

# VALSE 2016 武汉



## Food recognition in restaurants

Ruihan Xu, Luis Herranz, Shuqiang Jiang

Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS)

Institute of Computer Technology, CAS, Beijing, 100190, China

IEEE TMM2015 / ICME2015 (oral)

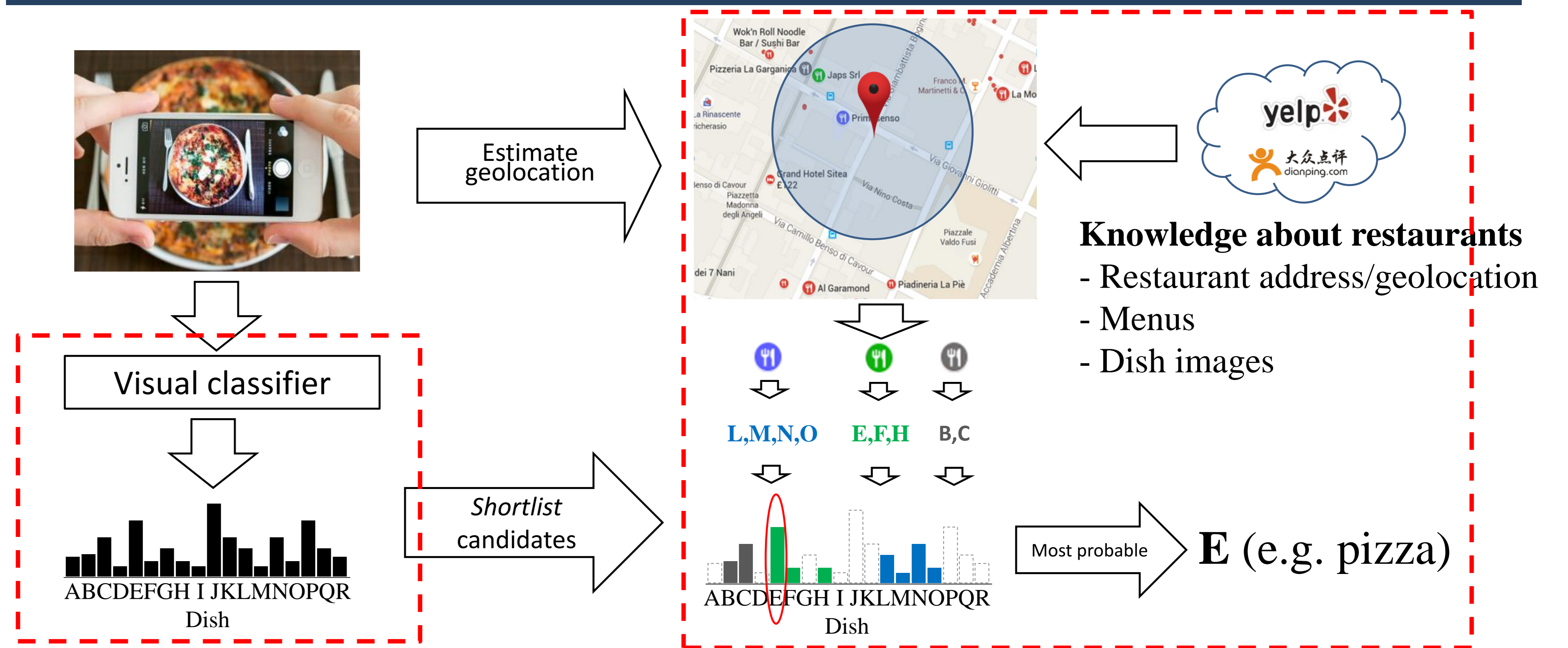
### Introduction

**Scenario:** lunch or dinner in a restaurant or similar. We have a mobile phone with GPS and access to internet.

**Problem:** recognize a dish from a photo and its context = **food recognition + context + external knowledge**

**Applications:** dietary assessment, food intake, eating habits monitoring, annotating features, food/restaurant recommendation, social networks, ...

### Framework

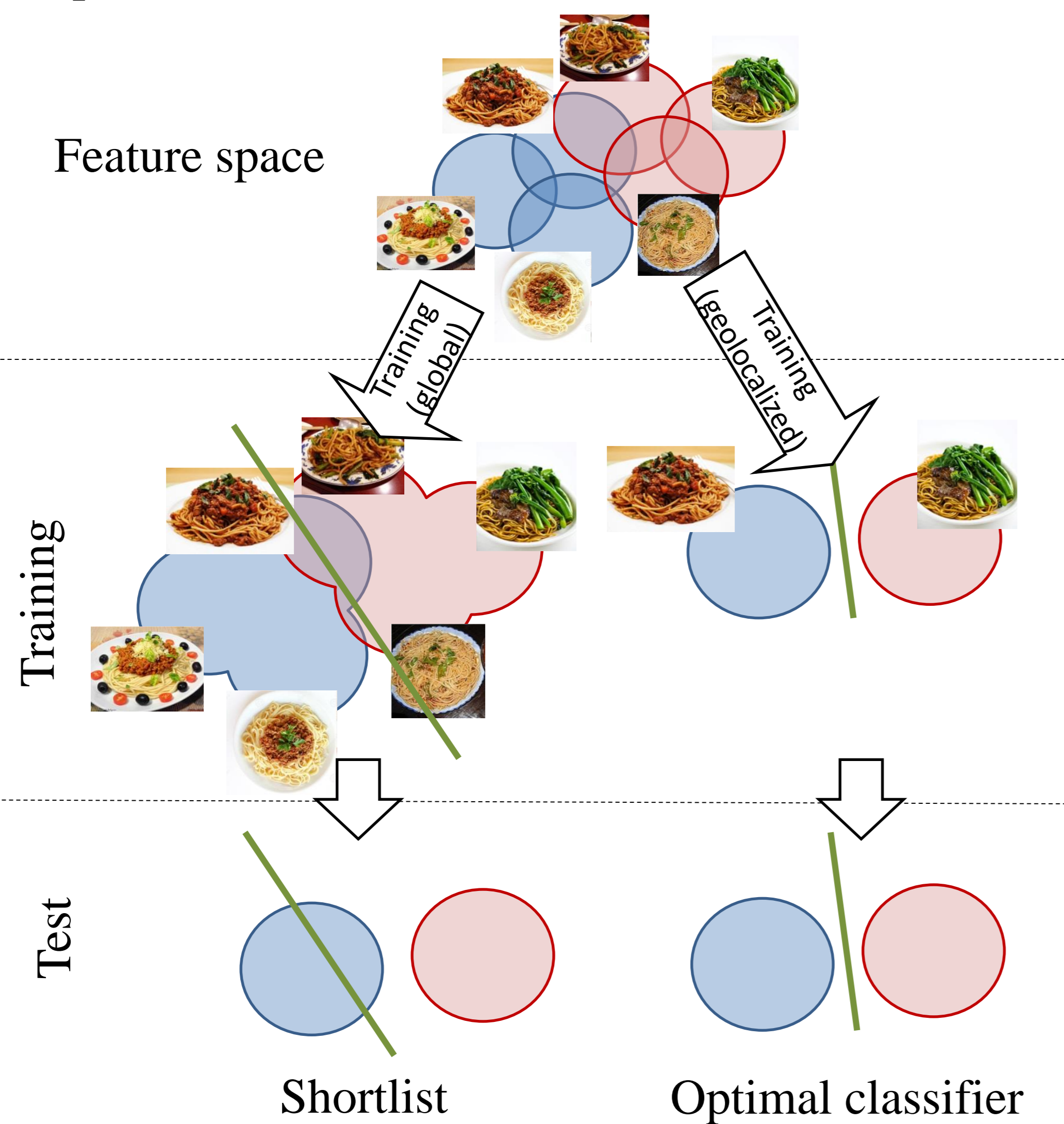


### Geolocalized models[1]

**Problem:** a global classifier ignores the location and solves a very complicated problem (all dishes). Filtering the classes during test helps but is **suboptimal**.

**Optimal classifier:** for *each* query image (and geolocation), select the neighboring restaurants and *retrain* a classifier only for the candidate classes.

**Advantages:** simpler classifier, lower bias, more efficient



However optimal classifier is **not feasible** in practice.

Approach: **geolocalized models**

Requires to geolocalize not only test, but **also training**.

Two strategies[1]:

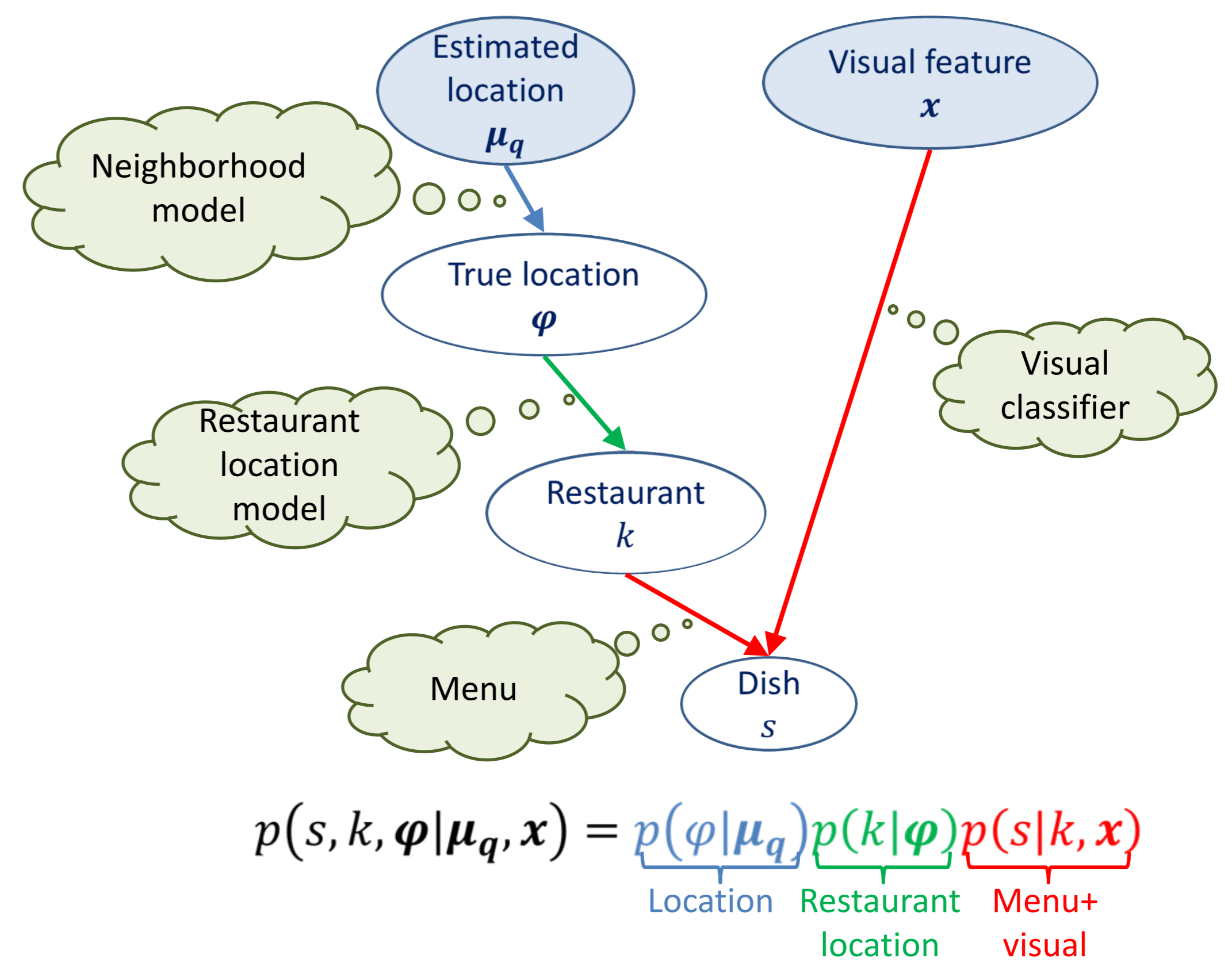
- Pairwise models (train) and geolocalized voting (test)
- Combination of bundled classifiers (CBC)

### Probabilistic framework[2]

Formulate the problem in terms of probabilities and dependencies between different variables.

**Observed** for query  $q$ : (estimated) geolocation  $\mu_q$ , visual feature  $x$

$$\text{Joint distrib.: } p(s, k, \varphi | \mu_q, x) = p(\varphi | \mu_q) p(k | \varphi) p(s | k, x)$$



$$p(s, k, \varphi | \mu_q, x) = \underbrace{p(\varphi | \mu_q)}_{\text{Location}} \underbrace{p(k | \varphi)}_{\text{Restaurant location}} \underbrace{p(s | k, x)}_{\text{Menu+ visual}}$$

By marginalizing we can infer either  $s, k, \varphi$ , leading to different tasks:

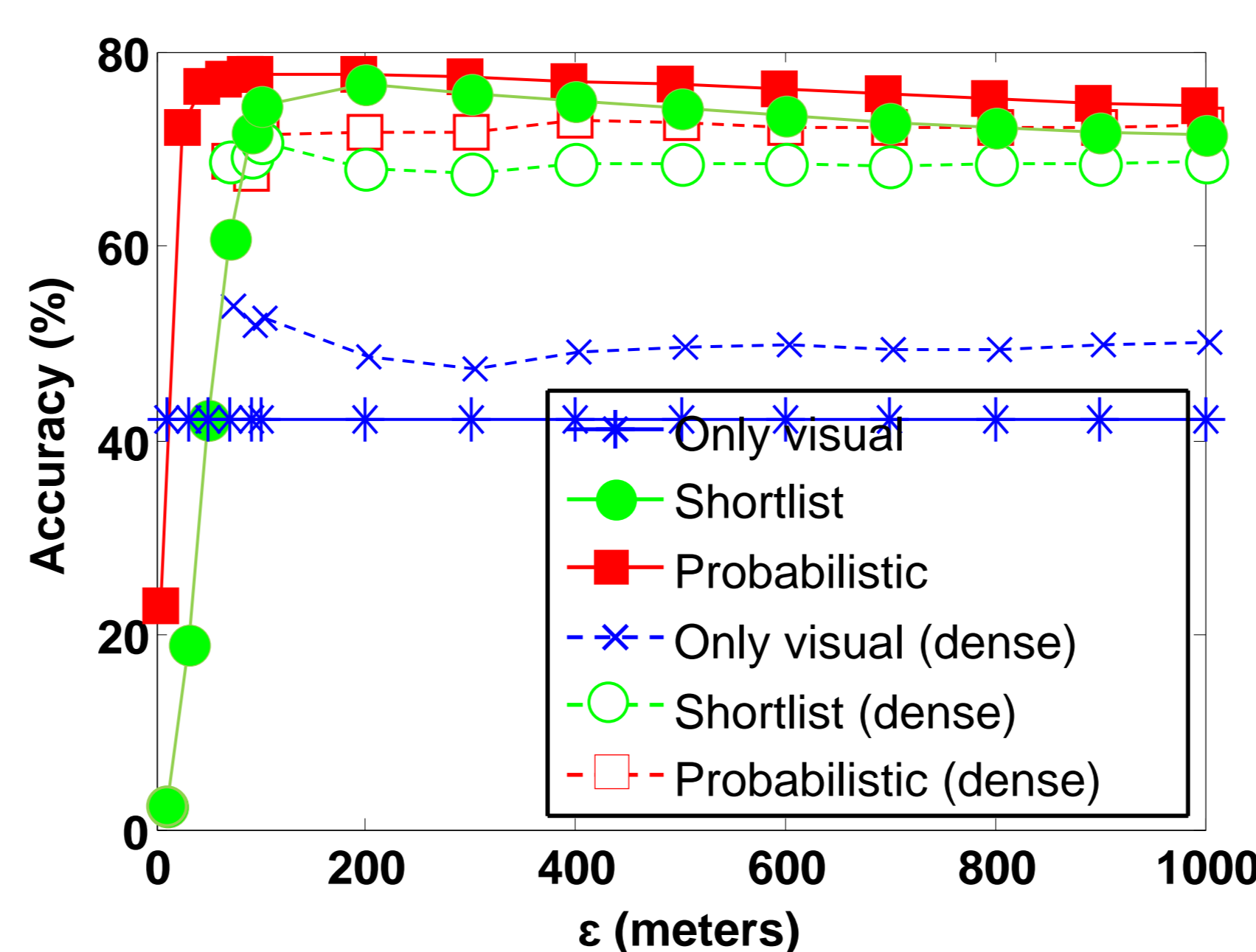
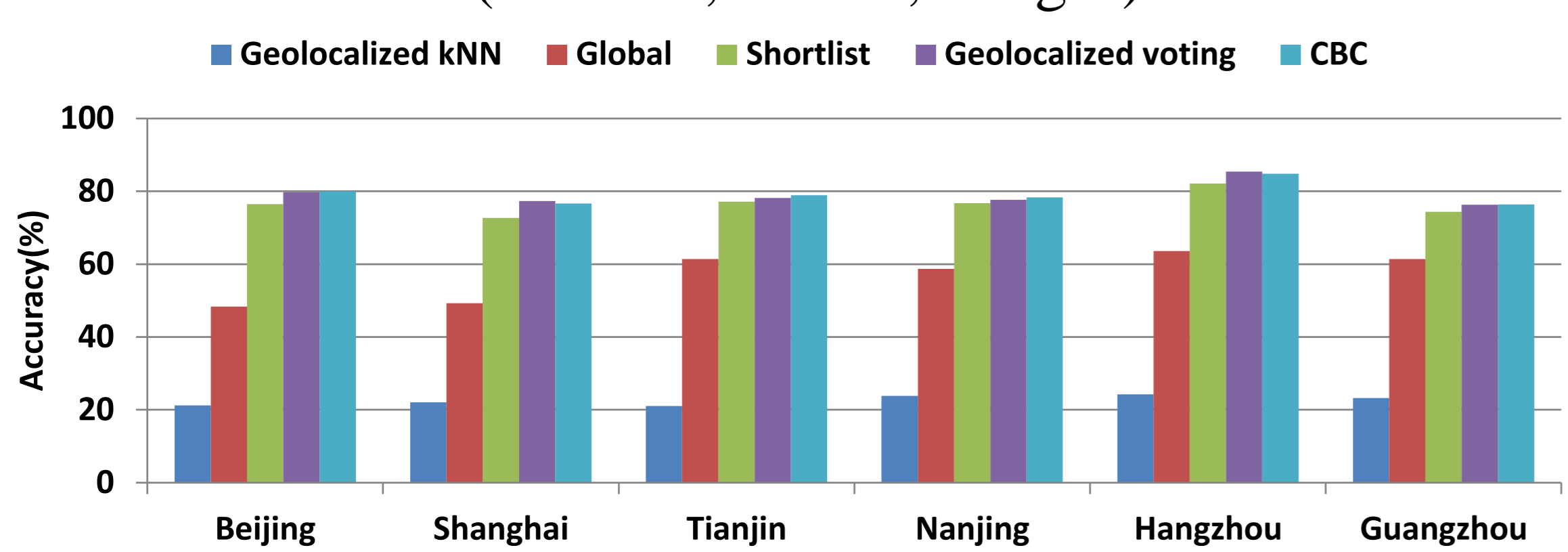
- Dish recognition  $p(s | \mu_q, x)$
- Restaurant recognition  $p(k | \mu_q, x)$
- Location refinement  $p(\varphi | \mu_q, x)$

Approach	Neighborhood model $p(\varphi   \mu_q)$	Rest. location model $p(k   \varphi)$	Visual classifier + menu $p(s   k, x)$
Shortlist	$\mathbb{I}[\ \varphi - \mu_q\  \leq \epsilon]$	$\delta(\ \varphi - \mu_k\ )$	$p_G(s   x) [s \in M_k]$
Probabilistic	$\mathcal{N}(\varphi; \mu_q, \Sigma_q)$	$\mathcal{N}(\varphi; \mu_k, \Sigma_k)$	$p_G(s   x) [s \in M_k]$
Geolocalized	$\mathbb{I}[\ \varphi - \mu_q\  \leq \epsilon]$	$\delta(\ \varphi - \mu_k\ )$	$p_k(s   x)$

### Experiments

Dishes datasets

- Collected from dianping.com
- Six cities
- Restaurants (location, menus, images)



### References

- [1] R. Xu, L. Herranz, S. Jiang, S. Wang, X. Song, R. Jain, "Geolocalized modeling for dish recognition", *IEEE Transactions on Multimedia*, vol. 17, no. 8, pp. 1187-1199, August 2015
- [2] L. Herranz, R. Xu, S. Jiang, "A probabilistic framework for food recognition in restaurants", *Proc. International Conference on Multimedia and Expo*, Torino, Italy, June 2015



MiniEye    Simple Eye    松下新加坡研发中心    三星    光电高斯    理光软件研究所    腾讯    华富睿智