

where $p(f_k|S_n)$ is the probability of the food label f_k for segment S_n

$\phi(f_{k,n}, f_{i,j})$ is the co-occurrence probability between two nodes

Z(.) is the normalization constant obtained by summing the numerator over all nodes in the same clique

Food Co-Occurrence



Temporal Context

Recursive Bayesian Model of Food Consumption Frequency

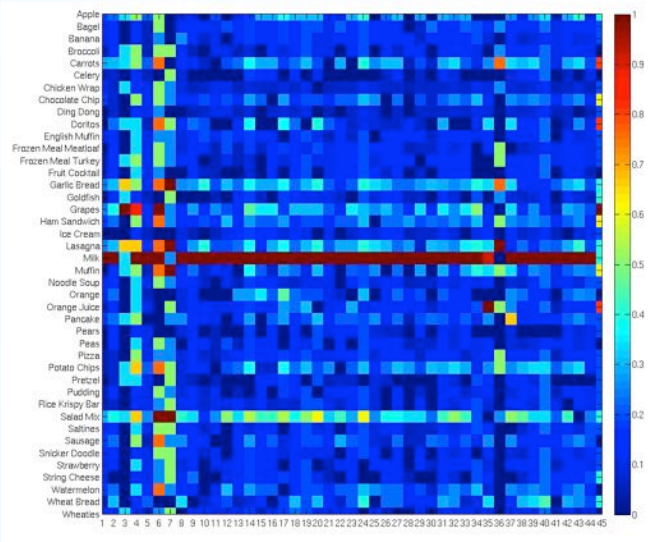
- **Let $p_{\lambda i}(x^k)$ be the probability density function (PDF) representing S_j consumes λ_i on the k^{th} day, and z^k be the observation whether S_j consumes λ_i on the k^{th} day**
- **z^k is obtained from the user feedback in the TADA system**
- **Posteriori update:**

$$p_{\lambda_i}(x^k \mid z^{1:k}) = \frac{p_{\lambda_i}(x^k \mid z^k) p_{\lambda_i}(x^k \mid z^{1:k-1})}{p_{\lambda_i}(z^k \mid z^{1:k-1})} = \frac{likelihood \times prior}{normalization}$$

- **On the $k+1^{th}$ day, $P_{\lambda i}$ is computed as $P_{\lambda i} = \arg\max p_{\lambda i}(x^k \mid z^{1:k})$**

The Use of Temporal Context

- **Selected participants with similar food consumption patterns were used to build personalized eating datasets for a month**
- **Three separate datasets (i.e. dataset 1, 2 and 3) with a total of 358 food images from a free-living study**



statistics	user ID	with context	without context
average daily classification accuracy(%)	user1	61.88	53.23
	user2	65.25	62.90
	user3	59.69	53.28
average daily accuracy improvement(%)	user1	18.45	
	user2	3.85	
	user3	12.39	

Food Portion Estimation: Single-View

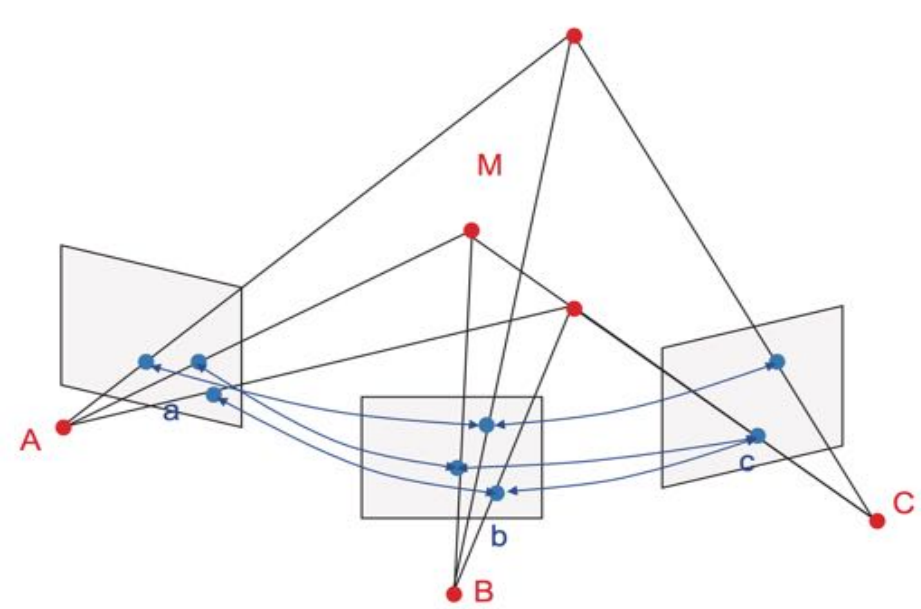
- To reduce a user's burden, our work has focused on the use of a single image to estimate food portion

- Food portion estimation based on a single-view is an ill-posed inverse problem
  - Most 3D information has been lost during projection process

- We use pre-defined geometric models to estimate food portion size

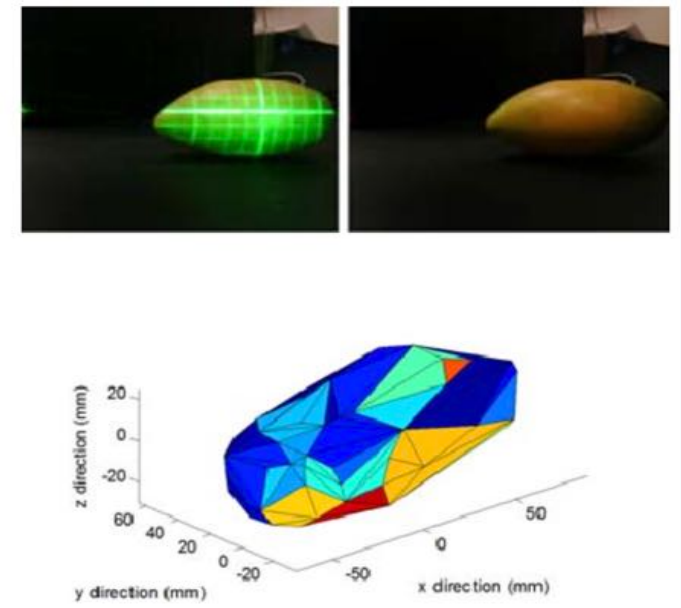
Food Portion Estimation: Stereo and 3D Approaches

multiple images



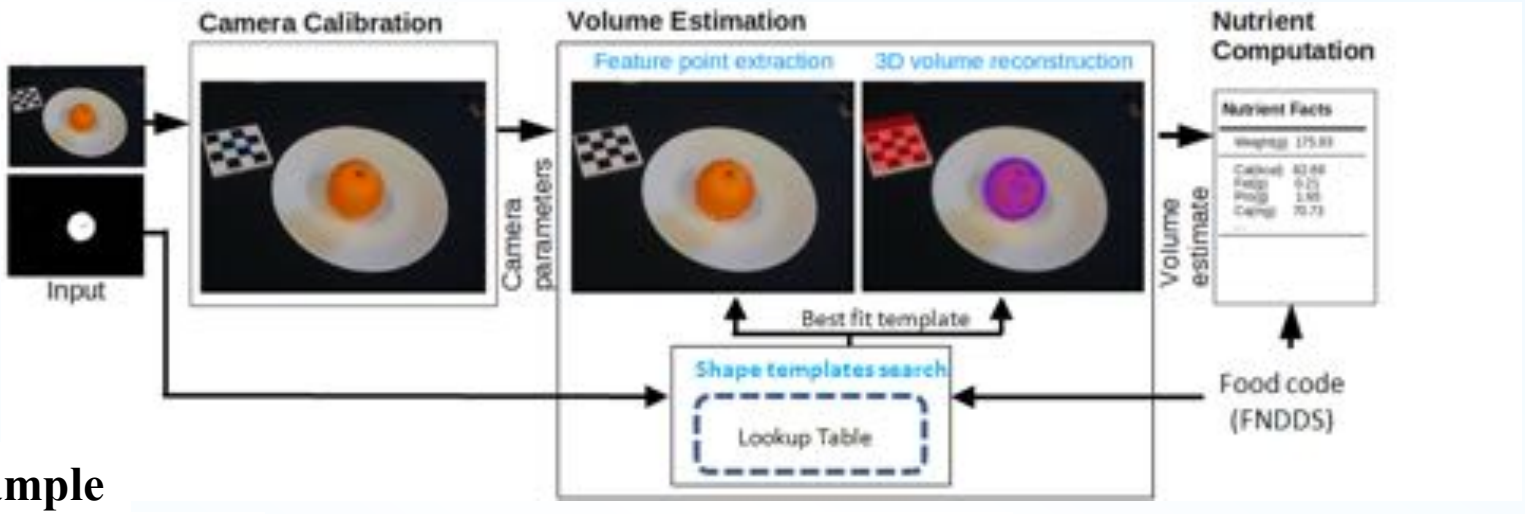
(C. Xu, Ph.D. thesis at Purdue University, May 2014)

3D range images



(J. Shang et al., ICCV Workshop '11)

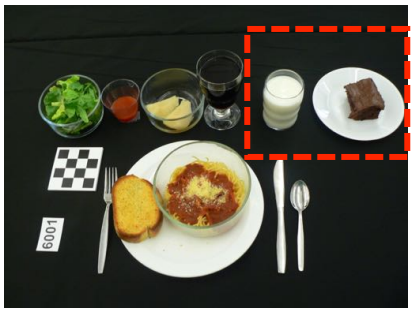
Food Volume Estimation



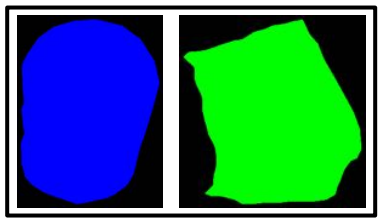
Example

Input datasets

Meal Image



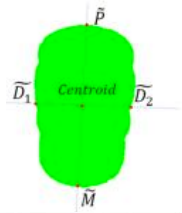
Segmented Images



Food Code
- 11112110 (Milk)
- 53105500 (Chocolate cake)

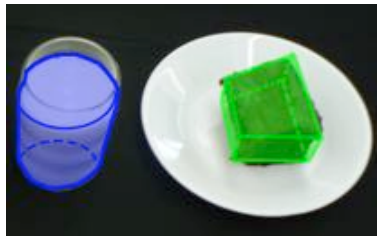
Results

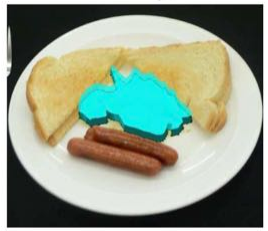
Feature Points



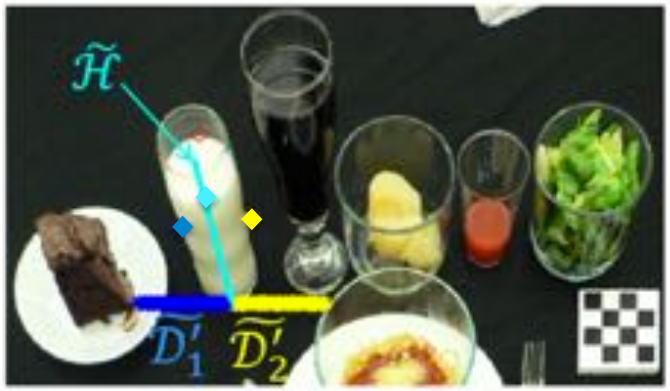
Best-Fit Geometric Model
- Cylinder, Box, Prism

Estimated volumes

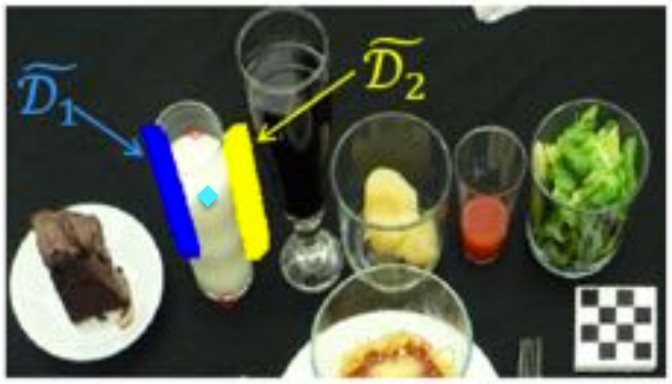




Iterative Points Search for Cylinder Model



(a) Initial search region for radius and height in rectified image coordinates.



(b) Refined search region for radius in rectified image coordinates.

Candidates obtained → Project candidate points → Adjust the increment along radius and height, till projection error has been minimized → Candidates obtained

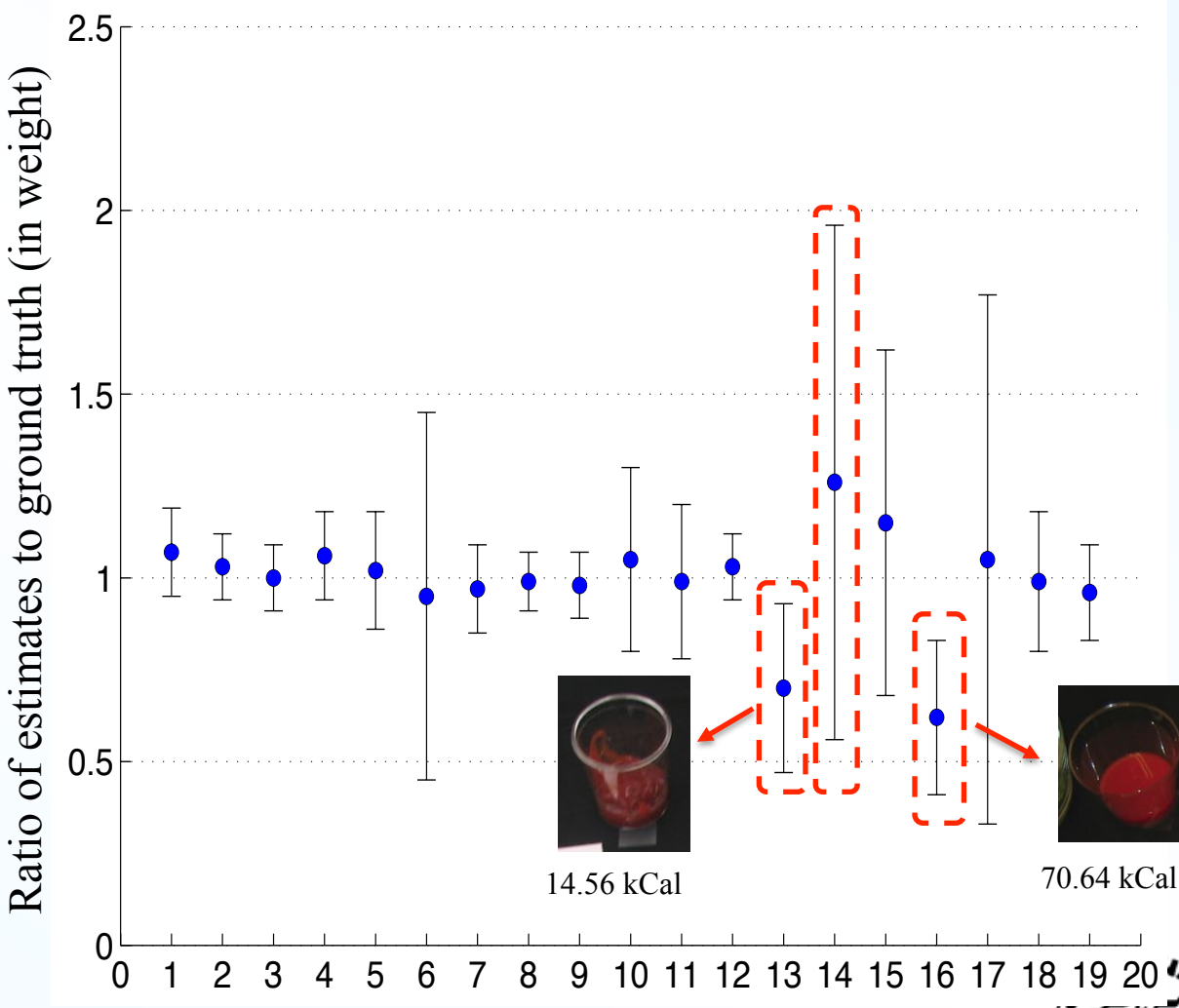
Food Portion Estimation: Points Search

- For example, many food containers are either cylinders or can be approximated to be cylinders
- To estimate portion size only radius and height are required
- A points search technique is designed to have radius and height estimated by minimizing projection error across coordinates

Food Items

1. 2% Milk
2. Sausage links
3. Scrambled eggs
4. Toast
5. Garlic bread
6. Chocolate cake w/ icing
7. Sugar cookie
8. Spaghetti w/ sauce, cheese
9. Orange juice
10. Peach slices
11. Pear, canned halves
12. French fries
13. Ketchup
14. Lettuce (salad)
15. Margarine
16. French dressing
17. Strawberry jam
18. Coke
19. Cheeseburger sandwich

Ratio greater than one, overestimated
Ratio less than one, underestimated



Fang *et al.* ISM 2015

Food And Nutrient Database For Dietary Studies (FNDDS)

- **FNDDS is a database that provides energy, nutrients for typical portions for foods/beverages**

- **The newest version of FNDDS (FNDDS 2011-2012) contains 7,600 main foods/beverages and 9,900 additional foods/beverages associated with a specific main food/ beverage**

- **More information can be found at: www.ars.usda.gov/ba/bhnrc/fsrg**

Food Density

$$\text{Density}\left(\text{g/cm}^3\right) = \frac{\text{Weight (g)}}{\text{Volume (cm}^3\text{)}}$$

- **True density: density of pure substance or material calculated from its component densities**
- **Apparent density: density of a particle including all pores**
- **Bulk density: density when particles are packed or stacked in bulk including void spaces**

The Estimated Energy For A Meal

- **Three sample meals[1]**

Combination type A: average ratio of estimated energy to ground truth: **1.01**

Combination type B: average ratio of estimated energy to ground truth: **0.97**

Combination Type C: average ratio of estimated energy to ground truth: **1.06**

[1]: "Single-View Food Portion Estimation Based on Geometric Models", S. Fang *et al.* ISM 2015

TADA Databases

Users/Patients/Participants

Wi-fi/3G/4G

Internet

Server

Communications Layer

Web Server

TADA Databases

I-TADA

T-FNDDS

E-TADA

Image Analysis and Volume Estimation

Multi-Food Image Analysis With Deep Learning

- **TADA system has been tested and validated by more than 800 users who took 60,000+ food images**
- **VIPER-FoodNet (VFN) dataset: 300,000+ food images from the net verified by crowdsourcing – ongoing**
- **Hierarchical, auto-associative food recognition**
- **Semantic segmentation**
- **Depth prediction from a single image**
- **From supervised learning to unsupervised learning (handle large number of classes)**

Food Image Analysis: The Big Data Problem You Can Eat!

Edward J. Delp and Fengqing Zhu
Video and Image Processing Laboratory
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana

Carol J. Boushey
Cancer Epidemiology Program
University of Hawaii Cancer Center
Honolulu, Hawaii

www.viperlab.org