A PROBABILISTIC MODEL FOR FOOD RECOGNITION IN RESTAURANTS

Luis Herranz, Ruihan Xu, Shuqiang Jiang

Institute of Computing Technology
Chinese Academy of Sciences

中国科学院计算技术研究所
Institute of Computing Technology,Chinese Academy of Sciences

Outline

- Introduction
- Proposed approach
- Experiments
- Conclusions

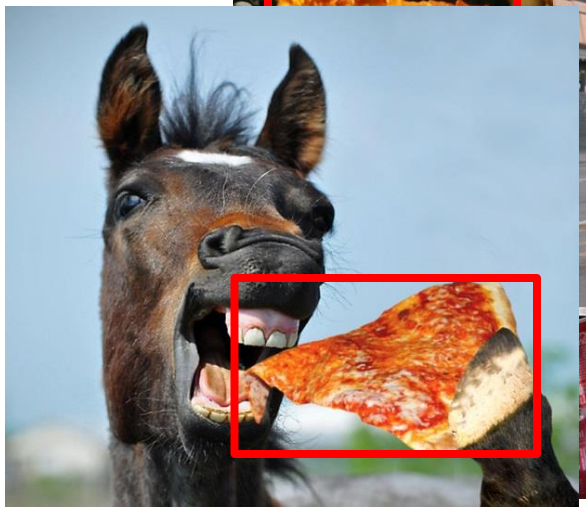
Food recognition

Pepperoni pizza

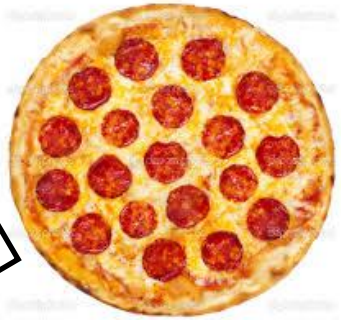
Why food recognition?

- Food images are everywhere
- Many potential applications

Annotation



Retrieval



Pepperoni pizza

Nutrition Facts

Calories 298		(1246 kJ)
Calories from fat 109		
		% Daily Value
Total Fat	12.1g	19%
Sat. Fat	5.3g	26%
Cholesterol	29mg	10%
Sodium	683mg	28%
Total Carbs.	34g	11%
Dietary Fiber	1.6g	6%
Sugars	4.1g	
Protein	13.3g	
Calcium	168.5mg	
Potassium	185.8mg	

Dietary/medical

Similarity and variability

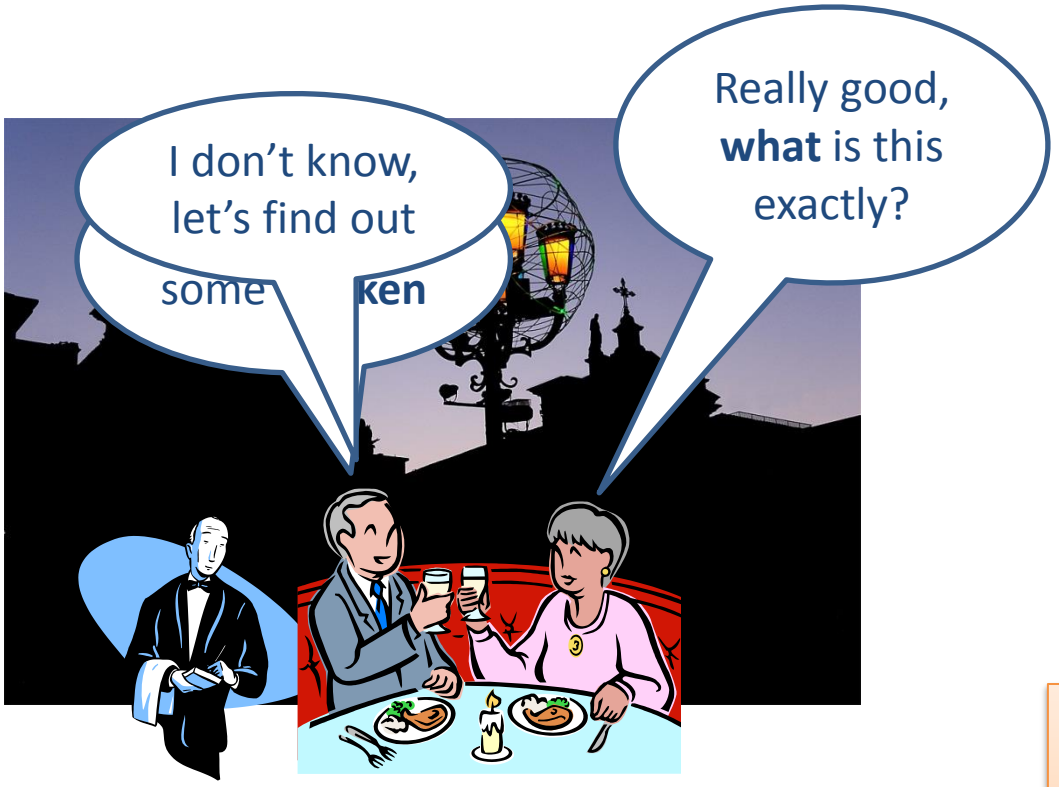
Beef vegetable stir-fry noodles

Spaghetti bolognese

Interclass similarity

Intraclass variability

Dish recognition



More accurate than user

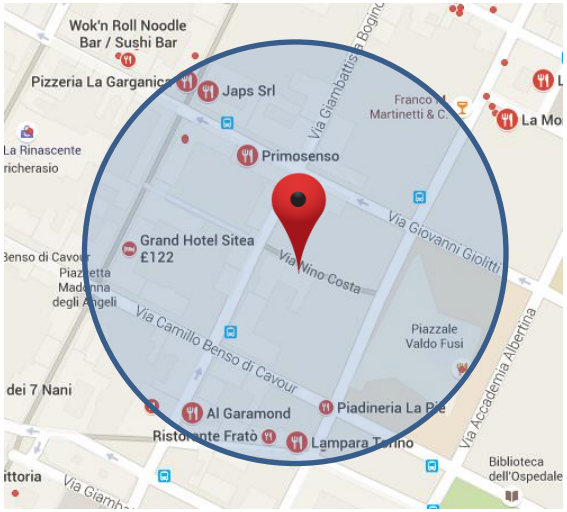
Requires expertise

Automatic (no interaction)

Requires user interaction

Exploiting context and external data

Geo-location

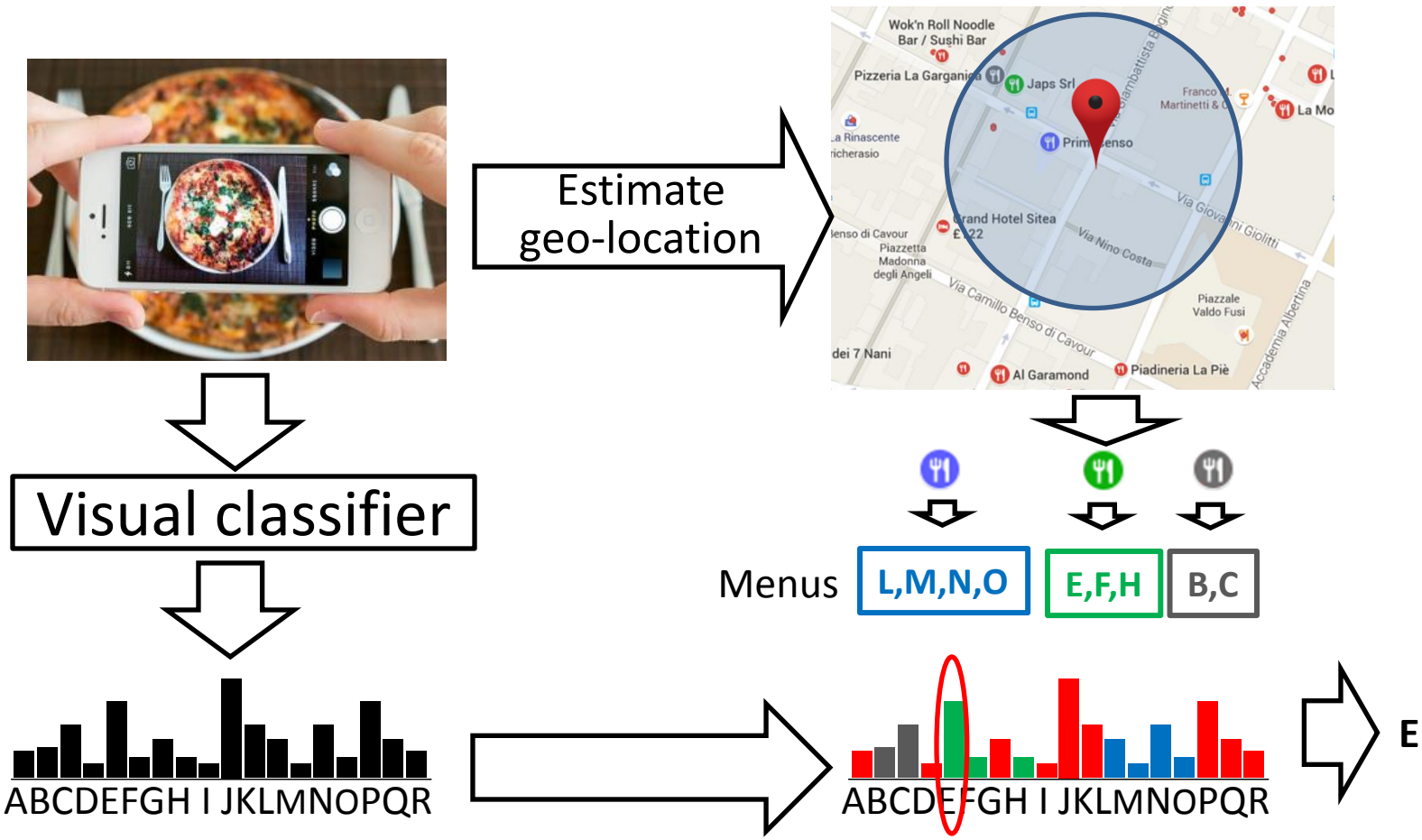


Knowledge about restaurants

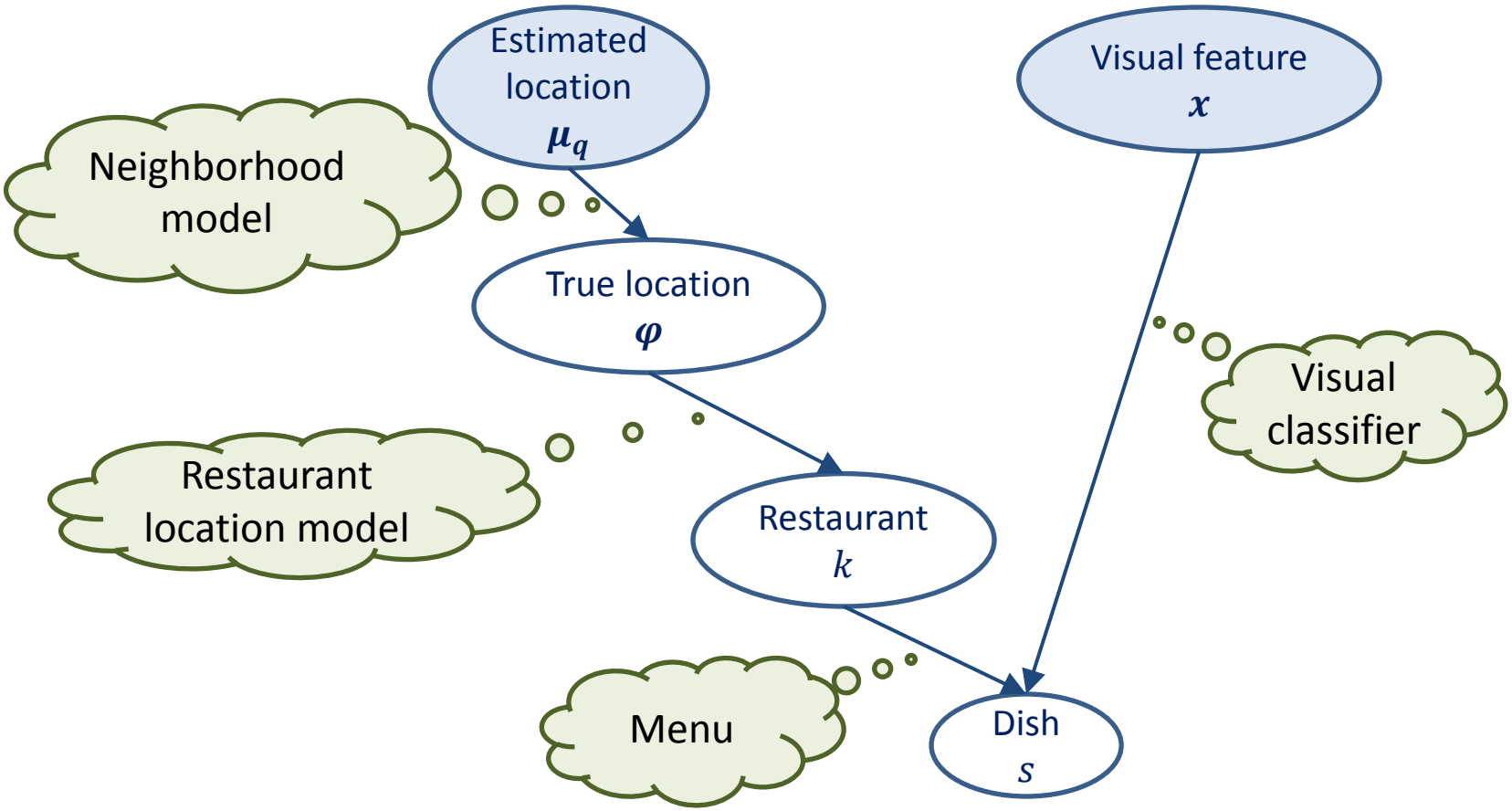
yelp

大众点评
dianping.com

Shortlist approach for dish recognition



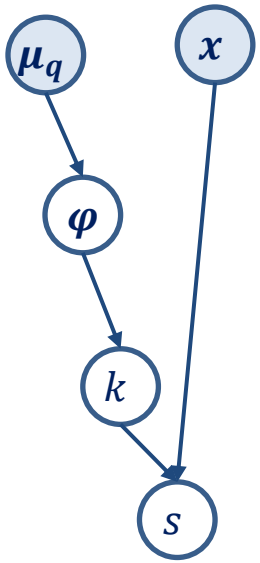
Probabilistic graphical model



Probabilistic graphical model

- If we model $p(s, k, \boldsymbol{\varphi} | \boldsymbol{\mu_q}, \boldsymbol{x})$ (joint distribution)

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



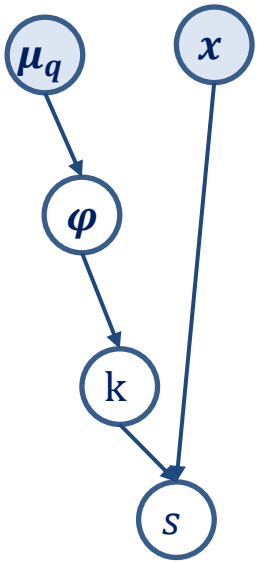
- By marginalizing we can infer either $s, k, \boldsymbol{\varphi}$
  - Dish $p(s | \boldsymbol{\mu_q}, \boldsymbol{x})$
  - Restaurant $p(k | \boldsymbol{\mu_q}, \boldsymbol{x})$
  - Location $p(\boldsymbol{\varphi} | \boldsymbol{\mu_q}, \boldsymbol{x})$

Predicting dish

- Marginalizing $p\left(s, k, \boldsymbol{\varphi} | \boldsymbol{\mu_q}, \boldsymbol{x}\right)$ over k and $\boldsymbol{\varphi}$

$$p\left(s | \boldsymbol{\mu_q}, \boldsymbol{x}\right) = \sum_{k=1}^{K} \int_{\varphi} p\left(s, k, \boldsymbol{\varphi} | \boldsymbol{\mu_q}, \boldsymbol{x}\right) \mathrm{d}\varphi$$

$$= \sum_{k=1}^{K} p(s | k, \boldsymbol{x}) \int_{\varphi} p\left(\varphi | \boldsymbol{\mu_q}\right) p(k | \boldsymbol{\varphi}) \mathrm{d}\varphi$$



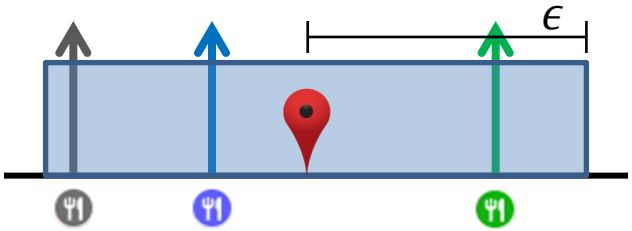
- And the predicted dish is

$$s^* = \mathrm{argmax}_{\mathrm{s}}\, p\left(s | \boldsymbol{\mu_q}, \boldsymbol{x}\right)$$

Revisiting shortlist

- Neighborhood model: hard **circular neighborhood**

$$p(\boldsymbol{\varphi}|\boldsymbol{\mu_q}) = [\|\boldsymbol{\varphi} - \boldsymbol{\mu_q}\| \leq \epsilon]$$



- Restaurant location model: **point**

$$p(k|\boldsymbol{\varphi}) = \delta(\|\boldsymbol{\varphi} - \boldsymbol{\mu_k}\|)$$

- Visual classifier: **global, filtered by menu**

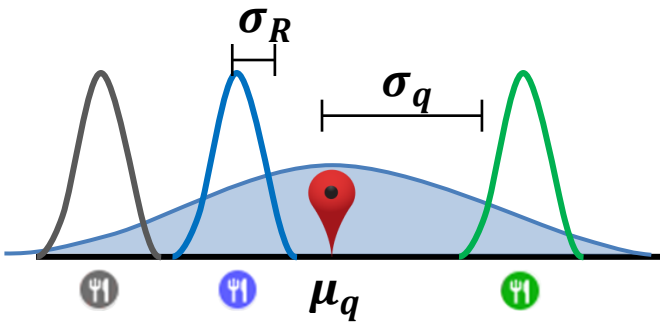
$$p(s|k, \boldsymbol{x}) \propto p_G(s|\boldsymbol{x})[s \in M_k]$$

- Dish conditional probability

$$p(s|\boldsymbol{\mu_q}, \boldsymbol{x}) \propto p_G(s|\boldsymbol{x}) \left[s \in \bigcup_{k \in H_q} M_k \right]$$

with $H_q = \{k|\boldsymbol{\mu_k} \in B_q\}$ and $B_q = \{\boldsymbol{\varphi}|\|\boldsymbol{\varphi} - \boldsymbol{\mu_q}\| \leq \epsilon\}$

Better models (Gaussian)



- Neighborhood model: **Gaussian**

$$p(\boldsymbol{\varphi}|\boldsymbol{\mu_q}) = \mathcal{N}(\boldsymbol{\varphi}; \boldsymbol{\mu_q}, \boldsymbol{\Sigma_q})$$

$$\boldsymbol{\Sigma_q} = \begin{pmatrix} \boldsymbol{\sigma_q} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma_q} \end{pmatrix}$$

- Restaurant location model: **Gaussian**

$$p(k|\boldsymbol{\varphi}) = \mathcal{N}(\boldsymbol{\varphi}; \boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})$$

$$\boldsymbol{\Sigma_k} = \boldsymbol{\Sigma_R} = \begin{pmatrix} \boldsymbol{\sigma_R} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma_R} \end{pmatrix}$$

- Visual classifier: global, filtered by menu

$$p(s|k, \boldsymbol{x}) \propto p_G(s|\boldsymbol{x})[s \in M_k]$$

- Dish conditional probability

$$p(s|\boldsymbol{\mu_q}, \boldsymbol{x}) \propto p_G(s|\boldsymbol{x}) \sum_{k=1}^{K} \frac{[s \in M_k]}{|M_k|} \mathcal{N}(\boldsymbol{\mu_k}|\boldsymbol{\mu_q}, \boldsymbol{\Sigma_q} + \boldsymbol{\Sigma_k})$$

Estimating restaurant and location

- Marginalizing $p\left(s, k, \boldsymbol{\varphi} | \boldsymbol{\mu_q}, \boldsymbol{x}\right)$ over s and $\boldsymbol{\varphi}$

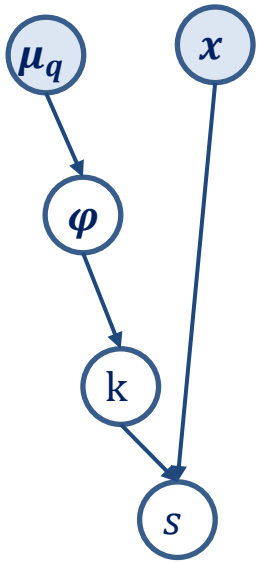
$$p\left(k | \boldsymbol{\mu_q}, \boldsymbol{x}\right) \propto \mathcal{N}\left(\boldsymbol{\mu_k} | \boldsymbol{\mu_q}, \boldsymbol{\Sigma_q} + \boldsymbol{\Sigma_k}\right) \frac{\sum_{s \in M_k} p(s|\boldsymbol{x})}{|M_k|}$$

$$k^* = \operatorname{argmax}_k p\left(k | \boldsymbol{\mu_q}, \boldsymbol{x}\right)$$

- Marginalizing $p\left(s, k, \boldsymbol{\varphi} | \boldsymbol{\mu_q}, \boldsymbol{x}\right)$ over s and k

$$p\left(\boldsymbol{\varphi} | \boldsymbol{\mu_q}, \boldsymbol{x}\right)$$

  - $\boldsymbol{\varphi}$ is continuous
  - Maximum likelihood estimation
  - Not closed form. Iterative procedure



Experiments: Dishes dataset

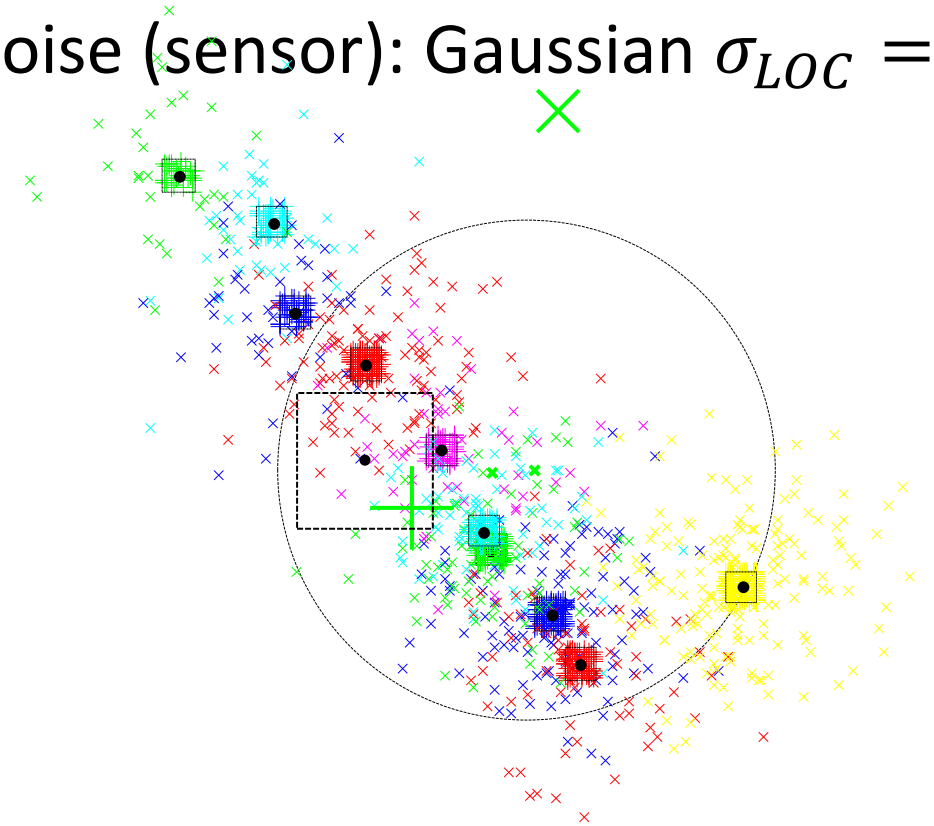
- Restaurants and dishes collected from dianping.com
  - Beijing city
  - Mostly Chinese dishes
- 187 restaurants
  - Menu (≥3 dishes/restaurant)
  - Dish images (≥15 images/dish)
  - Geo-location
- 701 unique dish categories (1173 in total)

Experiments: settings

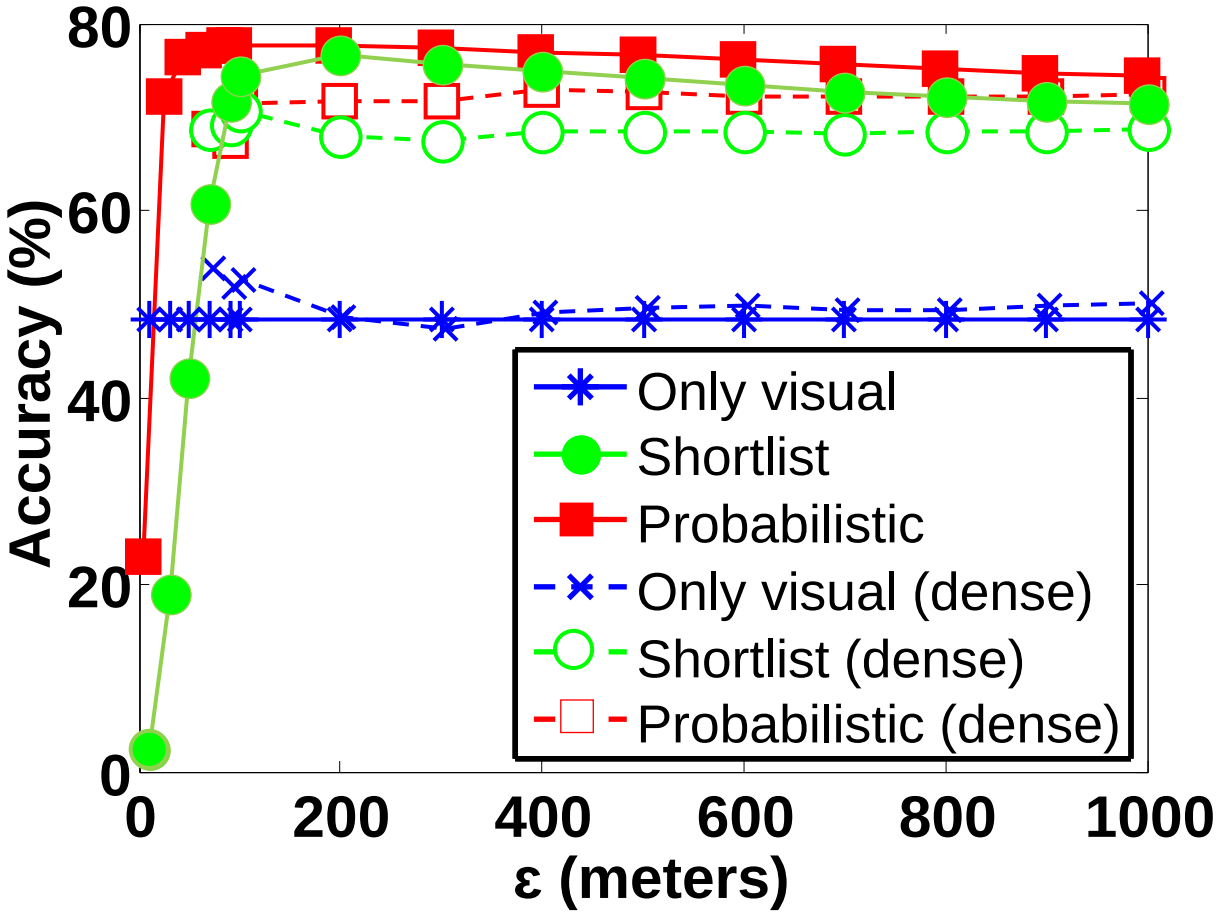
- Visual features: CNN FC7 (DeCAF)
- SVM: #training 10 images/dish (#test ≥ 5 img/dish)
- Dense setting: queries with ≥ 5 restaurants in neighborhood
- To compare shortlist/probabilistic: align parameters as $\epsilon = 3\sigma_q$

Simulating geo-locations for test

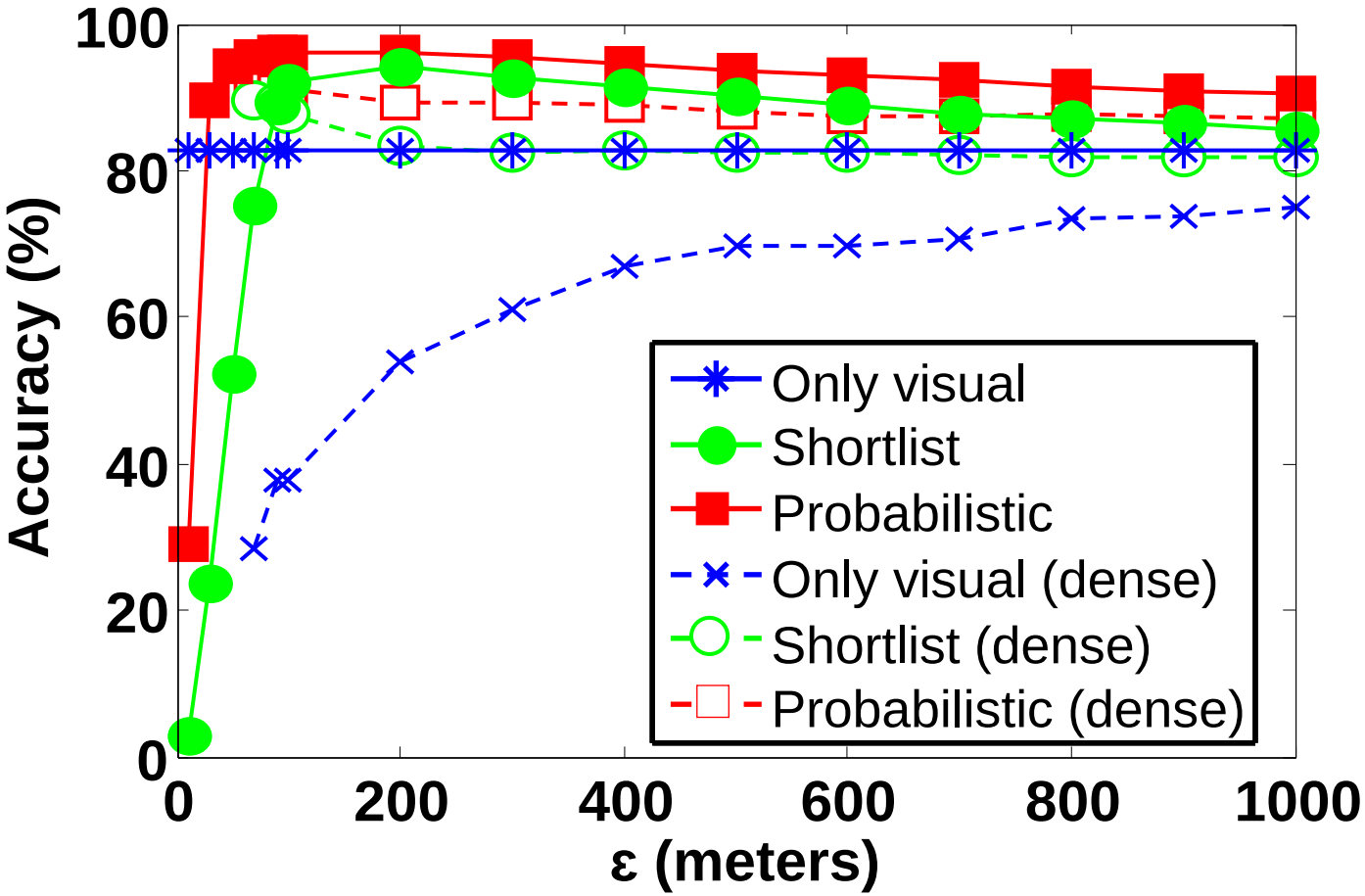
- Restaurants modeled as 25x25 m^2 squares
- Random location in the restaurant: uniform
- Random noise (sensor): Gaussian $\sigma_{LOC} = 40$ m



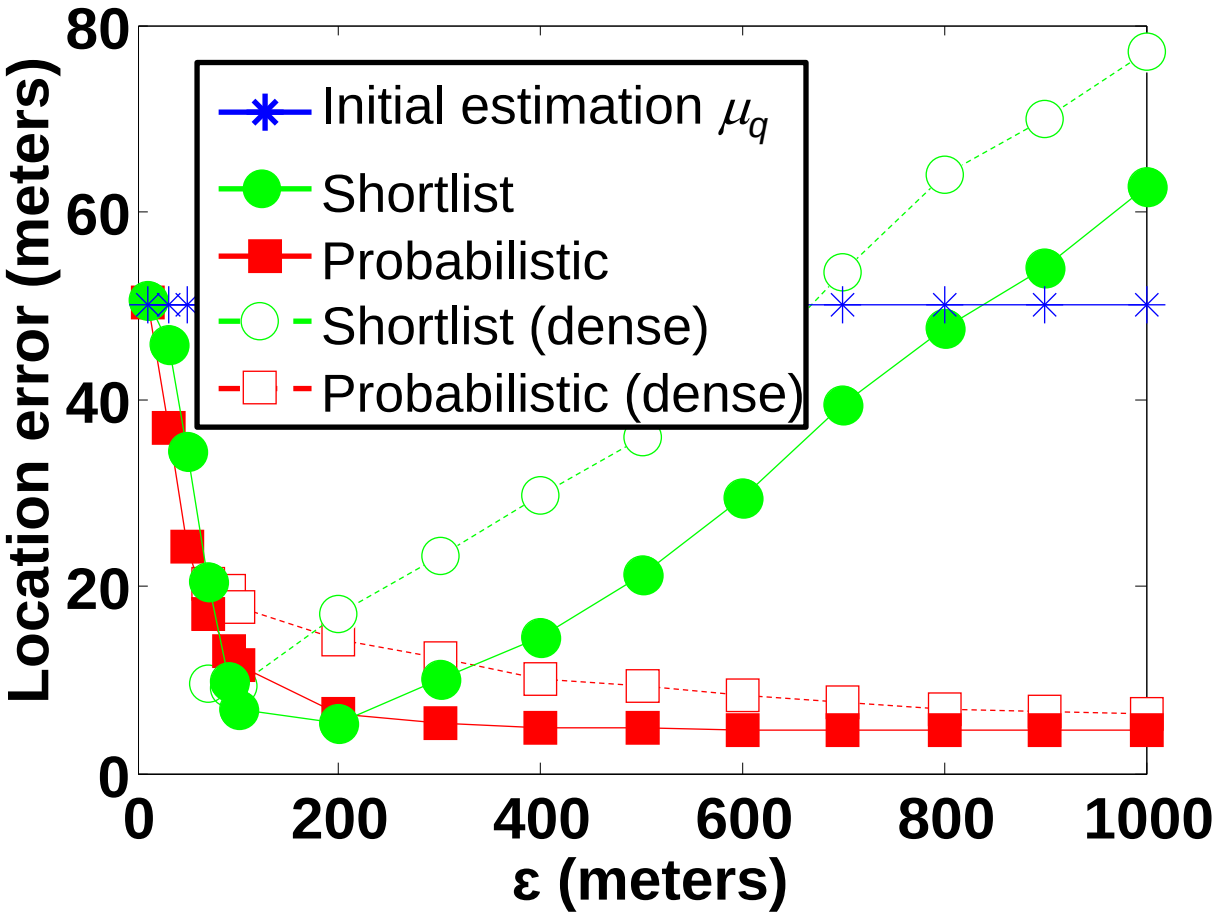
Results: dish recognition



Results: restaurant recognition



Results: geolocation refinement



Conclusions

- Dish recognition in restaurants
  - Food recognition+context+external knowledge
- Better using a probabilistic framework
  - Improves performance
  - More robust to the choice of parameters
- Restaurant recognition and geolocation refinement as byproducts

THANK YOU!