Hybrid Maximum Depth-kNN Method for Real-Time Node Tracking using Multi-Sensor Data

Sudhir Kumar, Abhay Kumar, Akshay Kumar and Rajesh Hegde

Indian Institute of Technology, Kanpur, India

{*sudhirkr, abhayk, akshakr, rhegde*}*@iitk.ac.in,*

IEEE International Conference on Communications (ICC) 2014, London, UK

Overview

Introduction

Motivation and Contributions

Notations and Known Definitions
- Statistical Properties of Depth Functions
- Depth functions
  - Types of depth function

Hybrid Maximum Depth-kNN Method
- Generalized Maximum Depth Method
- Generalized kNN depth based Method
- Hybrid MD-kNN Method using LHFs
- Statistical Properties of the Hybrid MD-kNN Method
- Algorithm for Hybrid MD-kNN for Sensor Localization

Performance Evaluation
- Experimental Conditions
- Experimental Results

Conclusion

Introduction

- Sensor node localization and tracking
- Combines two individual location hypothesis functions (LHF)
- Obtained from generalized maximum depth and generalized kNN methods
- Multiple sensors measure visible light, humidity, temperature, acoustics, and link quality
- Lower complexity of maximum depth and outlier rejection ability of kNN method

Motivation

- Univariate data is widely analysed using various statistical moments, like location, scale, skewness, and kurtosis
- Distance estimation using received signal is erroneous under NLOS conditions
- Moments are difficult to compute for multivariate data
- Generally, we analyse a multivariate data on the presumption of a normal distribution

What is data depth?

- Data depth is an important technique for non-parametric analysis and inference from multivariate data
- Facilitates systematic way of ordering the multivariate data (center-outward ordering)
- Contours based on data depth analysis provide more intuitive visualisation of distributional properties as compared with statistical moments.

Contributions

- Localization using depth as a measure without any assumption of the distribution of multivariate data
- Developed for the multiple hypothesis testing framework
- We generalize the maximum depth and kNN depth based method to account for a vector of multivariate observation under multiple hypothesis
- Generalized maximum depth method requires lesser computation time
- Generalized kNN method is robust in the non-line of sight conditions

Notations and Known Definitions

- Let L distinct grid points in the network are denoted by $\{G_l \triangleq (\alpha_l^1, \alpha_l^2, \ldots, \alpha_l^q)\}_{l=0}^{L-1}$, q dimensional coordinate system..
- At each grid, G_l, n_l number of training points are collected during offline phase.
- N training observation points, $\mathbf{O} = \{X_1, X_2, \ldots X_N\}$.
- Observations are categorised into L grids namely, $G_0, G_1, \ldots, G_{L-1}$ containing $n_0, n_1, \ldots, n_{L-1}$ points respectively such that $\sum_{l=0}^{L-1} n_l = N$.

- Given an observation vector, $X = \{x_1, x_2, \ldots, x_m, \ldots, x_M\}$, where $x_m \in \mathbf{R}^5$, it is localized to one of the grid point.
- Function $l : X \rightarrow Y$ where, $Y \in \{0, 1, \ldots, L-1\}$ that associates an observation vector X with its corresponding nearest grid Y.
- Indicator function is denoted by $\mathbb{I}$

Statistical Properties of Depth Functions

A random vector Z in $\mathbf{R}^d$ is said to be symmetric with respect to the following types of symmetricity.

- Centrally symmetric: About θ if $Z - \theta \stackrel{d}{=} \theta - Z$, where $\stackrel{d}{=}$ denotes equal in distribution.
- Angularly symmetric: About θ if $\frac{(Z-\theta)}{\|Z-\theta\|} \stackrel{d}{=} \frac{(\theta-Z)}{\|Z-\theta\|}$.
- Halfspace symmetric: About θ if $P(Z \in H \geq 1/2)$ for every closed halfspace containing θ.

Depth functions

- Used for center outward ordering of points in $\mathbf{R}^d$
- Let $D(.;.) : \mathbf{R}^d \times \mathscr{F} \rightarrow \mathbf{R}^1$ be mapping of depth of points in $\mathbf{R}^d$, whose distribution is given by $\mathscr{F}$.
- Properties: (a) Affine Invariance:

$$D(Ax + b; F_{Ax+b}) = D(x; F_x) \tag{1}$$

- Maximality at Center:

$$D(\phi; F) = \sup_{x \in \mathbf{R}^d} D(x; F) \tag{2}$$

- Monotonous with respect to the Deepest Point:

$$D(x; F) \leq D(\phi + \alpha(x - \phi); F) \qquad \text{holds for } \alpha \in [0, 1] \tag{3}$$

- Vanishes at Infinity:

$$D(x; F) = 0 \qquad \text{as } ||x|| \rightarrow \infty \tag{4}$$

Location Depth

- For an observation, $x \in \mathbf{R}^d$ with respect to a distribution having probability measure P on $\mathbf{R}^d$, location depth is defined as the minimum of probability mass of any closed halfspace containing x.
- $$\mathcal{L}(x; P) = \inf_H \{P(H) : H \text{ is a closed halfspace, } x \in H\} \qquad (5)$$
- It satisfies all the four desirable properties. Location depth of a point x with respect to an empirically distributed data set in $\mathbf{R}^d$ is defined as the minimum fraction of data points lying on either side of any possible hyperplane passing through x.

Simplicial Depth

- For an observation, $x \in \mathbf{R}^d$ relative to a probability measure P on $\mathbf{R}^d$, it is defined as the probability that x belongs to a random simplex in $\mathbf{R}^d$.

$$\mathcal{S}(x; P) = P(x \in S[X_1, X_2, X_3, ..., X_{d+1}]) \tag{6}$$

- Fail to satisfy the monotonicity property and maximality property for centrally symmetric and halfspace symmetric discrete distributions respectively.
- For an empirically distributed data set, it is the ratio of number of simplices containing x to the total number of possible simplices.

Mahalanobis Depth

- The Mahalanobis distance of a point x from **O** with mean μ and covariance matrix Σ is defined as

$$\mathcal{D}_m(x) = \sqrt{(x-\mu)^\top \Sigma^{-1} (x-\mu)} \tag{7}$$

- Assuming that second moment of **O** exists, Mahalanobis depth of x_i is defined as

$$\mathcal{M}_i = [1 + \underbrace{(x_i - \bar{X})^\top \Sigma^{-1} (x_i - \bar{X})}_{\mathcal{D}_m^2(x_i)}]^{-1} \tag{8}$$

where, $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ and $\Sigma = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(X_i - \bar{X})^\top$.

- Mahalanobis Depth function may fail to satisfy maximality property for angularly symmetric distributions.

L_1 Depth

- For an observation, x, with respect to $\mathbf{O}$ in $\mathbf{R}^d$, L_1 depth is defined as one minus average of the unit vectors, directed from x towards all observations points.

$$L_1(x, \mathbf{O}) = 1 - \left\| \frac{1}{N} \sum_{i=1}^{N} u_i(x) \right\| \tag{9}$$

where, $u_i(x) = \frac{x - X_i}{\|x - X_i\|}$ denotes the i^{th} unit vector.

$$0 \leq L_1(x, \mathbf{O}) \leq 1 \tag{10}$$

- It violates affine invariance property.

Oja Depth

- For an observation, $x \in \mathbf{R}^d$ with respect to **O**, whose distribution is given by F, it is defined as

$$\mathcal{O}(x; F) = [1 + E_F[v(S[x, X_1, \ldots, X_d])]]^{-1} \qquad (11)$$

  where, $S[x, X_1, \ldots, X_d]$ represents the closed simplex with vertices x and random d points from the **O**. Expected volume with respect to distribution, F, is represented by $E_F[v]$.
- For an empirically distributed sample in $\mathbf{R}^2$, Oja depth of a sensor node position x is defined as [**?**]

$$O(x, F) = \left[1 + \frac{1}{\binom{n}{2}} \sum_{\{i,j\} \in \{1,2,\ldots,n\}} \text{Area of traingle with vertices } x, X_i, X_j \right]^{-1} \qquad (12)$$

- Oja depth function may violate affine invariance property.

Generalized Maximum Depth Method

- Binary hypothesis problem,

$$\hat{h}_D(X; G_0, G_1) = \mathbb{I}\Bigg[\overbrace{\sum_{i=1}^{M} \mathbb{I}\left[D(x_i, G_1) > D(x_i, G_0)\right]}^{n(X,G_1)} > \underbrace{\sum_{i=1}^{M} \mathbb{I}\left[D(x_i, G_0) > D(x_i, G_1)\right]}_{n(X,G_0)} \Bigg] \tag{13}$$

where, $n(X, G_l)$ = number of points in observation vector X having higher depth with respect to G_l $\forall\ l = \{0, 1\}$

- For a multivariate observation vector and multiple hypothesis

$$\hat{h}_D(X; G_0, G_1, \ldots, G_{L-1}) = \underset{l}{\operatorname{argmax}} \Big[n(X, G_l) \Big] \tag{14}$$

Generalized kNN depth based Method

- Binary hypothesis problem,

$$\hat{m}_D^{\beta}(X; G_0, G_1) =$$

$$\mathbb{I}\Bigg[\sum_{j=1}^{M} \overbrace{\mathbb{I}\left(\sum_{i=1}^{N} \mathbb{I}[Y_i = 1]\, W_i^{\beta}(x) > \sum_{i=1}^{N} \mathbb{I}[Y_i = 0]\, W_i^{\beta}(x)\right)}^{\hat{m}_D^{\beta}(x_j)} > \sum_{j=1}^{M} \underbrace{\mathbb{I}\left(\sum_{i=1}^{N} \mathbb{I}[Y_i = 0]\, W_i^{\beta}(x) > \sum_{i=1}^{N} \mathbb{I}[Y_i = 1]\, W_i^{\beta}(x)\right)}_{1 - \hat{m}_D^{\beta}(x_j)}\Bigg] \quad (15)$$

with $W_i^{\beta}(x) = \frac{1}{K_x^{\beta}} \mathbb{I}\left[X_i \in \mathbf{R}_x^{\beta}\right]$ where $K_x^{\beta} = \sum_{j=1}^{N} \mathbb{I}\left[X_j \in \mathbf{R}_x^{\beta}\right]$ denotes total number of points in $\mathbf{R}_x^{\beta}$.

- For multiple hypothesis,

$$\hat{m}_D^{\beta}(X; G_0, G_1, ..., G_{L-1}) = \underset{l}{\operatorname{argmax}} \left[\sum_{j=1}^{M} \mathbb{I}\left(\sum_{i=1}^{N} \mathbb{I}[Y_i = l] W_i^{\beta}(x)\right)\right]$$

WSN Lab

Hybrid MD-kNN Method using LHFs

- $$\hat{H}_D(X; G_0, \ldots, G_{L-1}) = \overbrace{\hat{h}_D(X; G_0, \ldots, G_{L-1})\mathbb{I}\Big[D(X) > \tau * MD\Big]}^{LHF_1} + \underbrace{\hat{m}_D^{\beta(n)}(X; G_0, \ldots, G_{L-1})\mathbb{I}\Big[D(X) \leq \tau * MD\Big]}_{LHF_2} \quad (17)$$

  Where MD denotes the maximum depth value obtained from the observation collected at grid points, G_l $\forall$ l.

Statistical Properties of the Hybrid MD-kNN Method

- Affine Invariance: **Proof**: Let $X \mapsto Y$ and $Y = AX + b$ represent the affine transformation. $\hat{h}_D(x; G_{1x}, G_{0x})$ denote the maximum depth method for original observation and $\hat{h}_D(y; G_{1y}, G_{0y})$ denote the maximum depth method for the affine-transformed observation.

$$\begin{aligned} \hat{h}_D(y; G_{1y}, G_{0y}) &= \mathbb{I}\left[D(y, G_{1y}) > D(y, G_{0y})\right] \\ &= \mathbb{I}\left[D(x, G_{1x}) > D(x, G_{0x})\right] \\ &= \hat{h}_D(x; G_{1x}, G_{0x}) \end{aligned} \tag{18}$$

- Robustness:

Algorithm for Hybrid MD-kNN for Sensor Localization

1: **Input** : Grid points, $\{G_l \triangleq (\alpha_l^1, \alpha_l^2, \ldots, \alpha_l^q)\}_{l=0}^{L-1}$ and τ

2: **Measurement Acquisition** : Each element of observation vector is a multivariate observation containing visible light, temperature, humidity, acoustic signal and link quality as labels.

3: **Offline Training** : Multivariate observation vectors are recorded at each grid point. At each of the grid, G_l, n_l training points are collected during offline phase.

4: **Iteration** : Repeat for all the nodes at each time instant.

5: **Online Testing** : Record online observation vector, $X = \{x_1, x_2, .., x_M\}$, where $x_m \in \mathbf{R}^5$, at time t corresponding to a particular sensor node.

6: **Node Localization** : Localize to one of the nearest grid point using the hybrid MD-kNN method in Equation 17.

7: **Termination** : Output, the location of all nodes.

Experimental Conditions

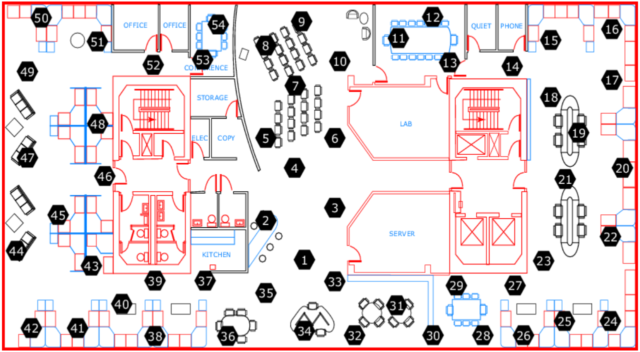

Figure 1: Figure illustrating the deployments of fifty four sensor nodes in Intel Berkeley Research lab. Reproduced from [?].

Experimental Set-up

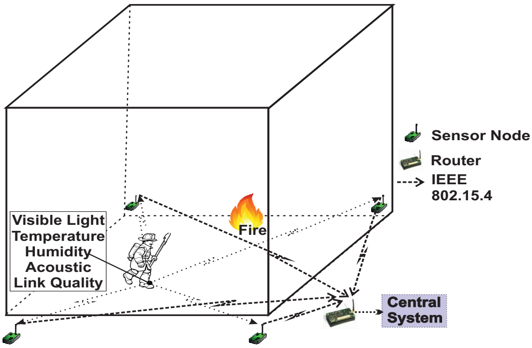

Figure 2: Figure illustrating the experimental set-up for firefighter tracking in indoor scenario.

Comparison of Localization Accuracy and Probability of Resolution

Table 1: Comparison of localization accuracy and probability of resolution for various methods using different depth functions, $\tau = 0.2$

	Localization Accuracy (m)			Probability of Resolution		
Depth Functions	**MD**	**MD-*k*NN**	***k*NN**	**MD**	**MD-*k*NN**	***k*NN**
Location Depth	1.1598	0.8805	0.5445	0.83	0.87	0.91
Simplicial Depth	0.8475	0.7556	0.5245	0.88	0.89	0.90
Mahalanobis Depth	1.4969	1.3014	0.5794	0.78	0.80	0.89
Oja Depth	2.1092	2.1092	0.5840	0.69	0.69	0.90
L_1 **Depth**	2.7202	1.8094	0.6370	0.63	0.75	0.89

Time Complexity Analysis for Hybrid MD-kNN Method

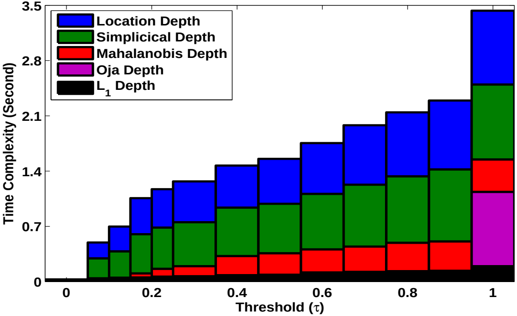

Figure 3: Figure illustrating the time complexity of hybrid MD-kNN method for various depth functions. The value of $\beta = \frac{k}{N} = 0.005$.

Variation of Localization Error with Threshold

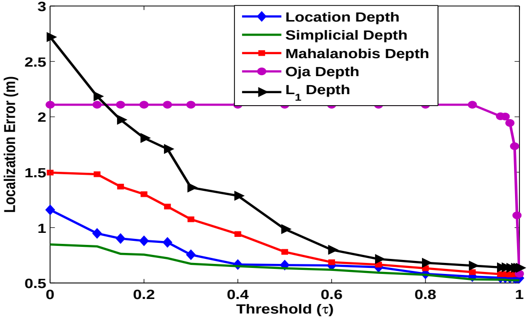

Figure 4: Figure illustrating the variation of localization error (m) with threshold of hybrid MD-kNN method for various depth functions. The value of $\beta = 0.005$

Variation of Probability of Resolution with Threshold and Time

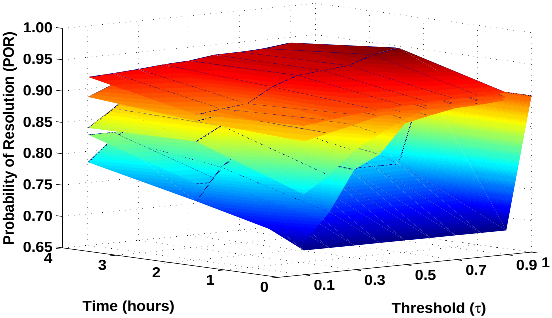

Figure 5: Figure illustrating the variation of probability of resolution with threshold and time of hybrid MD-kNN method, for $\beta = \frac{k}{N} = 0.005$. The depth functions used are simplicial, location, Mahalanobis, Oja and L_1 from top to bottom respectively on time axis.

Conclusion

- Hybrid MD-kNN method for real time sensor node tracking
- Inexpensive and more accurate for sensor node localization
- Future work includes the utilization of functional data with band depth techniques
- Accurately predicts the location of node with malicious sources is also currently being investigated

References I

[1] A. Baggio and K. Langendoen.
Monte carlo localization for mobile wireless sensor networks.
Ad Hoc Networks, 6(5):718–733, 2008.

[2] T. Camp, J. Boleng, and V. Davies.
A survey of mobility models for ad hoc network research.
Wireless communications and mobile computing, 2(5):483–502, 2002.

[3] C.-Y. Chang, C.-Y. Chang, and C.-Y. Lin.
Anchor-guiding mechanism for beacon-assisted localization in wireless sensor networks.
Sensors Journal, IEEE, 12(5):1098–1111, May 2012.

[4] http://adrianlatorre.com/projects/pfc/#/index.

[5] L. Hu and D. Evans.
Localization for mobile sensor networks.
In *Proceedings of the 10th annual international conference on Mobile computing and networking*, pages 45–57. ACM, 2004.

[6] S. Kumar, S. Tiwari, and R. Hegde.
Optimal anchor guiding algorithms for maximal node localization in mobile sensor networks.
In *Wireless Sensor (ICWISE), 2013 IEEE Conference on*, pages 13–18, Dec 2013.

Thank You!