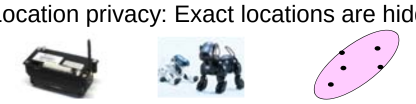
Processing Spatial Queries Based on Uncertain Location Information

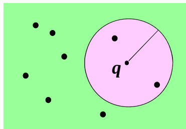
Yuichi Iijima and Yoshiharu Ishikawa (Nagoya University)

Background

- Uncertain Location Information
  - Sensor environments: GPS consumes batteries
  - Mobile robots: Localization may not be accurate
  - Location privacy: Exact locations are hidden

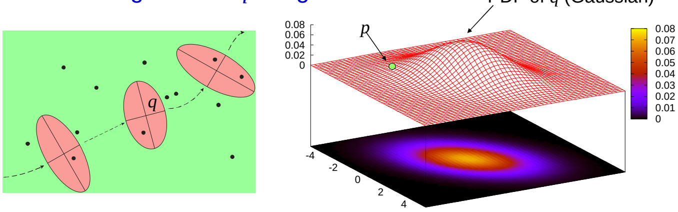


- Location-based Queries
  - Range queries, nearest neighbor queries
  - Spatial index-based processing
  - What's happen for uncertain locations?



Objectives

- Query Processing Based on Uncertain Location Information
  - Location of a query object is specified as a Gaussian distribution
  - Target data: spatial points
- Probabilistic Nearest Neighbor Query (PNNQ)
  - Find objects such that the probabilities that they are the nearest neighbors of q are greater than θ



- Gaussian distribution of query object q

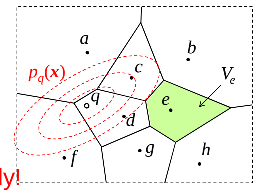
$$p_q(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{q})^t \Sigma^{-1}(\mathbf{x}-\mathbf{q})\right]$$

Naïve Approach

- $\Pr_{NN}(q, o)$: Probability that target object o is the nearest neighbor of query object q
  - Can be calculated by integrating $p_q(\mathbf{x})$ over Voronoi region V_o

$$\Pr_{NN}(q,o) = \int_{V_o} p_q(\mathbf{x})d\mathbf{x}$$

  - If the result is greater than θ, object o satisfies the condition
- Compute $\Pr_{NN}(q, o)$ for each object o using numerical integration: quite costly!



Our Approach

- Use of Filtering
  - Prune non-candidate objects using low-cost filtering conditions
  - Only the remaining candidate objects require numerical integration
  - Filtering should be conservative: no false negatives
- We propose two filtering strategies

Strategy 1: θ-Region-Based Approach

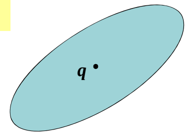
- θ-Region: Ellipsoidal region for which the integration of $p_q(\mathbf{x})$ becomes $1-2\theta$:

$$\int_{(\mathbf{x}-\mathbf{q})^t\Sigma^{-1}(\mathbf{x}-\mathbf{q})\le r_\theta^2} p_q(\mathbf{x})d\mathbf{x} = 1-2\theta$$

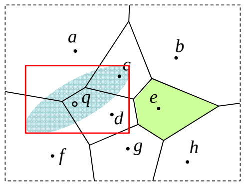
  - Ellipsoidal region

$$(\mathbf{x}-\mathbf{q})^t\,\Sigma^{-1}(\mathbf{x}-\mathbf{q}) \le r_\theta^2$$

  is the θ-region



  - θ-region can be derived using r_θ-table and transformation
- Query Processing
  - Given a query, derive its θ-region and its bounding box
  - Retrieve objects whose Voronoi regions overlap with the box
  - Perform numerical integration for each candidate objects

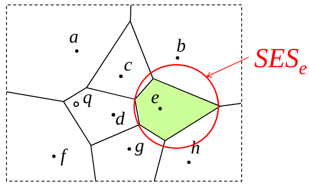


Strategy 2: Use of SES and $p_q^{\top}(x)$

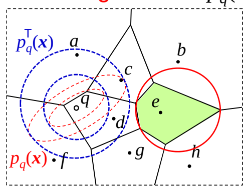
- Compute the smallest enclosing sphere (SES) for each Voronoi region beforehand
  - Idea: Calculate integration

$$\int_{SES_o} p_q(\mathbf{x})d\mathbf{x} > \Pr_{NN}(q,o),$$

  which overestimates $\Pr_{NN}(q, o)$



  - Integration over a spherical region is more easier to compute
- Additional approximation: Use of upper bounding function $p_q^{\top}(\mathbf{x})$
  - It gives the upper bound for $p_q(\mathbf{x})$, and has a spherical isosurface
  - Easy to compute integration using a pre-computed table

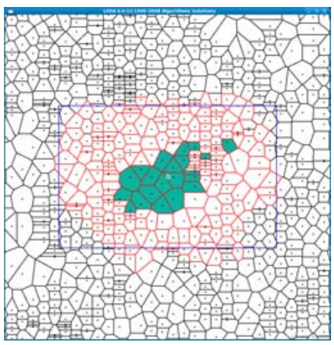


- In summary, we perform two-step approximations

$$\int_{SES_o} p_q^{\top}(\mathbf{x})d\mathbf{x} \ge \int_{SES_o} p_q(\mathbf{x})d\mathbf{x} > \Pr_{NN}(q,o)$$

Experimental Results

- Performance of two strategies depends on parameters and given queries
  - No apparent winner
  - The hybrid strategy shows the best performance
- Query Example (see figure)
  - Enclosing box: bounding box for the θ-region
  - Red cells: candidate cells
  - Green cells: answer cells



Our Related Work

- Y. Ishikawa, Y. Iijima, J.X. Yu, "Spatial Range Querying for Gaussian-Based Imprecise Query Objects", ICDE 2009.
  - Consider the case for range queries